Evolutionary and functional studies of p47 GTPases involved in cell autonomous immunity

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Cemalettin Bekpen

aus Bolu, Turkei

Köln 2005

Berichterstatter: Prof. Dr. Jonathan C. Howard
Prof. Dr. Diethard Tautz

Tag der mündlichen Prüfung: 14.02.2006

Dedicated to My Mother and True Love

OLD MAN

Old man is walking.
Has a dream in his pocket
which he eats when he is hungry.

Old man is thinking
Thinks to force impossibility
Which he couldn't do when he is child.

Old man is looking
Has a view in his mind
Which no body can see the same.

Old man is smiling
Smiles because he has just frozen the time
Which is not more than his age.

Cologne
21-11-03

TABLE OF CONTENTS

ABBREVATIONS

IFN-γ	Interferon-γ
IFN-α/β	Interferon-α/β
IFNGR	FN-γ receptor
IFNAR	IFN-α receptor
iNOS	Inducible nitric oxide synthetase
NRAMP1	Natural resistance associated membrane protein 1
PKR	Protein kinase R
IDO	Indolamine 2,3-dioxygenase
2'-5'-OAS	2'-5'-oligoadenylate synthetase
Phox	phogosome oxidase
GAP	GTPase activating protein
GBP	guanylate binding protein
GEF	guanine nucleotide exchange factor
GED	GTPase effector domain
ATP	adenosine triphosphate
GDP	guanosine diphosphate
GMP	guanosine monophosphate
GTP	guanosine triphosphate
BSA	bovine serum albumine
FCS	fetal calf serum
rpm	rounds per minute
RT	room temperature
OD	optical density
ON	over night
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PH	pleckstrin homology domain
EG	Effector genes
SDS	sodium dodecylsulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
U	unit
UV	ultraviolet
WT	wild type
IF	immunofluorescence
N-terminal	amino-terminal
C-terminal	carboxy-terminal

I.INTRODUCTION

Life on earth began about 3.5 billion years ago from a single replicating unit (Schopf, 1993). From the Precambrian period until now life is represented by more than 1.5 million described species and the actual number of species is expected to be more than 10 million (Wilson, 2000). From Larmarck and Darwin on, nearly all of the leading evolutionary biologists believed that main source of this complex diversity of life is evolutionary change. As summarized by Dobzhansky: "nothing makes sense in biology except in the light of evolution," Evolution in biology is defined as change in diversity and adaptation in populations of organisms (Mayr, 1978) (Dobzhansky, 1973). However, how the evolutionary changes have been maintained since the beginning of life is a hard question that may never be answered completely (Lewontin, 2002). Natural selection, the primary causal influence of phenotypic evolutionary changes, is the basis of adaptation (Dobzhansky, 1982), (Mayr, 2001), (Lewontin, 1978). Thus, organismal diversity is directly dependent on adaptation of organisms to different conditions occurring throughout the course of evolution.

Adaptive capability of organisms underlies the genetic composition of population as well as environmental interactions and varies from species to species (Dobzhansky, 1982), (Lewontin, 1978). The degree of adaptation is especially important in the co-evolutionary process where two organisms have direct and dynamic interactions with each other, as in host-pathogen interaction. In host pathogen interaction, two independent organisms with their specific adaptation capacity become adaptively interrelated and they start evolving under the selective conditions imposed by each on the other. As a result, both organisms pose continuous positive or negative selection force on each other. Perhaps this is one of the most effective processes at speciation, since fluctuation of two organisms in a population must be continuous from the beginning of their interaction (Haldane, 1949; Rausher, 2001b).

I.1.Host-Pathogen Coevolution

Host pathogen coevolution is the parasitic exploitation of one organism by another. This kind of coevolution requires direct interaction of two species with each other as in a never ending battle.

It was Haldane (Haldane, 1949) who first stated that host-pathogen interactions generate diversity both within and between species which not only keeps the species variable, but also leads to speciation. When the pathogen attempts to exploit resources of the host, it gets a tremendous selective pressure, and conversely defense against the pathogen drives

selective pressure on pathogen (Rausher, 2001b; Summers et al., 2003). This antagonistic, and direct relationship brings high fitness costs for both the pathogen and the host (Tian et al., 2003) (Rigby et al., 2002). To avoid this high cost for survival both organisms (pathogen and host) prefer to undergo coadaptation. It is expected that host and pathogen coadapt to each other in two different ways. The first way follows from directional selection leading to an arms races and the second way from heterozygote advantage or negative frequency dependent selection, leading to population diversity and transmission problems for the pathogen (Rausher, 2001a; Summers et al., 2003).

Arms races can be any type of adaptation in order to avoid or eliminate the pathogen, grouped under organisms undergoing external coevolution such as increase in fitness, change in structure, behavior, a robust immune system and internal coevolution such as recognition, destruction of pathogen at the cellular level (Fig 1). The escalating arms races generally lead to low level of polymorphism, whereas the negative frequency dependent selection gives higher level of polymorphism since it is working in a statistical way at the populational level.

Figure 1. Host-parasite interactions (Rigby et al., 2002),
Parasites infect a host by penetrating thorough the external defenses five different ways (indicated by the purple rugby shirts on the front line players). The internal defenses system represented by cell autonomous immunity; recognition system and effector system (see below) must be defeated (represented by the yellow rugby shirts on the back line) fight with the parasites before the onset of an infection. On the other hand, adaptive evolution by arms race can either be at the level of the above described external or internal defense systems. Cartoon by Neil Smith.

As it is already stated by Haldane "it is much easier for a mouse to get a set of genes which enabled it to resist *Bacillus typhimurium* than a set which enabled it to resist cat" (Haldane, 1949). Organismal internal coevolution, also named as molecular coevolution, is

the first step in the process of coadaptation. It is very different from the organismal external coevolution, which usually entails high fitness costs. During host pathogen interaction the host tries to reduce the attacks from the pathogen by building a proper defense system, which increases the organism's fitness. This leads to generation of organismal diversity as seen in the large differences in resistance between different breeds of mice to a variety of pathogens (Haldane, 1949).

Molecular coevolution (internal defense system) derived by arms races can be explained in two different ways: the first being gene to gene type of molecular coevolution mainly established by early studies on plant-pathogen interaction especially crop plants (Flor, 1971; Summers et al., 2003). In this type of interaction, there are multiple loci in both host and pathogen. For each locus in the host, there is a corresponding locus in pathogens. This type of interaction is usually related with low level of polymorphism. The second type is called as matching allele type molecular coevolution, mainly dependent on higher rate of polymorphism (Frank, 1994). Antagonistic host pathogen interactions are maintained by corresponding loci on each side. In both host and pathogen, there are multiple resistance alleles and virulence alleles respectively. If the pathogen allele matches with resistance allele in the host, then resistance to pathogen is induced. When the pathogens interact with the host, they have to be first recognized and then eliminated.

In a classical battle between two enemies, there are two crucial steps, information, and destruction of the enemy. In order to eliminate your enemy, you have to have better information (self-nonself discrimination) and available army to destroy your enemy. Thus, there must be two steps for direct or indirect antagonistic interaction suggesting that two steps for adaptation. As a result, arms races are performed in two steps; firstly, information exchange between two species (pathogen and host indicated by number 1 in figure 2) and second, is function of host effectors which induce the elimination of the pathogen (indicated by number 2 in figure 2) (Trowsdale and Parham, 2004) (Rausher, 2001a; Rausher, 2001b).

Pathogens evolve to defend themselves by various mechanisms such as specialized mechanisms forming a high rate of diversification, by mimicking the host system, escaping the host recognition system (antigenic drift) and interfering with the host defense mechanisms. Host learns to get the information from the pathogen to distinguish self from nonself by using combination of highly variable recognition systems which leads to elimination of the pathogen by generating powerful and alternative destruction system thorough subsequent signaling pathways (Fig 2) (Berriman et al., 2005; Borst, 2002; Rausher, 2001b; Trowsdale and Parham, 2004) (Charles A. Janeway 2005; Galan and Bliska, 1996).

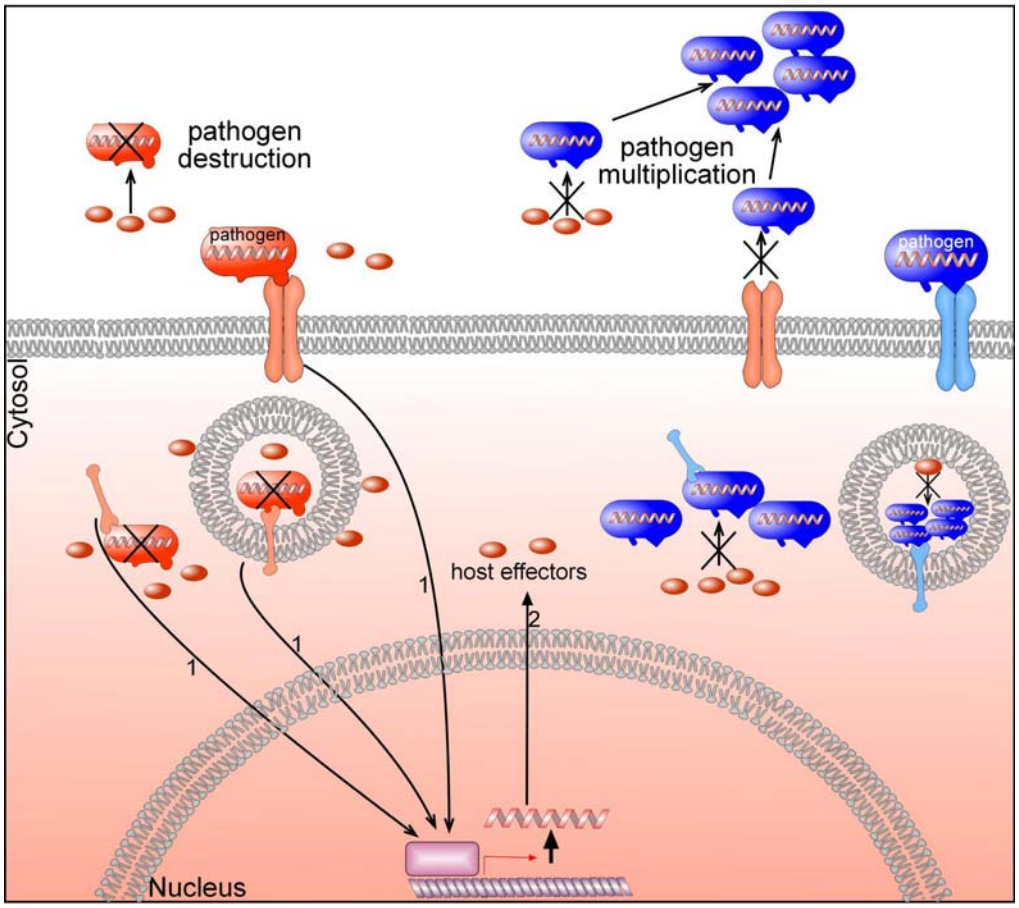


Figure 2. Simplified scheme of the host-pathogen coadaptation at the molecular level.
Arrows marked by number 1 indicates the recognition system and number 2 indicates the host effectors upregulated by recognition system (1). Pathogens are first recognized and the genes or gene families are induced by subsequent signaling pathways. Pathogen which is not recognized or destroyed by the host can multiply (blue). Whereas pathogen which is recognized or destroyed by the host cannot survive (red). By antagonistic direct interactions, host-pathogen coadaptation may occur as indicated in red colors for both pathogen and the host.

A striking example of arms races between a virus and host immunity is the murine cytomegalovirus; In susceptible mice, to turn off host NK cells, murine cytomegalovirus expresses a substitute class I molecule, m157, that binds to inhibitory receptor Ly49i, whereas resistant mice encodes an activating receptor, Ly49h, providing a counter strategy. These two receptors are highly homologous to each other suggesting that they have common evolutionary origin and evolved in response to selective pressure imposed by the pathogen (Vivier and Biron, 2002).

It is likely that the possession of a good destruction or invasion system for host and pathogen creates a high fitness cost under certain conditions. Otherwise one side would go to fixation, as is the case for fitness cost for having proper defense against pathogen (Rigby et al., 2002), (Burdon and Thrall, 2003). For example, plasma membrane protein in *A. thaliana*, RPM1, is responsible for recognition of *P. syringae* (pathogen for plants). Susceptible inviduals lack the entire coding region of *RPM1* and both susceptibility and resistance alleles

frequently occur together within natural populations. Tian et al., generated independent transgenic lines carrying *RPM1* and showed that all the transgenic plants have fitness loss such as 9% reduction in total seed production (Tian et al., 2003). Similarly, *Mx1* is resistance factor against variety of viruses in mouse such as influenza A and B (see below). The mouse carrying $Mx1^-$ allele is susceptible to influenza virus. The standard laboratory mouse strains all carry the $Mx1^-$ allele except A2G and SL/NiA mice. However, wild mice possess the Mx^+ and Mx^- alleles at roughly equal frequencies (Staeheli et al., 1988), (Haller et al., 1987), (Jin et al., 1998a). This suggests that *Mx1* gene like *RPM1* might create high fitness cost. However, there is no direct evidence for the fitness cost specific to *Mx1* gene in mice.

The battle between host and pathogen is mainly carried out by molecular interactions. These interactions reflect the co-evolutionary balance that the host and pathogen must reach in order to secure their survival. Such interactions are usually maintained by the proteins which are encoded on single genes or gene family in one or multiple locus in both. The genes underlying the host defense includes the substantial proportion of the genome. It is estimated that in Arabidopsis, 14 % of the 21000 genes are directly related to pathogen resistance, and in mice 50 loci distributed over 17 chromosomes are known to be involve in resistance against retroviruses alone. It is reported that at least 1000 genes are upregulated upon interferon stimulation in mouse (Bevan et al., 1998),(Bishop et al., 2000), (O'Brien, 1988), (Boehm et al., 1997).

The fate of a gene sequence or gene family through evolutionary time is determined by a combination of processes. Random genomic events, mutation, recombination, duplication, transposition and loss under the selective processes whether neutral or natural, determine the trajectory of the sequence and its derivatives through the generations. Since neither the genomic processes, nor the selective fate of their derivatives are replicated in multiple evolving lineages, the representation of an ancient gene in modern descendent groups of organisms can be surprisingly various. This is extreme in immunity related genes whose products contribute to host pathogen resistance

I.2.Immunity and Immunity Related Genes

Immunity is the state of protection from infections and tumors. The recognition of the pathogen by the immune system results in the induction of defense mechanisms leading to the destruction of the infectious agents. The defense mechanism is highly dependent on the infectious agent, usually fast evolving, because of the short generation time and high adaptive

capacity. Therefore, immunity related genes are frequently rapidly evolving, resulting in formation of different mechanisms and complex systems to fight against infectious agents.

Organisms especially the higher eukaryotes have generated two distinct types of immune systems. The effector immune mechanism, clearly first recognized by Janeway (Janeway, 1989) we call today as innate immunity. Recent studies reveal that some forms of innate immunity are present almost in all types of eukaryotes (Medzhitov and Janeway, 2000) (Hoffmann et al., 1999) (Janeway, 1989). Vertebrates have an additional highly sophisticated immune mechanism, generating adaptive immunity (Pancer et al., 2004), (Flajnik and Du Pasquier, 2004). In adaptive immunity, specified cell clones devoted to defense have an ability to recognize different subtypes of pathogens. Although, it is not possible to separate these two immune systems completely, the most striking difference between adaptive and innate immunity is in the generation of recognition systems. In the innate immune system recognition is mediated by germ line encoded receptors (e.g. TLR (Medzhitov and Janeway, 1999) (Kimbrell and Beutler, 2001), NOD (Ogura et al., 2001), Scavenger Receptors (Pearson, 1996)). This means that the specificity of receptors is genetically predetermined. These receptors can recognize patterns that are general to pathogens such as lipopolysaccaride (LPS)/Pathogen associated molecular patterns (PAMPs). Therefore, pathogen recognition receptors (PRR) are essential players in innate immunity. However, the recognition systems of adaptive immunity are generated during the development of T and B cell populations by somatic recombination. This process leads to the generation of very large and extremely diverse cell populations, which varies from individual to individual (Kimbrell and Beutler, 2001), (Flajnik and Du Pasquier, 2004).

To emphasize the distinction between these two mechanisms, I would like to pose a question: “What were the evolutionary pressure that selected for the development of these two sets of receptors and the two distinct recognition mechanisms they employ?” (Janeway, 1989) The answer lies with the terms for coadaptation of host and pathogen. If your enemy has a high capacity to change its strategy (for example, to escape host immune response, the African trypanosomes regularly changes their coat (antigenic variation) (Borst, 2002) (Berriman et al., 2005) (Charles A. Janeway Jr., 2005)), you need to have such system to be ready for the new approach followed by your enemy. Host must have enough genetic variation so that the species can change as fast as the pathogen. As a rule for adaptation “if the genetic variation is inadequate, the species will become extinct (Lewontin, 1978).” It is impossible to code for such a wide variety of receptors genetically. Additonally, as mentioned earlier, the genes encoding receptors for adaptive immunity are assembled during the

development of T and B cells. The enormous amount of the varaible regions of these receptors could potentially recognize many very different molecules or proteins (antigen) which are usually specific to the pathogens. The adaptive immune response very specific when compared to innate immunity which is known as non-specifically acting immune mechanism. Therefore, to generate wide variety of receptor repertoire by somatic recombination must be a big advantage for the host to fight against pathogens, which usually have high evolving capacity.

A recent study shows that the adaptive system arose at two time points during the course of evolution (Beutler, 2005), (Pancer et al., 2004). Immunoglobulins (IGs) are the effector molecules of adaptive immune system. They occur either as membrane-bound cell surface receptors or as free antibodies. T cell (TCR) and B cell (BCR) receptors are generated using the IGs during the development of lymphocytes. However, receptor components of innate immunity are composed of leucine rich repeats (LRRs) which are germline encoded. It was shown by Pancer et al., that like the IGs, LRRs are used to generate variable receptors by somatic recombination in lamprey fish. This clearly shows that receptor components for innate immunity can be also used for adaptive immunity and it suggests that during the host pathogen coevolutionary process lamprey fish used another evolutionary trajectory to generate its adaptive immunity to get high level information.

Innate immunity is present in all higher eukaryotes (Medzhitov and Janeway, 2000). Clearly, invertebrates and plants can survive without any adaptive immune mechanism (Hoffmann et al., 1999) and hence, innate immunity might be the most important immune system acting against pathogens in a wide range than adaptive immunity. One of the very well known families of innate immune receptors, which exist in vertebrates and invertebrates, are Toll like receptors (TLR). The Toll receptor was first discovered in a screen for dorso-ventral patterning in Drosophila (Anderson and Nusslein-Volhard, 1984). It took more than ten years to find out that in the adult fly the toll receptors have immune function especially to fungal infections (Lemaitre et al., 1996). Subsequently mammalian homolog, TLR was shown to be involved in immunity (Medzhitov et al., 1997). The analyses of genetic and physical mapping of LPS locus in C3H/HeJ and C57BL/10ScCr mice led to the discovery of TLR4 (Poltorak et al., 1998). Moreover, mice with a targeted deletion for the TLR4 gene were unresponsive to LPS (Hoshino et al., 1999) and it was shown that TLR2 and TLR4 play differential roles in the recognition of gram positive and gram-negative bacteria (Takeuchi et al., 1999). All members of the Toll family are single membrane-spanning proteins and their extracellular domains are composed of leucine rich repeats (LRR) which recognizes pathogen associated

molecular patterns (PAMPs) such as LPS, flagellin. So far, 23 members of the Toll family have been described both in vertebrates and invertebrates (Roach et al., 2005).

The plant recognition systems use similar receptors. The plant recognition receptors are classified as nucleotide binding receptors and leucine rich repeats (NBS-LRRs) which show significant similarity to NOD receptors in mammals (Ausubel, 2005). For example, FLS2 corresponds functionally to TLR5 in mammals which is flagellin receptors but FLS2 and TLR5 recognize different epitopes in flagellin protein. Similarly, RPM1, RPP5 (receptor proteins from Arabidopsis) are known to take part in immunity and have LRR containing domains (Staskawicz et al., 2001). However, the innate immune system in plants is more specialized than mammalia. As in the case of rice *Xa21* gene, a transmembrane protein containing extracellular LRRs recognize species specific secreted molecule from Xanthomonas oryzae rather than broadly conserved PAMPs (Ausubel, 2005) (Fritig et al., 1998) (Kimbrell and Beutler, 2001) (Kayihan et al., 2005). Innate immune system is the first system encounter for the pathogen (within first second to hours of entry) and is considered to be responsible for the induction of adaptive immunity (Medzhitov and Janeway, 1999) which takes several days longer.

Cell autonomous immunity, an effector mechanism in innate immunity is a newly introduced term, describing the ability of individual cells (including non-immune cells) to destroy intracellular pathogens in a cell autonomous manner. In the first instance, the pathogens are recognized by the PRRs. Recognition mobilizes specific destruction systems by the activation of several signaling pathways within cells. A variety of molecules protect cells in different ways. The mechanisms for cell autonomous regulation has been described for PKR (Tanaka and Samuel, 1994), 2'-5' Oligoadenylate synthesize (Mashimo et al., 2003), Mx (Schwemmle et al., 1995), IDO (Pfefferkorn, 1984), iNOS (MacMicking et al., 1995), LRG47 (Collazo et al., 2001), gp91-phox (Nathan et al., 1983) which act intracellularly. These genes will hence be referred to as effector genes (EGs) see fig 2 and 3.

These sets of EGs are either inducible by direct signaling events within the cell or by cytokines which activate signaling events in almost any type of cells in the host resulting in induction of EGs.

Cytokines are proteins secreted by cells upon infection or tissue damage. Interferons are a class of cytokines responsible for managing host defense against pathogens by activating cells upon infection (Fig 3). They can be classified into three kinds, type I IFNs (ifn-α,β,ω,τ), type II IFNs (ifn-γ) and recently identified type III IFNs (ifn-λ) (Stark et al., 1998), (Kotenko et al., 2003), (Boehm et al., 1997), (David, 2002). While IFN γ is mainly secreted by natural

killer and activated T cells (Th1 and Tc1), the interferon receptors are expressed in nearly all types of cells. Differential screens and expression analysis indicate (Boehm et al., 1997) that thereby regulating more than 800 genes constituting a specific and complex defense protecting cells (Boehm et al., 1997) (Dar et al., 2005) .

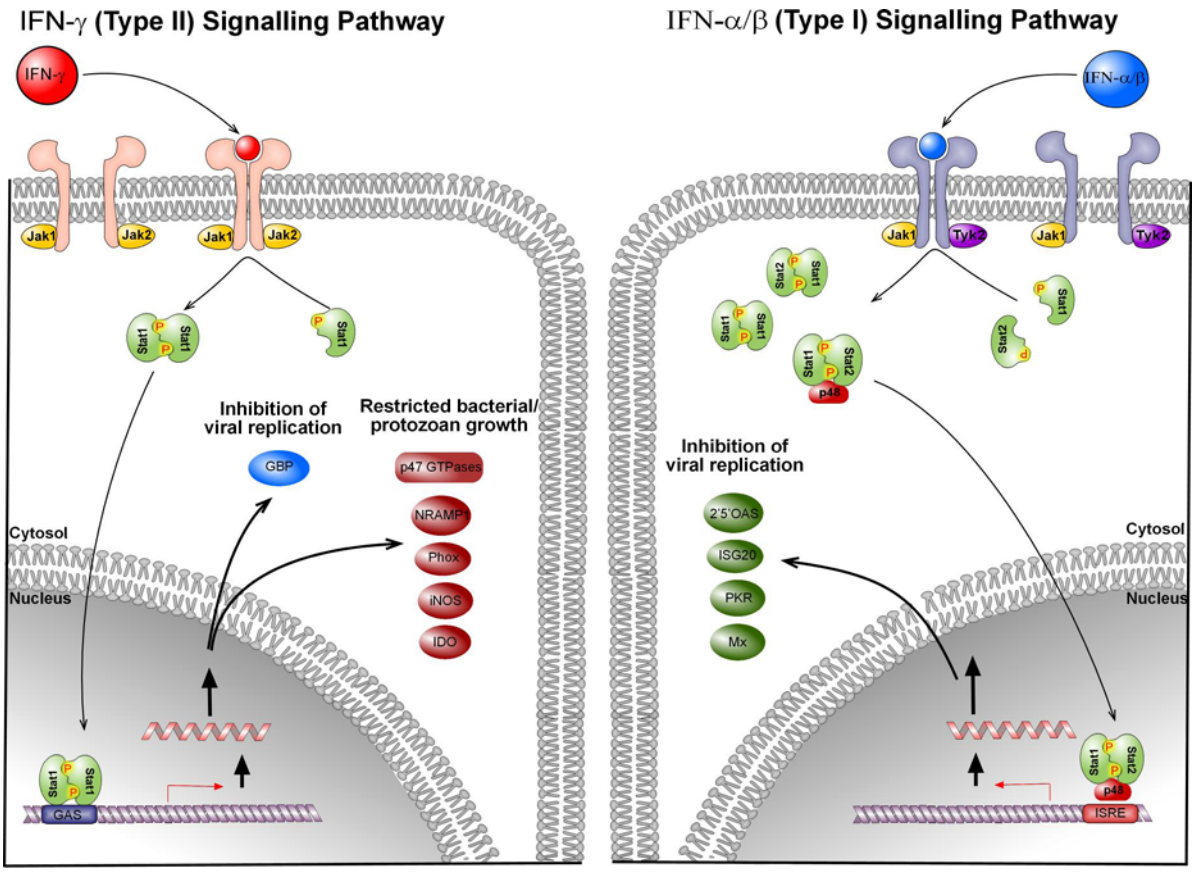


Figure 3. Simplified scheme for the type I and type II interferon signal transduction pathways.
Cytokines bind to their respective receptors and trigger the signalling pathways via phosphorylation of Stat1, Stat2 by Jak1, Jak2, and Tyk2. Activated Stat1 and Stat 2 homodimerize (Type II signaling) or hetoridimerize (Type I signaling) and interact with p48 (IRF9) to form ISGF3 complex. Activation of subsequent genes occurs via binding of Stat1 homodimer or the ISGF3 complex to GAS and ISRE sequences, respectively. IFNs induce set of genes or gene families, that are involved in inhibiting intra- and/or extracellular propagation of virus, bacteria and protozoa. Modified after (Taylor, 2004) and (Stark et al., 1998).

I.3.Interferon Inducible GTPases

Among the plethora of interferon-inducible genes, the importance of the GTPases will be emphasized because of their abundance as well as their functions. These include the Mx family of GTPases (Lindenmann et al., 1963), Guanylate binding protein (GBP) family (Cheng et al., 1985), very large inducible GTPase (VLIG) (Klamp et al., 2003) and the p47 GTPase family (Boehm et al., 1998). These GTPases have similar biochemical characteristics and functions to the dynamin family of GTPases. These proteins are characterized by their ability to oligomerize and can display oligomerization-dependent stimulation of GTP

hydrolysis (Warnock et al., 1996). Thus, here they will be grouped as dynamin like GTPases (Praefcke and McMahon, 2004). Though, their phylogenetic relationship is not resolved.

I.4.Dynamin Family of GTPases

Dynamins are GTPases with a molecular weight of about 100 kDa having an N-terminal GTP binding domain, a middle coiled coil domain, a pleckstrin homology (PH) domain involved in binding to phosphoinositides, a GTPase effector domain (GED) which is important for oligomerization, and a C-terminal proline rich domain (PRD) that interacts with SH3 domain containing proteins (Praefcke and McMahon, 2004), (Vestal, 2005), (Song and Schmid, 2003).

Dynamin and dynamin-like GTPases are involved in many processes in the cell. Dynamin1 plays a major role in the endocytic pathway by the scission of clathrin-coated vesicles from the plasma membrane. They are generally classified as large GTPases since they differ in size and function from the small GTPases like the Ras superfamily (p21 ras, an oncogene, which is very well characterized with respect to its GTPase properties and function). However, the mechanism of action of dynamin has not been resolved. Dynamin functions either as mechanochemical enzyme or regulatory enzyme or both; dynamin behaves as mechanochemical enzyme using the energy of GTP hydrolysis to sever vesicles. It differs from the other regulatory GTPases such as ras which upon GTP binding, interacts with several effector molecules, thereby inducing and performing their respective functions in the cell. Dynamin uses the PH domain to bind to membranes, and through the SH3 domains, binds to several effector molecules essential for its endocytic function (Song and Schmid, 2003).

The members of the dynamin family are found in prokaryotes and eukaryotes. Its function diversifies within the cell from cell division to vesicle scission. Drosophila dynamin was the first dynamin to be described, recognized via a temperature sensitive mutant in a locus called “shibire”. Since then many members of the family with similar characteristics have been discovered.

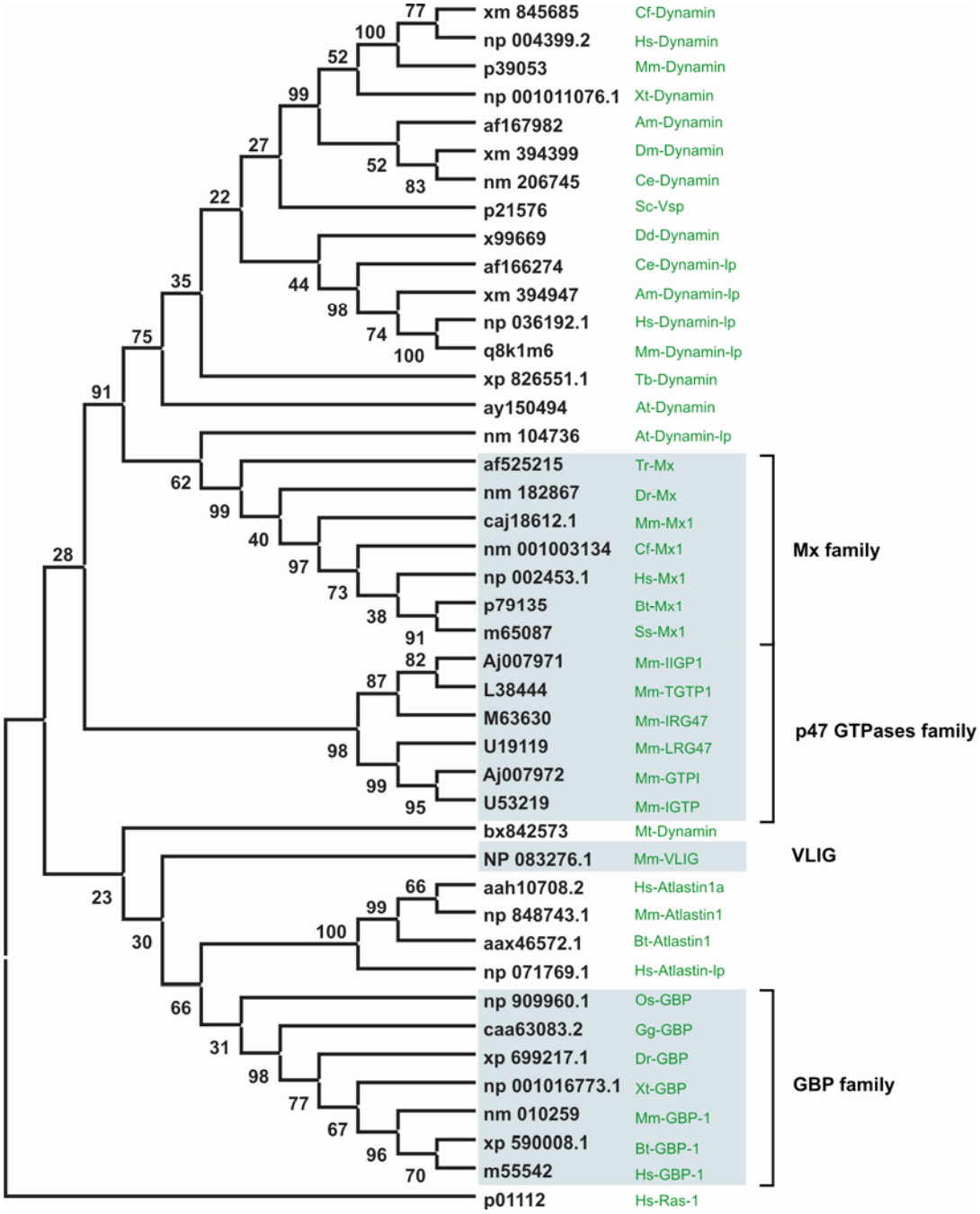


Figure 4. Phylogeny of dynamin and dynamin-like GTPases
Maximum Parsimony tree based on the G-domain of selected dynamin related proteins generated using clustal-X 1.83 (Matrix blosum) for multiple alignment, Mega3.1 for phylogenetic tree construction and bootstrap test. G-domain is defined according to Hs-Ras-1. Bootstrap values were indicated in black on the branch point. Black and green colored labels indicate the accession numbers and name of the gene, respectively. The blue highlighted genes are found to be inducible by interferon. The species names are abbreviated as Hs *(Homo sapiens)*, Mm *(Mus musculus)*, Bt *(Bos taurus)*, Cf *(Canis familiaris)*, Ss *(Sus scrofa)*, Gg *(Gallus gallus)*, Dr *(Danio rerio)*, Xt *(Xenopus tropicalis)*, Ce *(Caenorhabditis elegans)*, Am *(Apis mellifera)*, Dm *(Drosophila melonogaster)*, Mt *(Mycobacterium tuberculosis)*, Tr *(Takifugu rubripes)*, Dd *(Dictyostelium discoideum)*, Sc *(Saccharomyces cerevisiae)*, Tb *(Trypanosoma brucei)*, Ec *(Escherichia coli)*, At *(Arabidopsis taliana)*, Os *(Oryza sativa).*

All P-loop GTPases are classified into two main groups: TRAFAC-(GTPase similar to translation factors) and SIMIBI-(GTPase similar to signal recognition particle) according to their relationship with translation and signal transduction respectively (Leipe et al., 2002).

The dynamin family belongs to the TRAFAC family of P-Loop GTPases according to the Leipe classification (Leipe et al., 2002). They also showed that the dynamin like subfamily of P-loop GTPases is represented early in the eukaryotic branch. The branch reaches to the LUCA (Last Universal Common Ancestor of the extant life forms) suggesting that they emerged at the beginning of eukaryotic evolution. Based on the similarity in the mechanism of action (see below), the origin of the dynamins can be linked to the septin family of proteins, important for cell division (Field et al., 1996), (van der Bliek, 1999). Septins are necessary for cytokinesis in budding yeast and drosophila. They have an N-terminal GTP binding domain and C-terminal domain which show similar functions to the dynamins (Field et al., 1996).

The dynamin and dynamin like GTPases are also found in plants and invertebrates fig 4 (Praefcke and McMahon, 2004). At least four members of this family are massively inducible by interferons: The guanylate binding protein (GBP), Mx, VLIG and the p47 (IRG) family of GTPases (light blue highlighted in phylogeny Fig 4.). Since the inducible dynamin-like GTPases are major players in cell autonomous immunity, there is no reason to believe that these proteins should not be present in invertebrates and one expects that these genes act in the same way they act in mammals. In fact, we know that members of the GBP family are present in invertebrates, and some of the representatives of Mx GTPases have also been found in plants (Hong et al., 2003) (Dombrowski and Raikhel, 1995). Interestingly, for the p47 GTPases, no homologs have been found in invertebrates and plants (see discussion). It is most likely that genes evolved with the mechanism of immune response, under different selection pressures (coevolution) leading to their disappearance from some of the main branches of the eukaryotes.

GBPs are induced by type I and type II interferons. The first GBP members cloned were HuGBP-1 and HuGBP-2 (Cheng et al., 1991). This family now comprises five members described in human and mouse. GBP-1 has a mass of about 67-kDa and has a unique property of binding to GMP, in addition to GTP and GDP (Cheng et al., 1985). hGBP1 has the canonical GTP binding motifs important for coordinating the binding of guanine nucleotides except for G4 motif, which is different from other GTPases (Praefcke et al., 1999). Biochemically, it has an ability to oligomerize upon binding to GTP (dimer), GDP-AlFx (tetramer) and shows at least eight fold increase in GTP hydrolysis upon multimerization (Prakash et al., 2000) (Praefcke et al., 1999). Recent analysis showed that hGBP-1 can target specifically to Golgi membrane in its GDP-AlFx bound form (Modiano et al., 2005). Although GBPs are massively induced by interferons, their function as resistance factors has

not been established. However, it has been reported that hGBP-1 shows an inhibitory effect (40-60%) on EMCV and VSV replication in cultured HeLa cells (Anderson et al., 1999). However, 35 out of 46 different mouse strains showed inability to express murine GBP-1 upon induction by type I or type II interferon, and no viral susceptibility was observed between the expressing and non-expressing strains. Since the GBP family contains 5 members, no viral susceptibility can be linked to redundant function of the individual GBP proteins (Staeheli et al., 1984) (Vestal, 2005). Additionally, it is reported that hGBP-1 has growth inhibitory effect on endothelial cells and alter the adhesive invasive properties of the cells (Guenzi et al., 2001).

The Mx family of resistance GTPases, especially human MxA, has been shown to be involved in resistance against a wide variety of viruses such as bunyaviruses, orthomyxoviruses, paramyxoviruses, rhabdoviruses, togaviruses, picornaviruses and hepatitis B virus (Gordien et al., 2001; Haller and Kochs, 2002; Janzen et al., 2000; Kochs and Haller, 1999). Mx proteins are mainly induced by type I interferon. The mouse Mx1 gene encodes interferon inducible nuclear protein. As mentioned above, only two lab mouse strains (A2G and SL/NiA) carry the $Mx1^{+}$ allele and are resistance to the influenza virus. Whereas all the other laboratory mouse strains carry $Mx1^{-}$ allele and susceptible to influenza virus (Staeheli et al., 1988) (Jin et al., 1998a). In contrast to laboratory strains, it was shown by Haller et al., that wild mice carry the both alleles at equal frequencies (Haller et al., 1987). This suggests that Mx proteins are under the control of balancing selection which can possibly be explained by general fitness cost for resistance genes (Rigby et al., 2002). Like GBP-1, Mx proteins also contain an N-terminal GTP binding domain, middle domain and C-terminal domain which has GED activity. It has low affinity to GTP when compared to ras like GTPases and high rate of GTP hydrolysis following general characteristics of dynamin-like GTPases (Haller and Kochs, 2002; Schumacher and Staeheli, 1998). It has also been shown that human Mx1 and MxA protein can form higher oligomeric structures (Melen et al., 1992) (Kochs et al., 2002a). However, for the antiviral activity of MxA, formation of large oligomeric structures is not necessary (Janzen et al., 2000). Recent studies show that MxA specifically recognizes and sequesters the LaCrosse viral (LACV) N protein into large perinuclear complexes and oligomeric MxA/N complexes are formed in close association with COP-I-positive vesicular-tubular membranes (Kochs et al., 2002b),(Reichelt et al., 2004).

I.5.The Family of p47 GTPases

The p47 GTPases are a family of GTPases which is massively induced by interferon gamma (Boehm et al., 1998). The proteins have an N-terminal region, GTP binding domain and highly variable C-terminal region, which might be important for intracellular localization and oligomerization (see discussion). The GTP binding domain of the p47 GTPases has all three classical GTP binding motifs (Fig 5). Apart from G domain, p47 GTPases have no homology to other GTPases. Both N and C-terminal region have characteristic features which distinguish this family from other P-loop GTPases (Fig 4, 5 and see below). IRG 47, isolated as a cDNA from B cells, was the first member to be described (Gilly and Wall, 1992). So far, six of the members have been studied in some detail (Boehm et al., 1998) (Taylor, 2004). At least four of them have been analyzed functionally by targeted gene knockout experiments (Table 1) (Taylor et al., 2000) (Collazo et al., 2001) (Parvanova, 2005) (Taylor, 2004). Targeted gene knock-out experiments revealed that the p47 GTPases family is indeed involved in resistance against wide variety of pathogens in a non-redundant way (see Table 1) (Taylor, 2004). The phenotype of the knock out mice for LRG47 and IGTP were very striking, showing early death upon infection by *Toxoplasma gondii* (within first 10 days p.i.) suggesting that these p47 resistance proteins may be the strongest resistance system in mouse. They can be grouped into two structural subfamilies, named GMS and GKS, based on a remarkable substitution in the G1 motif (Fig 4 and 5). The GMS proteins LRG47, IGTP and GTPI (GMS subfamily) carry methionine instead of lysine in their G1 motif. This substitution is a unique feature of the p47 GTPases family. All the P-Loop GTPases have the canonical lysine important for the coordination of the phosphates in the nucleotide. The GMS subgroup of p47 GTPases also contains 12 additional specific amino acid substitutions in their G-domain relative to the members of the GKS subfamily; IIGP1, TGTP1, IRG47 (Fig 5). Biochemical analysis of recombinant IIGP1 shows low affinity to GTP, slow rate of GTP hydrolysis, co-operative GTP hydrolysis with ability to form oligomers in a GTP dependent manner (Uthaiah et al., 2003). The crystal structure of IIGP1 (Fig 6) (Ghosh et al., 2004) shows three N-terminal α-helices followed by a G-domain, which is structurally similar to GTPase domain of Ras. The G-domain is linked to the C-terminal domain by a short linker helix (αE) and the C-terminus contains seven α helices. Based on homology within the family and analysis by secondary structure prediction programs, we can clearly say that IIGP1 is likely to be structurally representative of all p47 GTPases. Granted, the similarity in the biochemical characteristics, and sequence analysis, the p47 GTPase family can be grouped into the dynamin like GTPases. The p47 GTPases are emerging as important cell autonomous

resistance molecules. LRG-47 deficient mice (LRG-47$^{-/-}$) have increased susceptibility to *M. tuberculosis*. Moreover, Macropahges isolated from LRG-47$^{-/-}$ mice showed arrested maturation of phagosomes containing *M. tuberculosis* (MacMicking et al., 2003). Recently, Martens et al., reported that astrocytes isolated from IIGP1 deficient mice have increased susceptiblity to *T. gondii* (Martens S, 2005). In the resting level, LRG-47 localizes to the Golgi apparatus and is recruited to the plasma membrane upon phagocytosis whereas IIGP1 is an endoplasmic reticulum associated protein in fibroblast, hepatocyte and macrophages (Martens et al., 2004). Recent analysis showed that upon infection by *T. gondii* TGTP, IIGP1, IRG47, GTPI and IGTP are accumulated on the parasitophous vacuole (Martens S, 2005).

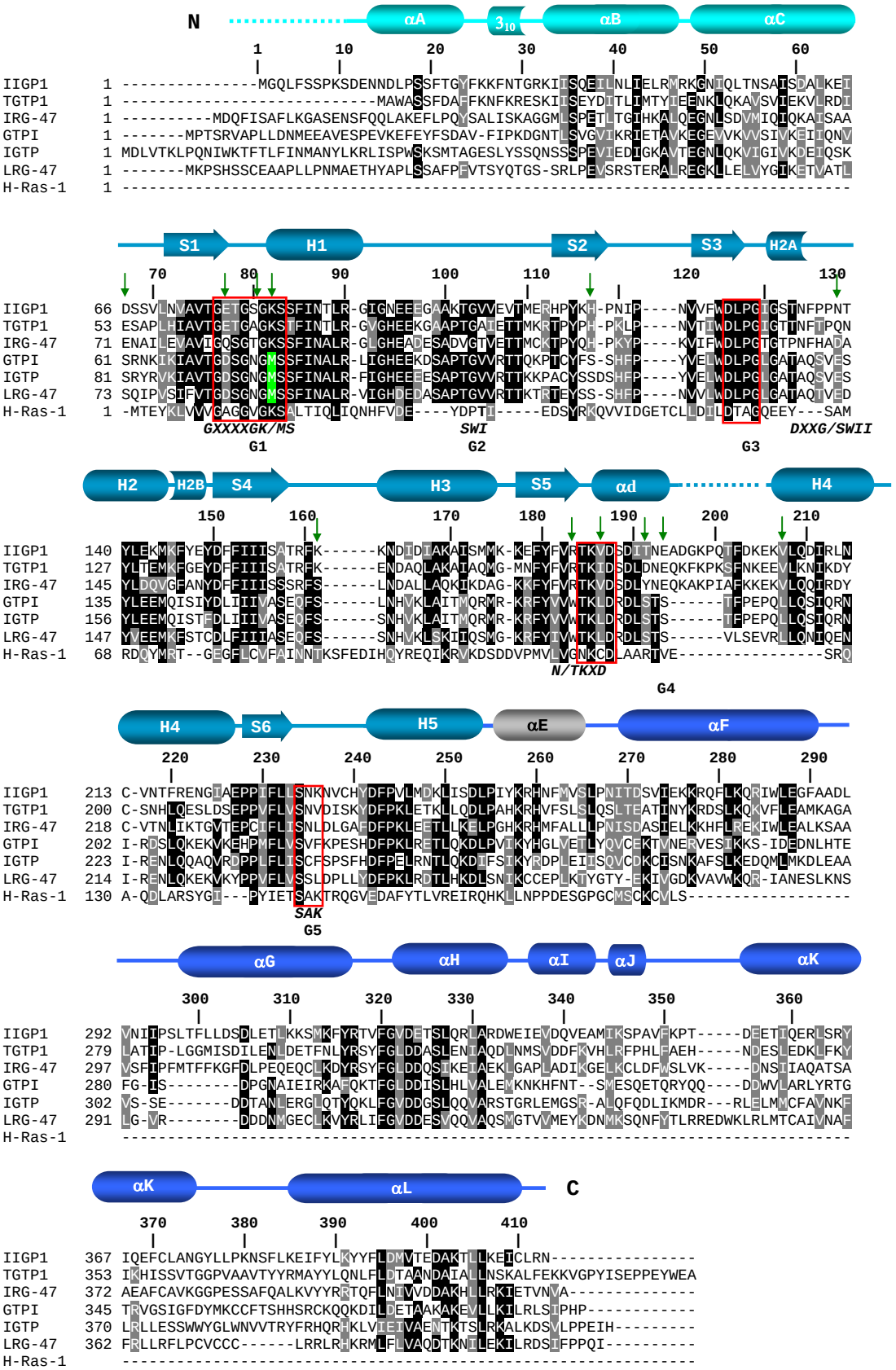


Figure 5. Sequence Alignment of identified p47 GTPases.
Sequences of 6 mouse p47 GTPases IIGP1 (AJ007971), TGTP1 (L38444), IRG47 (M63630), LRG47 (U19119), GTPI (AJ007972), IGTP (U53219) H-Ras-1 (P01112) showing close homology extending to the C-terminus, aligned on the known secondary structures of IIGP1 (Ghosh et al., 2004). The unusual methionine residues in the G1 motif of GMS proteins are highlighted in green and GMS specific a.a. substitutions are indicated with green arrow. Canonical GTPases motifs are indicated in red boxes.

Mouse	Intracellular protozoa			Intracellular bacteria				Virus
	T. gondii	*L. major*	*T. cruzi*	*L. monocytogenes*	*S. typhimurium*	*M. tuberculosis*	*M. avium*	*MCMV*
Wild type	R	R	R	R	R	R	R	R
IFN-γ knockout	S(acute)	S	S	S	S	S	S	S
LRG47 knockout	S(acute)	S	N.D.	S	S	S	S	R
IGTP knockout	S(acute)	S	R	R	R	R	R	R
IRG47 knockout	S(chronic)	N.D.	N.D.	R	R	R	N.D	R
IIGP1 knockout	R*	R	N.D.	R	N.D	N.D	N.D	N.D

Table 1. Summary of phenotype observed to different intracellular pathogens in mice lacking p47 GTPases.
S and R indicate susceptible and resistance respectively. N.D.: not determined. * The susceptibility effect was only observed in cultured cell lines (astrocytes) (Martens S, 2005) and (Parvanova, 2005). Modified after (Taylor, 2004)

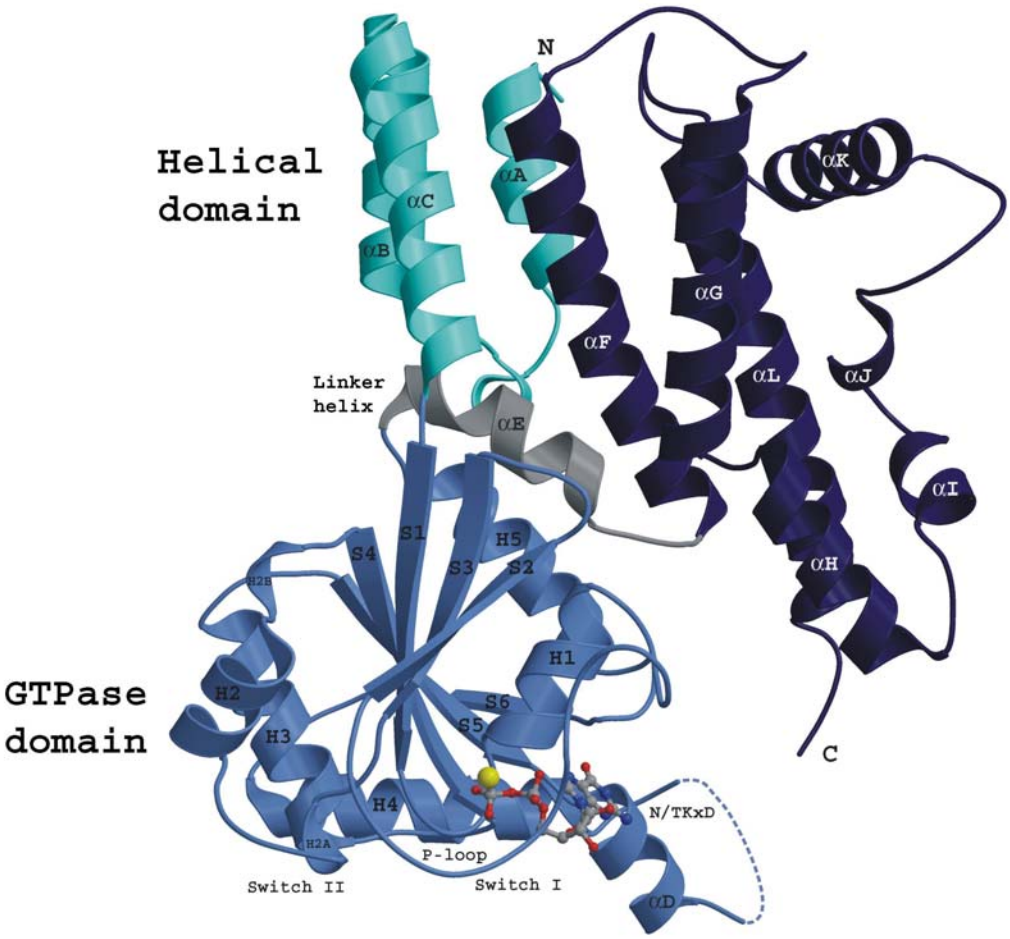


Figure 6. Crystal structure of IIGP1 in GDP bound form shown by ribbon presentation (Ghosh et al., 2004)
IIGP1 contains three domains, The N-terminal domain (cyan), G-domain (light blue) and C-terminal domain. The GTPase domain shows very similar features to the G-domain of H-Ras-1.

I.6.The Aim of This Study

Host-pathogen interactions generate powerful evolutionary forces. Therefore, genes or gene families related with immunity are known to be fast evolving. Involvement of interferon inducible large GTPases in immunity has been described. p47 GTPases, described above, is one of the interferon inducible large GTPases family thought to be involved in providing cell autonomous immunity in mouse. Detailed analysis of six of the family members revealed that p47 GTPases are indeed one of the most important resistance mechanisms of the mouse against variety of vacuolar pathogens. Having such resistance mechanism must be a big advantage for an organism. However, there are no reports of p47 GTPases in man. Hence, the importance of p47 GTPases as a resistance mechanism in mouse is a critical theme to be analyzed. The analysis of six p47 GTPases working in a non-redundant way as a family led us to analyze the evolution of p47 GTPases. Elucidation of the functional relationship between the species especially for mouse seemed imperative.

Detailed analysis of the whole p47 GTPases family was carried out both phylogenetically and experimentally. The conclusions reached were unexpected, to be presented and discussed in detail in the following sections.

II.MATERIALS AND METHODS

II.1.CHEMICALS, REAGENTS AND ACCESSORIES

All chemicals were purchased from Aldrich (Steinheim), Amersham-Pharmacia (Freiburg), Applichem (Darmstadt), Baker (Deventer, Netherlands), Boehringer Mannheim (Mannheim), Fluka (Neu-Ulm), GERBU (Gaiberg), Merck (Darmstadt), Pharma-Waldhof (Düsseldorf), Qiagen (Hilden), Riedel de Haen (Seelze), Roth (Karlsruhe), Serva (Heidelberg), Sigma-Aldrich (Deisenhofen). DNA size standards from Gibco-BRL (Eggenstein), electrophoresis chambers from FMC Bioproducts (Rockland Maine US), developing and fixing solutions for Western Blot detection from Amersham Pharmacia (Freiburg), Luminol from Sigma Aldrich (Deisenhofen), Coumaric acid from Fluka (Neu-Ulm). Deionised and sterile water (Seral TM) was used for all the buffers and solutions, Ultra pure water from Milli-Q-Synthesis (Millipore).

II.1.1.Enzymes/Proteins

Restriction Enzymes and T4 DNA polymerase from New England Biolabs (Bad Schwalbach)
"Complete Mini" protease inhibitor cocktail from Boehringer (Ingelheim).
Pyrococcus furiosus (Pfu) DNA polymerase from Promega (Mannheim)
Shrimp Alkaline Phosphatase (SAP) from Amersham
Thrombin from Serva, (Heidelberg)
RNase A from Sigma
1Kb ladder for Agarose gels from Gibco
Rainbow –Molecular weight marker-Precision protein standardsTM (Biorad)
Page Ruler Protein Marker from Fermentas
Wide Range Protein Marker from Sigma

II.1.2.Reagent Kits

Plasmid Mini and Midi kit from Qiagen
Sequencing Kit from ABI PRISM
Total RNA and mRNA isolation kit from Qiagen

II.1.3.Vectors

PGW1H from British Biotech (Oxford, England)

pGEX-4T-2 from Amersham Pharmacia (Freiburg)
pMALp2E from New England Biolabs (Bad Schwalbach)
pBlueScript II KS+ from stratagene
pGEMTeasy from Promega
pET28b+ from Novagen
pRSET (A,B,C) from invitrogen

II.1.4.Materials for Protein Isolation

Ni-NTA Superflow from Qiagen
Amylose Resin from New England Biolabs
GST beads from Amersham

II.1.5.Media

Luria Bertini (LB) Medium
10 g Bacto Tryptone, 5 g Yeast Extract, 10 g Nacl, Distilled water to 1Litre
LB Plate Medium
10 g Bacto Tryptone, 5 g Yeast Extract, 10 g Nacl, 15 g Bacto Agar, Distilled water to 1Litre
Terrific Broth (TB) Medium
12 g Bacto Tryptone, 24 g Yeast Extract, 0.17 mM KH_2PO_4, 0.072 mM K_2HPO_4, 4 ml Glycerol, Distilled water 1 Litre
IMDM (Iscove's Modified Dulbecco's Medium) from Gibco
10% FCS, 2 mM 1-Glutamine, 1 mM Sodium pyruvate, 100 U/ml Penicillin,100 µg/ml Streptomycin, 1x non-essential amino acids. Media mainly used for the growth of L929, T2 IRF9-/-, MEF and Hela cells.
DMEM (Dulbeco's Modified Eagle Medium) from Gibco.
10% FCS, 2 mM 1-Glutamine, 1 mM Sodium pyruvate, 100 U/ml Penicillin,100 µg/ml Streptomycin, 1x non-essential amino acids. Media mainly used for the growth of Hek293, HepG2, MCF-7, SW480, and Primary Foreskin Fibroblast (HS27) cells.
RPMI 1640 + L-Glutamine from Gibco.
10% FCS, 100 U/ml Penicillin,100 µg/ml Streptomycin, 1x non-essential amino acids.
Media mainly used for the growth of Thp1, primary foreskin fibroblast (HS27), and IRF 8 -/- (50uM 2ME, 6ng/ml GM-CSF, 6ng/ml M-CSF additionally required) cells.

II.1.6.Antibiotics

Ampicillin from Roth was prepared as a stock solution of 100 mg/ml in water, used as final concentration of 100 μg/ml and stored at 4°C. Kanamycin from Sigma stock solution was prepared as 30 mg/ml in water, used as final concentration of 30 μg/ml and stored at -20°C. Chloromphenicol from Sigma was prepared as 30 mg/ml in EtOH, used as final concentration of 30 μg/ml and stored at -20°C. Penicillin/Streptomycin from Gibco

II.1.7.Bacterial Strains

E. coli XL1-Blue: recA1, end A1, gyrA96, thi-1, hsdR17, supE44, relA1, lac,
[F', *pro AB*, *lacI*^{q}ZΔ*M15*, Tn*10* (Tet^{r})]
E. coli DH5α: *80dlacZ* ΔM15, recA1, endA1, gyrA96, thi-1, hsdR17 (r_B^- m_B^+), supE44, relA1, deoR, Δ*(lacZYA-argF)*U169
E. coli BL-21: *E. coli* B, F-, omp T, hsd S (r_B^- m_B^-), gal, dcm
E. coli NB42: (Cicchetti et al., 1999) kindly provided by Ralf Max Leonhardt

II.1.8.Eukaryotic Cell Lines

Hela (Human cervix Carcinoma)

MEFs (Mouse embryonic fibroblasts)

L929 (Mouse fibroblast cell line)

Thp1 (Human monocytic Leukumia)

GS293 (Human embryonic kidney)

HepG2 (Human primary liver cancer)

MCF7 (Human breast adenocarcinoma)

T2 (Lymphoblastoma cell line)

Hs27 (primary foreskin fibroblast)

IRF8 -/- (CL2 cells (Macrophage like cell lines))

IRF3 -/- (MEFs)

IRF9 -/- (MEFs)

MDCK II (Madin-Darby canine kidney cells)

II.1.9.Antibodies

<u>Primary antibodies and antisera</u>

α4181, human IRGM recombinant protein, Rabbit polyclonal antibody, dilution 1:5000 for IB (Immunoblot), 1:500 for IF (immunoflourescence), generated in this study (see below)

Ctag-1 (Natasa Papic personal communication), peptide (KLGRLERPHRD), Rabit polyclonal antibody, dilution 1:5000 for IF and IB, from Eurogentec. αIGTP, mouse IGTP (283-423), mouse mono clonal antibody in a concentration 0,25 μg/ml, dilution 1:250 for IF from BD Transduction Laboratories. A19, mouse LRG47 (N-terminal), goat polyclonal, IF dilution 1-100 from Santa Cruz.

II.1.10.Secondary Antibodies and Antisera

For westernblot;
IgG anti-Mouse Horseradish peroxidase coupled from Goat (Pierce)
IgG anti-Rabbit Horseradish peroxidase coupled from Donkey (Amersham)
For immunofluorescence;
Goat anti-mouse 1-2000 Alexa 488/546 from Molecular Probes, Donkey anti rabbit 1-2000 Alexa 488/546 from Molecular probes and DAPI (Sigma)

II.2.MOLECULAR BIOLOGY

II.2.1.Culture of Eukaryotic Cells

Cell lines (see above) were grown in 75cm^2 polystrene tissue culture flasks (Sarstedt) with 5% CO_2 at 37°C in a humidified incubater with suitable media appropriate for each cell line (see above). When the cells reached 80% confluency medium was removed and the cells were washed once in 1XPBS then detached from the plastic by trypsinisation (1X Trypsin). To prepare frozen stocks, cells were resuspened in freezing medium (FCS with 10 % di-methyle sulfoxide (DMSO)) in a final cell number (>10^6 cells/ml) then cells were kept over night at -20°C and next morning transferred to -80°C for longer storage cells transferred liquid nitrogen. Transient transfection was performed in Hela, MEFs, L929 and Hek293 cells. Cells were grown up to 80 % confluence in 60mm dishes and transfection with Fugene (FuGENE^{TM} Roche applied sciences) was performed according to manufacturer conditions; 6 μl of fugene mixed with 90 μl of serum free medium, appropriate amount of DNA was added (minimum 1μg of DNA) mixture incubated at room temperature for 15min- 45min and added to the cells in drop wise manner.

II.2.2.Preparation of IRGM(a) Specific Polyclonal Antisera (α4181)

A rabbit antiserum against IRGM(a) protein was prepared, The rabbit was immunized subcutaneously with 200 μg purified recombinant MBP-IRGM(a) fusion protein (Figure 21)

which was diluted in up to 500 µl PBS in equal amount of complete Freund's adjuvant (DifcoLab., Detroit, MI). 2nd injections with same protein were given subcutaneously after four weeks with 200 µg in same conditions. Two weeks later 2° bleed was collected and tested versus pre-immune serum by westernblot. 3rd injection was performed after 4 months in same conditions but the preparation of recombinant IRGM(a) protein was different; The MBP-IRGM(a) fusion protein was digested over night with thrombin (see above). The digested protein was subjected to Gel-filtration column. The IRGM(a) containing fractions were collected, concentrated with Vivaspin (Vivascience) centrifugal concentrator with 10000 MW cut off and redisolved with 500 µl of resuspension buffer (6M GnCl, 50mM Hepes, 4mM DTT, pH:7.5). Resuspended IRGM(a) recombinant protein was subjected to Gel-filtration coloumn. The fractions were collected and dialyzed against PBS in a volume ratio 1 to 500 over night. Dialyzed fractions were checked on the gel and stored at -80°C. 4th injection was performed after 4 months in the same conditions above. Amount of the protein were determined using the Bradford assay. Antisera from all the bleeds, the pre-bleed prior immunisation, first, second, third, and the fourth bleeds kept at room temperature for over night and next morning were obtained by centrifugation of the clotted blood at 3000g for 7 min at 4°C and stored at -20°C. Western blots were done on transfected and un-transfected cell lysates using different dilutions of the pre-bleed, and partially depleted 2nd, 3rd and 4th bleed to check the specificity for pre-bleed control.

II.2.3.Western Blot Analysis

Proteins were run on SDS-PAGE gel and transferred to nitrocellulose membrane by electroblotting. Ponceau-S (0.1% Ponceau-S (w/v) (Sigma), in 5% acetic acid) staining was used to define the place of the proteins on nitrocellulose membrane. Membrane was blocked with 5% milk powder, 0.1% Tween 20, for 15 hours at 4°C. Antisera/antibody was diluted in PBS, 10%FCS, 0.1 Tween20, and protein bands visualized using the enhanced chemiluminescence (ECL) substrate.

II.2.4.Immunofluorescence

Appropriate cell lines (see above) grown on 22X22 mm coverslips in 6 well plates were induced, left uninduced with interferon γ or transfected with GTPase constructs. After 24 hours medium was removed. The cells were washed with 2ml of PBS and fixed with 2 ml of PBS/3%Paraformaldehyde for 20 min at RT. Cells were washed 3times with PBS and washed

with 2 ml PBS/0.1% Saponin incubated for 10min at RT. Wash buffer was removed and immediately cells were blocked by adding PBS/0.1 %saponin/3% BSA and incubated for 1hour at room temperature in 6 well plates. Coverslips were incubated with 100 µl of PBS/0.1 %Saponin/3% BSA which contains appropriate antibody dilution (see above) on parafilm in humid environment for 1 hour at RT temperature or over night at 4°C. Coverslips were put in to the 6 well plates and washed with 3X 5 ml of PBS/0.1% Saponin. Coverslips incubated with 100 µl of PBS/0.1 %Saponin/3% BSA which contains appropriate secondary antibody dilution (see above) or dapi (1:1000) on parafilm in humid environment for 30 min at RT temperature in dark conditions. Coverslips were put into the 6 well plates and washed with 3X 5 ml of PBS/0.1% Saponin. Finally, coverslips were put on to the slide with 20 µl of ProLong Gold antifade reagent (Molecular Probes). After over night incubation cells were observed with a Zeiss Axioplan II fluorescence microscope equipped with a cooled CCD camera (Quantix) using the Metamorph software (version 4.5r3, Universal Imaging Corp.)

II.2.5.Oligonucleotides

Oligonucleotides were designed using the programs, primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi),Netprimer (http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html).

All oligonucleotides were from invitrogen, nucleotides were supplied as powder and resuspended to a final concentration 100pmol/µl then diluted 1 in 10 to 10pmol/µl as final concentration. For each reaction 1µl was used.

List of Primers

Gene	Primers 5' to 3'
Irgc	GCCTCTAGCTGCTGGGACCTGTCTCAGGTCACATCTGAG
	GCGGGTGGCCGCCAGATCCTCGTCCACC
*IRGC (*human*)*	GGAGATCCTATCAGTGGGGAGAGTGTGAGGG
	CCTCTCTGAAGCCCGACGGCCG
*GBP-1 (*human*)*	CTGTATCCGGAAATTCTTCCCAAAG
	CTTCAATGGCCTCTCTCTCACTGTC
*GAPDH (*human*)*	ATGACAACTTTGGTATCGTGGAAGG
	GAAATGAGCTTGACAAAGTGGTCGT
	ATATTTCTGGGCCTTGTGGAATTCAC
Irgb1-3-8	AAAGTTCTACTTTGTCCGAACCAAGATAGATCAAGAT
	CTCTTCCTTATTAAGGAGAGACTTGGCATCACTTG
Irgb2-5-9	GGTACATACAACCACTGAGAGAACACCATACACTTACA
	ATGGTATGGTAGCCCATGCTCTTGGCA
Irgb6	TCTACTTTGTCAGAACCAAGATAGACAGCGACTTAGA
	GCCATGCGATAGTAAGTGACTGCAGCG
Irgb7	TCATTATTGTCTCTGCTGGACGCATTAAACAT
	TTAGAGACTAAGAAGACTGGAGGCTCCTGGTG
Irgb8	CGCTTATCTAGACCAAGTGGGATTTGCCA

	CAACCACTATGTTAAGGAACTGTGTGCGCC
Irgm1	CGGAATCAAGGAGACTGTGGCAACATTG
	TCCTGGGCAACTAAGAAAAGCATGCGTT
Irgm2	GATCTCGGATCCGGGTAACGCGAT
	TAACAGAACTTCCTTGGCTTTGGCAGCAG
Irgm3	CTGGAGGCAGCTGTCAGCTCCGAG
	GTCCTTTAGAGCTTTCCTCAGGGAGGTCTTG
Irgb10	TGCTGCCTTGACAGACATTGAGAAAGCC
	GCTGCGTTAGCATTCTGCAGATTCTTTACAC
Irga1	TTCCCTTGTCAATGTGGCTGTCACTGG
	AGAAAGGTCAGTGAAGGTATGATAGTAGCACACCAG
Irga2	CACAGGTGGACTCTGACTTAAGAAATGAAGAGGATT
	ACTTTCTAAGAAGAAGCTCAGTAGCCCATCTGC
Irga3	AGCTATGCTTGAAAAGGGGGGACTTTCAG
	TATGGGCAAAATGAGTTTGAGCCGCTT
Irga4	GCTGAAGTTGGAGTAATAGAGACAACTATGAAGAGAACTTCT
	TGTGTCAAGGATATGAAGTTGTAAATAATAGATTGCAGG
Irga5	CTGACACTAGGAGATGTTCAGCAAGCAAATAATG
	AAGGAAAAGAAGTGGTAAGTAAAAGGCTTTCTCCATATA
Irga6	ATCAGTGATGCATTAAAAGAAATCGATAGTAGTGTGC
	GTCAGAGAAGGGATGATATTCACTAGGTCAGCAG
Irga7	CACAATTTTATGCTTTCTCTGCCTGGCATT
	TCAGCAAATGAGGGGACTTCATTATTTCTTTTACTT
Irga8	GAGTTATGCCTGAAGAAGGGGGACATTCA
	TGAGTTTGAGCTATTTTTTTGAATATGCCTCTTTAAGG
IRGC	GCTGGCAAGTCCTCCCTCATCAAC
	GAGAGGTTGGACACGAGGAAGATGC
IRGB2	TCCTTTCTCAGGAGGCCATCACTTC
	GCCAGTTGTGCATCATTGATTGTGA
IRGM5	GAAAAGGCATTGGGAGATGGGAAGT
	AACCTTTTCCCCTGTCTTTGGATGG
IRGM6	GAGAGAGCATCCAGTGTCCCATTGA
	GATGGGTTCGAAAACCCTCTCCTTC
IRGM4	ACCCAGTCCCTTCACACTCCATCAC
	TAGCAAGTGGGAATCTGGGTGGTTC
*IRGMs1-r1 (*human*)*	CAGGACACCAGTTAACATCACTATG
	GATTTTCCAGGACATTTTCTCTGAT
IRGM-rGMS *(*human*)*	ATATTTCTGGGCCTTGTGGAATTCAC
IRGM(b-e) f10-r60	GAGAAAGCCTCAGCAGATGGGAACTTG
	GCACTGGCTAGCTAGCTGTTGAATATCCTGA
IRGMmn	GGACTCTGGCAATGGGATGAACACCTTCATCAGTGCCCTTCG
	CGAAGGGCACTGATGAAGGTGTTCATCCCATTGCCACAGTCC
IRGM5'(HindIII)	CCCCCAAGCTTATGAATGTTGAGAAAGCCTCAGCAG
IRGMctag1(EcoR1)	CCCCCGAATTCTTAGTCACGATGCGGCCGCTCGAGTCGACCTAGTTTG TATTCACATACCCGCTCCTTCTGG
IRGMhistag(EcoR1)	CCCCCGAATTCTTAATGATGATGATGATGATGGTATTCACATACCCGC TCCTTCTGG
IRGM(b)3' (HindIII)	GCGCAAGCTTCTAGCTGTTGAATATCCTGAGCAGATTTAC
5' Anc	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGII
Est Stop	CCCCCAAGCTTCAGGATCCTTTCAGCAAGCAAGAGG
*IRGMr1 (*human*)*	GATTTTCCAGGACATTTTCTCTGAT
5'(1-2)	CCCCCGGATCCATGAATGTTGAGAAAGCCTCAGC
AP	GGCCACGCGTCGACTAGTAC$(T)_{17}$
AUAP	GGCCACGCGTCGACTAGTAC
UAP	$(CUG)_4$GGCCACGCGTCGACTAGTAC
*GAPDH (*mouse*)*	GTCTACATGTTCCAGTATGACTCCACTCACGG
	GTTGCTGTAGCCGTATTCATTGTCATACCAGG

II.2.6.Preperation of mRNA and cDNA synthesis

Oligotex mRNA isolation kit (Qiagen) was used to isolate mRNA from total RNA (isolated by using total RNA isolation kit Qiagen). mRNA was stored at -80°C. cDNA synthesized using the mRNA obtained by SuperScript First Strand Synthesis System (Invitrogen) according to manufacturers instructions. Synthesis was primed by Olig-dT primers. 1µl of cDNA is used for each reaction.

II.2.7.RT-PCR on Cells and Tissues

Mouse L929 fibroblasts or appropriate cell lines were stimulated for 24 h with 200U/ml IFN-γ or 200U/ml IFN-β (R&D and Calbiochem respectively). Human cell lines (Hela, HEK293, HepG2, T2, THP1, MCF-7, SW-480, Primary foreskin fibroblast-HS27) were stimulated for 24 h with 2000 U/ml Interferon-β or 200 U/ml Interferon-γ (PBL Biomedical laboratories and Peprotech respectively).Total RNA was extracted from tissues and cells using the "RNeasy mini kit" (QIAGEN, Hilden, Germany), except for testis, where the "RNeasy Lipid Tissue Kit" (QIAGEN) was used. Poly (A) RNA was isolated from total RNA using the Oligotex mRNA kit (QIAGEN). Total RNA from human tissues was purchased from Biochain (Hayward, CA, USA). cDNA was generated from mRNA and total RNA using the "Super Script First-Strand Synthesis System for RT-PCR" (Invitrogen, Carlsbad, CA, USA). The generated cDNAs were screened for the presence of p47 GTPase transcripts by PCR. The amplified fragments were confirmed by sequencing.

II.2.8.5'and 3' RACE (Rapid amplification of cDNA ends) PCR

5'RACE PCR; cDNA synthesized using the *IRGM*-rGMS primer (see above), after preparation of cDNA, cDNA made single stranded and purified using the rapid PCR purification kit (Boehringer). Terminal deoxy transferase Rection maintained on the purified cDNA, (16.5µl cDNA, 5µl TdT + Reaction buffer (Amersham), 2.5µl dCTP (2mM)) incubated for 3min at 94°C, 1µl of Tdt was added and incubated for 15min at 37°C, fallowed by inactivation step for 5 min 65°C. PCR reaction was performed on the (cDNA+polyC) using the primer *5'Anc*. PCR product was purified by using the rapid PCR purification kit and second round nested PCR was performed using the primers *UAP* and *IRGMr1*. 1.7 kb PCR product was cloned to PGEM-T easy and positive clones were determined by sequencing (1.6-5'Race-hGMS and 3.6-5'Race-hGMS).

3'RACE PCR; cDNA synthesized using the *AP* primer (see above), after preparation of cDNA, cDNA made single stranded and purified using the rapid PCR purification kit (Boehringer). First PCR was performed using the primers Hgms5'(1-2) and UAP. PCR product was prufied by using the rapid PCR purification kit and second round nested PCR was performed using the primers *IRGMr1* and *AUAP*. The PCR product was double digested with HincII and SpeI (50µl purified PCR product, 7µl digestion buffer (NEB), 3µl HincII, 3µl SpeI, 0.7µl BSA, 6.3µl H_2O incubated at 37°C for 10hour) and same time pBlueScript II KS+ was double digested with HincII and SpeI (6µl pBlueScript II KS+ (200 ng/ml), 1µl digestion buffer (NEB), 0.1µl BSA, 1µl HincII, 1µl SpeI, 0.9µl H_2O incubated at 37°C for 10 hours). After restriction digestion, both products were purified and ligated (1µl digested purified pBlueScript II KS+, 7µl purified PCR product, 1µl T4 DNA ligation Buffer(10X), 1µl T4 DNA ligase and incubated 15 hours at 16C). Positive clones were screened by sequencing and named (3'-9.5-Race-hGMS, 3'-9.6-Race-hGMS, 3'-9.8-Race-hGMS, 3'-R.1-Race-hGMS, 3'-R.6-Race-hGMS, 3'-R.9-Race-hGMS, 3'-R.3-Race-hGMS)

II.2.9.Site Directed Mutagenesis

Site directed mutagenesis was carried out with the modification of "QuickChangeTM XL Site-Directed Mutagenesis" Kit from strategen. Modifications; amount of plasmid used as template 20-60µl, amount of primers 100-125ng, DpnI digestion at least four hours.

II.2.10.Real-Time-PCR on Cells

Mouse L929 fibroblasts were stimulated for 24 h with 200U/ml IFN-γ or 200U/ml IFN-β (R&D and Calbiochem respectively). The induction ration between induced and uninduced p47 GTPases (Irgm1 and Irga6) was detected by a quantitative PCR assay using the LightCycler System (Roche). cDNA synthesized using the mRNA prepared from the induced and un-induced cells was used as a template. The amount of measured transcripts was normalized to the amount of the mouse GAPDH transcript in the probes. The sequences of all primers are listed in List of primers (see above). The reaction performed using Quantitect SYBR Green (Qiagen, Hilden) according to the manufacturer instructions with the modifications; addition of extra 1U taq polymerase (Rita Lange personal communication) and PCR program (95°C, 3 min denaturation step.). Melting curve analysis was performed after each run to analyse specificity of primers. To generate regression curve as standard for calculation of molucules, pGEMT-easy+Irgm1 and pGEMT-easy+Irga6 were used in serial dilutions.

II.2.11.Quantification by UV Spectroscopy

DNA concentrations were determined by using UV/Vis Spectrophotometer (Biomate 3 Thermo spectronic). 1 µl of genomic DNA was diluted in 1 ml of TE buffer (pH: 7,5) and DNA quantification was done at 260 nm and 280 nm wavelengths. 260 nm wavelengths show the concentration of nucleic acid in the sample. 1 OD (Optic Density) at 260 nm approximately is equal to 50µg/ml double helical DNA in the sample. The ratio of two values that were read at 260 nm and 280 nm measures the purity of the nucleic acids. The ratio of OD_{260}/OD_{280} must be between 1.8 and 2.0. This ratio is drastically decreased if protein or phenol remains in the solution.

II.2.12.Checking the Presence of DNA on Agarose Gel Electrophoresis

Agarose gel electrophoresis was used for qualitative analysis of extracted DNA. Agarose gels were prepared in respective percentage (0.6%- 1%) by boiling agarose in 1X TAE or 0.5XTBE buffer. The gel was poured onto an electrophoresis plate and gel was left in room temperature for 30 minutes for polymerization. 1.0 µl of genomic DNA, 6.0 µl of 6X bromophenol blue dye and 6.0 µl of dH_2O mixed (for PCR products DNA quantities were varied from 5µl to 20µl according to experiment) and the gel was run at 70-120 V for 60-30 minutes respectively, and stained in 0,5µl/ml ethidium bromide (EtBr) solution. It was visualised under UV light. The quality of DNA was determined by looking at the migration patterns of the bands on the gel and the presence or absence of smears.

II.2.13.Preparation of Competent Cells I

The preparation was started with 5 mL overnight culture of E. coli cells in a LB medium which contains 0.02M $MgSO_4$, 0.01M KCl at 37°C with 180rpm. Next day it was inoculated in a ratio 1:10 in fresh LB with same contents above up to approx. OD_{600} = 0.1 and incubated until reaching OD_{600} = 0.45 at 37°C with 180rpm. The flask was put on ice or in cold room for 10 min. Cells were pelleted at 5000 g for 10 min, 4 °C. pellet was resuspended in 100 mL cold TFB I and incubated on ice for 10 min, Cells again spun down at 4000 g for 5min, 4 °C. The cells were resuspended carefully in 20 mL TFB II. Finally, the cells were aliquoted into precooled 1.5 mL tubes (100-300µL/tube) and frozen in liquid nitrogen, finally aluquots were transferred to -80 °C. 100 µL/ is used for transformation.

Buffers:

TFBI: 30 mM KOAc; 50 mM $MnCl_2$; 100 mM KCl; 10 mM $CaCl_2$; 15 % (w/v) Glycerin.

TFBII: 10 mM Na-MOPS pH 7.0; 75 mM $CaCl_2$; 10 mM KCl; 15 % (w/v) Glycerin.

II.2.14.Preparation of Competent Cells II

5-10 E. coli DH5α colonies were inoculated into 250ml of LB + 50mM $MgCl_2$ and incubated at 24°C (optimal 18°C) to OD_{600} = 0.450-0.600 (it takes 10 to 12 hour at 24°C several days at 18°C). Cells were cooled on ice for 10 min and pelleted by centrifugation 8min at 4500g at 4°C in 50ml tubes. Cells were resuspended with 80ml ice-cold TB buffer (total) and incubated for 10 min on ice. The cells were centrifuged for 4500g for 5 min. and pellet was resuspended again in 20 ml ice-cold TB buffer. DMSO (room temperature) was added to 7% final concentration and incubated on ice for 10min. The cells were aliquoted in pre-cooled eppendorf tubes and shock frozen in liquid nitrogen. Competent cells were kept in -80°C.

TB buffer

10mM Pipes, 55mM $MnCl_2$, 15mM $CaCl_2$, 250mM KCl everything added except $MnCl_2$ adjusted to PH:6.7 with KOH. Finally, $MnCl_2$ is added and stored at 4°C.

II.2.15.*E. coli* Transformation

Cells were thawed on ice for 10 min. then plasmid DNA or ligation reaction was mixed with 100 µl competent cells in 1.5 ml tube (usually to 5/10 µl of ligation reaction used). Competent Cells and DNA mixture was incubated on ice for 20 min. followed by heat shock at 42 °C for 2 min. in water bath. 500 or 1000 µl of fresh LB was added and incubated on rotater for 45-60 min. then plated 250 - 500 µl on selection plate.

II.2.16.Prufication of IRGM(a) Protein, N-Terminally Fused to GST Protein

With this method it was possible to get low amount (approx. 100µM) of IRGM(a) N-terminally fused with GST protein which was going to the inclusion bodies.

Construct= PGEX 4T-2 + hGMS

Bacteria= E.Coli (BL-21)

(1). A 20 ml over night culture was incubated at 37 °C, 180 rpm **(2).** The O/n incubated culture was diluted 1:100 in 200 ml Terrific Broth, $Amp_{100µg/ml}$ and incubated at 37 °C, 180 rpm around 2 hours until the growth reaches between 0.200 and 0.400 **(3).** The E. coli culture was stored at cold room (4 °C) around 1 hour to cool down temp. before the induction with

IPTG, and was induced with 50µM IPTG (10 µl from 1M to 200ml) **(4).** Incubation was maintained over-night at 18 °C, 180 rpm. **(5).** The culture was separated into 100ml centrifugation tubes (250ml) and centrifuged at 5000g, 15min. at 4 °C .*After this step everything was maintained on ice* **(6).** Each pellet was weight (for my case=1.390gr) and washed 1time with Pre-Washing Buffer, (Resuspending by pipetting up and down slowly) **(7).** Centrifuged at 5000g, 15min. at 4 °C, supernatant was removed and samples were immediately frozen in Liquid Nitrogen (-80) **(8).** Each frozen pellet was resuspended in 20 ml Sonification Buffer by pipetting up and down slowly.1 tablet of Protease inhibitor Tablet was added in 20ml solution (Before adding Sonification buffer, Tubes were kept in room Temperature for 10min.) **(9).** When two pellet completely resuspended, each solutions were mixed together (40 ml) before sonification **(10).** Total 40 ml Solution was sonificated 15 times 30sec. with 30sec. break on ice. (Sonifier 450= Output Control (5), Duty Cycle (Constant), Timer (0)) (Bramson TM2= Time (max.), Temp (max.), 1X) (During the sonification, temperature was always checked to keep sample cold, some times was waited longer then 30sec.) *After this step everything was maintained in Cold Room* **(11).** An equal volume (40ml) of 20% Glycerol + sonification buffer was added on sonificated culture in drop wise (Using Gravity Column with flow rate 1ml/min) by slowly stirring. **(12).** The sonificated culture was centrifuged at 100.000g (25000rpm in Beckmann Class H with SW 41 Ti rotor), for 30min, at 4 °C. **(13).** The supernatant was directly added on Gravity Column containing 3ml of GST beads which were previously prepared by washing with 10% Glycerol + sonification buffer several times of column volume. The flow rate was adjusted to 1ml/min. **(14).** The GST-beads was washed with 3times 10% Glycerol+sonication buffer and each wash fractions were collected to check on SDS-PAGE. **(15).** The column was washed with 2times wash buffer and each wash were collected to check on SDS-PAGE **(16).** 2ml of 20mM reduced glutathione (In Wash Buffer) was added on the column with 5 ml glass pipet and pipetted several times to mix GST beads with glutathione solution properly. **(17).** Elution of recombinant protein was further performed with additional 4times, 2ml of reduced glutathione in a same way above. (In each step 5min incubation is made to elute protein) **(18).** GST beads were washed with 2 times with 20 ml of wash buffer and 2times of 20ml water **(19).** 10ml of 6M Guanidium Cloride was added to the column and washed extensively 2times in a 100ml of volume **(20).** GST-beads were stored in 30% EtOH + water

Solutions

Pre-Washing Buffer

50mM Hepes-NaOH	10ml	(0.5 M PH:8.0)
100mM NaCl	10ml	(1M)
H_2O to final Conc.	100ml	PH adjusted to 7.5

Sonificatin Buffer

50mM Hepes-NaOH	10ml	(0.5 M PH:8.0)
100mM NaCl	10ml	(1M)
5mM DTT	500µl	(1M)
5mM $MgCl_2$	500µl	(1M)
30µM GDP	100µl	(30mM)
H_2O to final Conc.	100ml	PH adjusted to 7.5

20% Glycerol + Sonificatin Buffer

50mM Hepes-NaOH	10ml	(0.5 M PH:8.0)
100mM NaCl	10ml	(1M)
5mM DTT	500µl	(1M)
5mM $MgCl_2$	500µl	(1M)
30µM GDP	100µl	(30mM)
20% Glycerol	20ml	
H_2O to final Conc.	100ml	PH adjusted to 7.5

10% Glycerol + Sonificatin Buffer

50mM Hepes-NaOH	10ml	(0.5 M PH:8.0)
100mM NaCl	10ml	(1M)
5mM DTT	500µl	(1M)
5mM $MgCl_2$	500µl	(1M)
30µM GDP	100µl	(30mM)
10% Glycerol	10ml	
H_2O to final Conc.	100ml	PH adjusted to 7.5

Wash Buffer

50mM Hepes-NaOH	10ml	(0.5 M PH:8.0)
100mM NaCl	10ml	(1M)
1mM DTT	100µl	(1M)
5mM $MgCl_2$	500µl	(1M)
30µM GDP	100µl	(30mM)
10% Glycerol	10ml	
H_2O to final Conc.	100ml	PH adjusted to 7.5

II.2.17.Prufication of IRGM(a) Protein, N-Terminally Fused to MBP Protein

After cloning the gene into the expression vector pMAL-p2E + Thrombin digestion site, clones for expression of the fusion protein were incubated overnight 24°C in NB42. The

MBP-IRGM(a) fusion protein was purified from a bacterial lysate by binding to an amylase resin. After washing the resin of bacterial impurities, the fusion protein was eluted off the amylase resin with 10 mM maltose. Using this method more then 10mg/ml MBP-IRGM(a) fusion protein was purified. (This protocol was adapted from original protocol of Donald Ria, University of California Berkeley, Bio reagents and Chemicals)

Construct= pMAL-p2E +Tr+hGMS

Bacteria= E.Coli (NB-42)

(1). A 6L of rich media was inoculate with 10ml/L (1 to 200) of an overnight culture with strain expressing the MBP-IRGM(a) fusion protein. The culture was incubated at 37°C temperature until culture reaches to an optical density of 0.5 at 600nm (OD_{600} of 0.5) approx 3 to 4 hour. **(2).** Culture was induced with 400µl of 0.5 M IPTG (0.1 M IPTG for final concentration) for over night at 24°C. **(3).** The culture was centrifuged for 15 min. by using Beckmann 1L rotor at 6000 rpm (5000g) at 4°C **(4).** The supernatant was removed and the cell pellet was resuspended in 20 ml of ice-cold lysis buffer. 1X Protease inhibitor cocktail was added on pellet and the tablet was solubulized together with pellet. (2 tablet from 1 to 10 Complete mini Roche). **(5).** The cell suspension was transferred to 15 ml falcon tubes and was snap frozen in liquid nitrogen. **(6).** The cell suspension was thawed in cold water. **(7).** The thawed cell suspension was sonicated by using the 30 second burst with 30 second break at setting of 5 with the tip of a cell sonicator probe in 120 ml beher. (Be sure to minimize foaming, sample should be kept in ice water bath during sonication)The sonication bursts were repeated until no more protein was released. The protein release was mnitored by 10 µl aliquots of the lysate by bradford assay (2ml). **(8).** The cell suspension was centrifuged at 4°C for 30 min at 50 000 g. **(9).** At the same time, the amylase resin column was prepared by pouring 60 ml of resin into 1 g coloumn 2.5 cm in diameter. The capacity was expected to be 3 mg of maltose binding protein/ml resin in theory but for IRGM(s) fused protein capacity was around 1.5 mg/ml. **(10).** The column was equlibrated by lysis buffer with 2-5 column buffer (column buffer should be around 35 ml). **(11).** The flow rate was adjusted to 1ml/min but this flow rate was reduced during the process. **(12).** the sample was loaded and the flow though was collected and kept at 4°C so that it can be used again to load the column several times. Because it is observed that there was always protein not bound to the coloumn. **(13).** Once loaded, the column was washed with 10 column volumes of elution buffer. **(14).** MBP-IRGM(a) fusion protein was eluted by elution buffer with 5 column volume, 5 ml each fractions. Usually within first 3 fractions elution of MBP-IRGM(a) protein should be observed . This was very much depends on how old the column material was. **(15).** A 50 µl

of thrombin (5 unit/ml) was added to each elution fractions containing 10ml of highly concentrated protein (average 2mg/ml protein) and kept at 4°C o/n for complete digestion. **(16).** The digested protein was concentrated by using Vivaspin centrifugal concentrator with 10000MW cut off with 20 ml capacity up to appropriate volume. **(17).** The concentrated protein was subjected to gel filtration column (Hi-Load 26/60 superdex 75 prep grade, resolution 3000-70 000 or).

Solutions

Elution Buffer
20 µl of (GDP, GTP or GTPγS)
5mM $MgCl_2$
200mM NaCl (No difference is detected between 150-300mM)
50mM Tris-HCl, PH:8.0
2mM $Na_2S_2O_5$ (Sodium metabisulfide)
10mM Maltose
1 mM DTT
10% (v/v) Glycerol
PH: 8.0
Add the DTT and $Na_2S_2O_5$ fresh before use

Wash Buffer
1mM PMSF
5mM $MgCl_2$
300mM NaCl
50mM Tris-HCl, PH:8.0
2mM $Na_2S_2O_5$ (Sodium metabisulfide)
10mM Maltose
1 mM DTT
10% (v/v) Glycerol
PH: 8.0
Add the DTT, PMSF and $Na_2S_2O_5$ fresh before use

Lysis Buffer
1mM PMSF
5mM $MgCl_2$
1 M NaCl
100mM KCl
50mM Tris-HCl, PH 8.0
10mM Maltose
0. 5 mM DTT
PH: 8.0
Add the DTT, PMSF and protease inhibitor cocktail fresh before use

II.3.EVOLUTIONARY AND PHYLOGENETICS ANALYSIS

II.3.1.Use of Database Resources

All available public databases were extensively screened by BLAST and related searches for sequences belonging to the IRG family. In the case of the mouse, transcript sequences derived from the C57BL/6 strain were given preference over sequences of other and undefined strain origin and compared in all cases with genomic sequence available via the ENSEMBL and NCBI. A systematic study of polymorphism has not yet been completed, but it is already clear that nearly all IRG sequences derived from the CZECHII cDNA libraries (*Mus musculus musculus*) differ from C57BL/6 sequences. These differences make allocation of many CZECHII sequences to individual clade members of the C57BL/6 mouse problematical. Identification of certain *Irg* sequences with recognised gene symbols was achieved through the Mouse Genome Initiative web resources at http://www.informatics.jax.org/.

Human and dog IRG sequences were identified from the available public databases (ENSEMBL, NCBI) and confirmed wherever possible by multiple sequence comparisons at transcriptional and genomic level. Fugu material was obtained and analysed through the BLAST server at http://fugu.hgmp.mrc.ac.uk/ and ENSEMBL web site at http://www.ensembl.org/Fugu_rubripes/. Tetraodon sequence was initially assembled from the GSS sequence database at NCBI and subsequently from the UCSC compiled genome database via the BLAST server at http://genome.ucsc.edu/cgi-bin/hgGateway. Zebrafish sequence was obtained from zebrafish genome resources at http://www.sanger.ac.uk/Projects/D_rerio and analysed in an Acedb database using the Spandit annotation tool.

Chromosomal locations and synteny analysis of mouse and human chromosomes was initiated through http://www.ensembl.org/Mus_musculus/syntenyview. Further details were obtained through http://www.sanger.ac.uk/Projects/M_musculus/publications/fpcmap-2002/mouse-s.shtml. Protein molecular weight calculations is maintain by using available free calculation program at (http://bioinformatics.org/sms/prot_mw.html)

II.3.2.Phylogeny and Alignment Protocols

Routine sequence analysis and local sequence database management was handled using DNA-Strider 1.3f12, Vector-Nti and MacVector 7.2. The identity and similarity matrix of protein and nucleotide sequences (Table 2) are based on GeneDoc version (# 2.6.0002). Phylogenetic analysis was conducted using the neighbor-joining (NJ) method (Saitou and Nei,

1987), as implemented in the MEGA2 program (Kumar et al., 1994). We used p-distances for constructing the phylogenetic trees. Reliability of the NJ trees was examined by the bootstrap test (Felsenstein, 1985).

Alignments were performed via the BCM multiple alignment programme suite (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) and EBI-Clustalw (http://www.ebi.ac.uk/clustalw/) using the default options and manipulated according to the crystal structure of IIGP1 (Ghosh et al., 2004). Shading of alignments was performed with Boxshade (http://www.ch.embnet.org/software/BOX_form.html) and additional sequences were shaded manually according to the default options of Boxshade. Contig assembly was performed either by using the TIGEM, Cap3 (http://fenice.tigem.it/bioprg/interfaces/cap3.html) or Infobiegen (http://www.infobiogen.fr/services/analyseq/cgi-bin/cap_in.pl).The ests for contigs were edited for sequences error when necessary. Ka/Ks (codon based selection test) analysis was performed using the program K-Estimator 6.0 (Comeron, 1999).

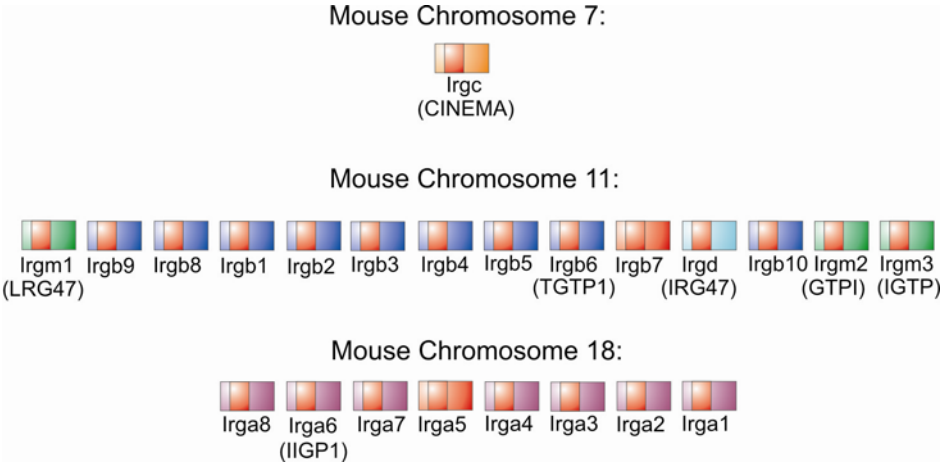
II.3.3.Identification of Transcription Factor Binding Sites

Promoter regions (2 kb upstream of putative transcription start point) were screened for putative transcription factor binding sites with the Transcription Element Search System (TESS, http://www.cbil.upenn.edu/tess) and the results were further analysed and confirmed manually (Schug and Overton, 1997). Additional promoter analysis of Irgc (mouseCinema) and IRGC (humanCINEMA) was performed with ConSite (Lenhard et al., 2003) based on phylogenetic footprinting (http://www.phylofoot.org).

III.I.RESULTS I.

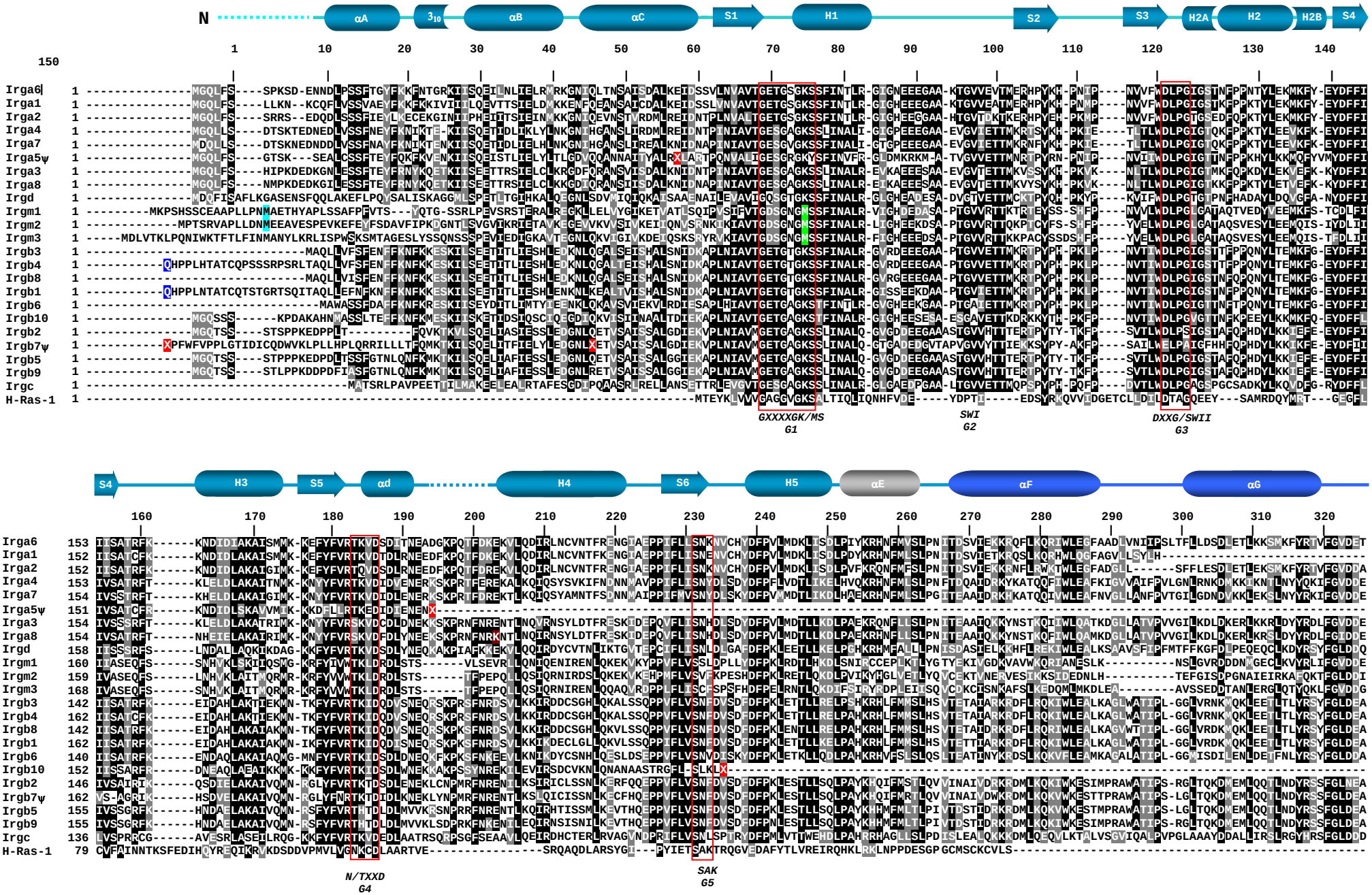
III.I.1.Genomic organization of the p47 (IRG) GTPase genes of the C57BL/6 mouse.

Using a combination of screens and supplementary analysis (see Materials and Methods), the following genomic representation for the p47 GTPases of the C57BL/6 mouse was established. The p47 (IRG) GTPases form a well-defined family of 23 members distributed on mouse chromosomes 7, 11 and 18:



A general nomenclature on phylogenetic principles is introduced for the p47 GTPases, based on the stem name IRG (immunity-related GTPases). The sources of all Irg sequences in database and assignment of genes with previously published names are listed in appendix table 1. However, from now on, the name “p47 GTPases or p47 GTPase family” will be used to describe the family considering historical reason. ORFs of individual members of p47 GTPases can be found in our p47 GTPases database (http://www.genetik.uni-koeln.de/groups/Howard/index.html). From the open reading frames of these genes protein sequences were predicted and aligned in fig. 7 (see below). Among the 23 p47 homologous genes, two are putatively pseudogenes based on criterion of inability to code for a functional GTPase domain. Irga5 is highly degraded pseudogene resulting in putative loss of coding full length p47 GTPases and this appears to be a recent event. Irga1 has a perfect open reading frame from the putative initiator methionine until residue 298 of the sequence and runs out of frame through a 4 base pair deletion followed by a single base loss (See Fig. 7). Although transcribed, Irga1 appears to be further damaged by an unexplained failure to splice correctly from exon 1 to exon 2. Both donor and recipient splice sites appear normal (see below). Thus, Irga1 is expected to be expressed normally in the cell based on the information both from general principles and from homology to closely related p47 genes such as Irga2 and Irga6. However only a single correctly spliced transcript is found in the Est database (BI658674).

Irga8 is assigned pseudogene status in C57BL/6 mice because of a single base insertion at position 204 in the second exon, resulting in a frame shift at amino acid lysine (K) (see Fig 7). The strongest evidence for this insertion to be a recent event is the presence of an intact version of Irga8 in the closely related mouse species, *Mus musculus musculus*, represented by the Czech II strain for which an extensive EST database is available. With this exception of a single base insertion, the open reading frame of Irga8 is complete and shows close homology to other p47 GTPases. In the p47 gene clusters on chromosome 11, Irgb7 is identified as a pseudogene on the bases of a single base change which mutates residue glutamine (Q) of the putative open reading frame to a stop codon (X) (Fig 7). In addition, no transcript of Irgb7 has yet been found either as an EST or by RT-PCR studies (see below). Irgb10, is another truncated p47 GTPase despite being transcribed and interferon-inducible, because its ORF terminates shortly after those regions homologous to other p47 GTPases, breaking off at amino acid 232. Additionally, no homologous sequence is to be found in the underlying DNA. Thus, Irgb10 appears to be a relatively recent 5‘gene fragment coding only for the G-domain of p47 GTPases. The remaining 19 p47 genes appear to be intact in the open reading frame. Thus, a minimum estimate of the number of potentially functional p47 GTPases in mouse is not just six, as previously described, but rather 20.


N
αA
3₁₀
αB
αC
S1
H1
S2
S3
H2A
H2
H2B
S4
GXXXXGK/MS
G1
SWI
G2
DXXG/SWII
G3
H3
S5
αd
H4
S6
H5
αE
αF
αG
N/TXXD
G4
SAK
G5
Irga6
Irga1
Irga2
Irga4
Irga7
Irga5ψ
Irga3
Irga8
Irgd
Irgm1
Irgm2
Irgm3
Irgb3
Irgb4
Irgb8
Irgb1
Irgb6
Irgb10
Irgb2
Irgb7ψ
Irgb5
Irgb9
Irgc
H-Ras-1

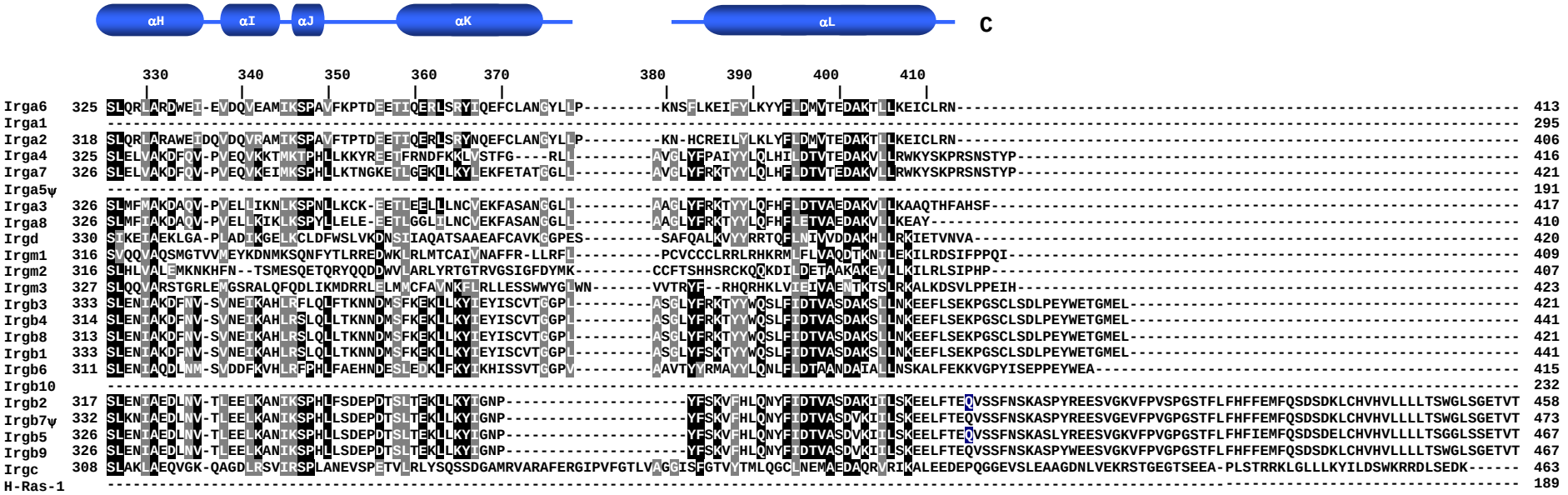


Figure 7. Amino acid alignment of the mouse Irg GTPases.
Sequences of all 23 mouse Irg GTPases showing the close homology extending to the carboxyl-terminus, are aligned on the secondary structure of Irga6 (indicated by the secondary structure elements drawn in blue above the sequence alignment). The sequences of notional products of the two pseudo-genes Irga5 and Irgb7 have been partially reconstructed; premature terminations are indicated in red. In the C57BL/6 mouse, the sequence of the Irga8 gene is altered by an adenine insertion, indicated by the red highlighted lysine (K) at position 204. (The sequence after this position is given after correcting the frameshift, and is identical to that of the CZECHII (*Mus musculus musculus*) sequence BC023105 that lacks the extra adenine.) The turquoise-highlighted M in M1 and M2 are initiation codons that are dependent on alternative splicing (also see Figure 9); the unusual methionine residues in the G1 motif of GMS proteins are highlighted in green. The blue background Q residue of Irgb5 and Irgb2 at positions 405 and 396 represents the point at which tandem splicing occurs to Irgb4 and Irgb1, respectively. Canonical GTPase motifs are indicated by red boxes. The nucleotide and amino acid sequences themselves can be obtained in the p47 (IRG) GTPase database from our laboratory website (http://www.genetik.uni-koeln.de/groups/Howard/index.html).

Analysis of the relationship between the p47 GTPases based on nucleotide sequence delivers suggestive clues to understand phylogenetic events that generate complexity in gene families. The multiplex block of 13 genes on chromosome 11 contains most divergent sequences, including all three representatives of the GMS GTPases, LRG-47 (Irgm1), IGTP(Irgm3) and GTPI (Irgm2), and the singlet sequence for IRG-47 (Irgd), as well as the previously isolated TGTP (Irgb6) sequence now accompanied by 8 further representatives. Fig 8a shows a phylogeny generated from the full length nucleotide alignment of the p47 GTPases and Fig 8b shows an alignment generated from the G domains alone (according to structure of IIGP1(Irga6)). The deep roots connecting the p47 GTPases on chromosome 11 suggests that this cluster is relatively ancient. In contrast, all eight genes clustered on chromosome 18 show a degree of homogeneity, suggesting relatively recent divergence, with a plausible ancestral relationship to a member of the TGTP (Irgb6) cluster on chromosome 11. In contrast, the isolated p47 gene on chromosome 7 seems to represent an ancient root with no obvious systematic relationship to any of the other subfamilies. Within the chromosomal clusters, more recent duplication events are apparent, thereby linking Irga1, Irga2 and Irga6, Irgb1, Irgb3, Irgb4 and Irgb8, Irgb2, Irgb4, Irgb7 and Irgb9. The open reading frame of the adjacent sibling pair Irgb3 and Irgb4 differ only by nine nucleotides. Table 2 gives the nucleotide and protein sequence identities across the aligned open reading frames of the 23 complete genes of the p47 family, in order of dissimilarity, showing the wide evolutionary divergence between the more distant branches of the tree. The pattern of divergence in the p47 tree suggests a relatively old gene family that has undergone a succession of duplication-divergence cycles over time, a pattern of evolution, which is still actively continuing in several of the subfamilies (see discussion).

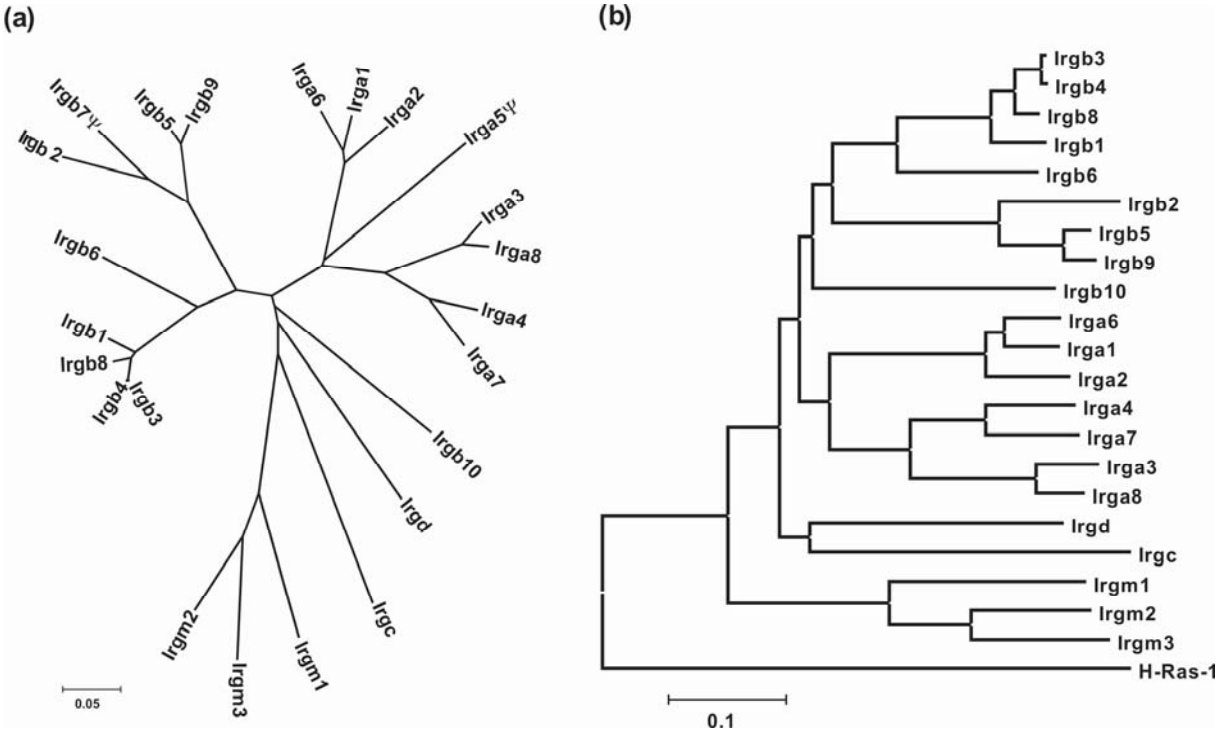


Figure 8. Phylogenetic relationship of mouse Irg GTPases.
(a) Unrooted tree (p-distance based on neighbour-joining method) of nucleotide sequences of the G-domains of the 23 mouse Irg GTPases, including the two presumed pseudogenes Irga5 and Irgb7. **(b)** Phylogenetic tree of the amino acid sequences of the G-domains of 21 mouse Irg GTPases rooted on the G-domain of H-Ras-1 (accession number: P01112). The products of the two presumed pseudo-genes Irga5 and Irgb7 are excluded from the analysis.

Irg	m1	m3	m2	b3	b4	b8	b1	b6	b2	b7	b5	b9	b10	a2	a6	a1	a5	a3	a8	a4	a7	d	c
m1		0.71	0.75	0.53	0.53	0.52	0.51	0.53	0.45	0.46	0.48	0.49	0.42	0.47	0.48	0.48	0.47	0.46	0.45	0.46	0.47	0.50	0.51
m3	0.64		0.80	0.49	0.49	0.48	0.49	0.49	0.44	0.45	0.47	0.47	0.40	0.47	0.48	0.48	0.45	0.46	0.46	0.46	0.46	0.47	0.46
m2	0.65	0.77		0.53	0.52	0.53	0.53	0.53	0.46	0.47	0.49	0.50	0.41	0.50	0.51	0.51	0.48	0.48	0.49	0.49	0.49	0.50	0.49
b3	0.40	0.38	0.40		0.99	0.96	0.94	0.80	0.68	0.69	0.74	0.73	0.53	0.64	0.64	0.65	0.60	0.62	0.63	0.63	0.62	0.61	0.55
b4	0.40	0.37	0.40	0.99		0.96	0.94	0.80	0.68	0.69	0.74	0.73	0.53	0.64	0.64	0.65	0.60	0.62	0.62	0.63	0.62	0.61	0.55
b8	0.40	0.37	0.39	0.95	0.95		0.95	0.81	0.68	0.69	0.73	0.73	0.53	0.65	0.65	0.66	0.61	0.63	0.63	0.64	0.62	0.61	0.55
b1	0.40	0.37	0.41	0.89	0.89	0.91		0.81	0.66	0.68	0.72	0.71	0.52	0.64	0.64	0.65	0.61	0.62	0.63	0.65	0.63	0.61	0.54
b6	0.41	0.34	0.39	0.73	0.73	0.75	0.75		0.65	0.66	0.69	0.69	0.53	0.64	0.64	0.66	0.59	0.62	0.62	0.63	0.63	0.63	0.56
b2	0.34	0.31	0.33	0.55	0.55	0.57	0.52	0.52		0.85	0.82	0.80	0.45	0.57	0.57	0.57	0.54	0.56	0.57	0.56	0.56	0.52	0.46
b7											0.83	0.82	0.47	0.58	0.57	0.58	0.56	0.58	0.58	0.59	0.58	0.54	0.47
b5	0.35	0.33	0.36	0.59	0.59	0.60	0.56	0.54	0.82			0.96	0.50	0.62	0.61	0.62	0.59	0.62	0.62	0.61	0.60	0.57	0.53
b9	0.35	0.33	0.35	0.57	0.57	0.58	0.54	0.54	0.80		0.94		0.51	0.61	0.61	0.62	0.59	0.62	0.62	0.62	0.60	0.57	0.52
b10	0.33	0.31	0.31	0.47	0.47	0.48	0.47	0.49	0.39		0.40	0.40		0.50	0.49	0.50	0.47	0.48	0.48	0.50	0.49	0.49	0.43
a2	0.37	0.34	0.35	0.54	0.54	0.55	0.55	0.57	0.47		0.48	0.48	0.44		0.89	0.90	0.69	0.70	0.71	0.70	0.70	0.61	0.51
a6	0.36	0.34	0.35	0.57	0.56	0.57	0.57	0.59	0.47		0.49	0.50	0.44	0.84		0.92	0.70	0.69	0.70	0.70	0.70	0.61	0.52
a1	0.38	0.37	0.37	0.56	0.57	0.56	0.56	0.60	0.47		0.50	0.50	0.43	0.85	0.90		0.70	0.69	0.70	0.71	0.71	0.62	0.51
a5																		0.68	0.68	0.68	0.67	0.57	0.49
a3	0.34	0.33	0.34	0.51	0.51	0.52	0.50	0.53	0.48		0.51	0.51	0.39	0.56	0.54	0.55			0.95	0.80	0.80	0.59	0.52
a8	0.35	0.33	0.35	0.54	0.54	0.54	0.54	0.55	0.47		0.50	0.50	0.39	0.57	0.56	0.58		0.90		0.81	0.81	0.59	0.51
a4	0.35	0.35	0.35	0.56	0.56	0.56	0.56	0.54	0.46		0.49	0.49	0.43	0.58	0.59	0.60		0.68	0.69		0.90	0.62	0.51
a7	0.36	0.34	0.36	0.55	0.55	0.55	0.55	0.56	0.45		0.49	0.48	0.42	0.59	0.59	0.61		0.69	0.71	0.84		0.60	0.49
d	0.37	0.35	0.38	0.52	0.52	0.53	0.53	0.55	0.40		0.44	0.44	0.44	0.52	0.52	0.53		0.48	0.49	0.50	0.49		0.56
c	0.35	0.36	0.35	0.46	0.46	0.45	0.44	0.45	0.37		0.39	0.39	0.36	0.44	0.45	0.44		0.45	0.45	0.44	0.43	0.50	

Table 2.Nucleotide and amino acid identities based on the G-Domain of the mouse Irg family.
Identity matrix of pairwise aligned nucleotide (gray background) or amino acid (white background) sequences of mouse Irg family members. Matrix was generated using the GeneDoc program. Pseudogenes, Irgb7 and Irga5 are excluded from protein analysis.

III.I.2.The structure of p47 GTPase genes and their splicing patterns.

The genes of the p47 family have a distinctive signature common to the whole family (Fig 9). The entire open reading frame is encoded on a single long exon with the initial ATG close to the splice acceptor site for one or more untranslated 5‘ exons. All the splicing acceptor and donors are listed in appendix table 2. In two cases (LRG47 and one of the splice forms of GTPI), the methionine is encoded at the 3‘ end of the previous exon, giving 3 or 4 N-terminal amino acid residues encoded by the upstream exon. In the case of GTPI, a second methionine classically positioned at the 5‘ end of the long exon 3 is used as the initiator codon in the most common form splicing directly from exon 1 to exon 3. Three genes of the chromosome 18 cluster have unusual genomic structures. The strongly expressed Irga6 (IIGP1) gene has two alternative untranslated 5‘exons (exon 1A and exon 1B) each independently furnished with a functional promoter (see below).

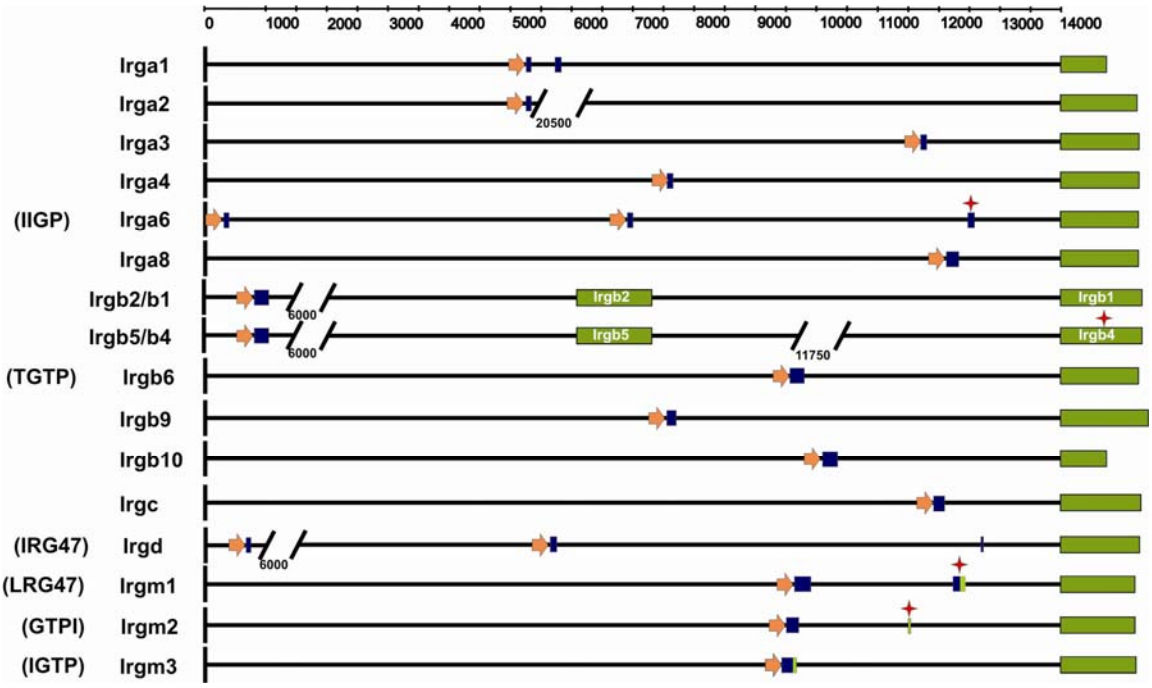


Figure 9. Genomic structure of mouse Irg GTPases.
Genomic structure of mouse Irg genes. Green blocks indicate coding exons and blue blocks indicate 5’-untranslated exons. Orange arrows identify putative promoter regions. Stars represents exons shown to be excluded in alternative splice forms. The scale bar is measured in base pairs up to the first base of the long coding exon. Note the presence of two promoters for Irga6 and Irgd.

The close homologues Irga1 and Irga2 are closely related to each other (Fig. 7 and Table 2) and exon 1 of Irga1 is used as the first exon of Irga2, entailing an intron length of 35 kb containing the Irga1 putative pseudogene as well as the completely intergenic interval between Irga1 and Irga2. A genomic sequence apparently homologous to exon 1 of Irga1 is present 7 kb upstream of the coding exon 2 of Irga2 but is not apparently provided with an adequate promoter and has not yet been observed in a cDNA. Exon 1 of Irga1 also splices to

acceptors upstream of the coding exon of Irga1, but only a single cDNA is recorded where the correct 5‘splice acceptor site is used (BI658674). A recorded Est of Irga1 contain long genomic sequences upstream of the coding exon resulting in multiple starts in incomplete reading frames (BG915086).

A further splicing anomaly found in the public database connects Irgb1-Irgb2 and Irgb5-Irgb4 which is indicated in fig 9 and 10a. These genes are adjacent and in the same polarity on chromosome 11, with Irgb2 upstream of Irgb1 and Irgb5 upstream of Irgb4. The only transcripts seen containing Irgb1 are tandem structures in which the long exon 2 of Irgb1 is preceded in frame by the long exon 2 of Irgb2. This appears to result from sporadic use of a cryptic splice donor site near the termination codon of Irgb2 resulting in splicing with the splice acceptor site of the long coding exon of Irgb2. RT-PCR analysis using primers from the 5‘ end of Irgb1 and the 3‘ end of Irgb2 results in an amplification of a long, interferon-inducible cDNA consistent with a fusion transcript (Fig 10b). A full-length cDNA representing the Irgb2-Irgb1 tandem sequence derived from “Mammary tumor metastatized to lung” is present in the NCBI public databases. The corresponding tandem sequence was amplified and shown to be inducible by interferon γ on the cDNA synthesized using RNA prepared from L929 cells (Figure 10b). Since no ESTs of Irgb1 alone have been reported so far and due to inability to detect Irgb1 without the Irgb2 tandem, it may perhaps be reasonable to consider Irgb1 simply as a second long coding exon of Irgb2 rather than as a gene by its own right. However, the situation is different for Irgb5-Irgb4 tandem, since single Est AK037088 can be found in public databases, which does not splice into Irgb4, thus Irgb5 can exist as a single gene or as a tandem gene together with Irgb4 (See Fig 9 and 10a). In my RT-PCR analysis, Irgb4 (or Irgb3, these two sequences only differ by nine nucleotides from each other especially in 5’ prime region) is constitutively transcribed in mouse L929 cells. However, it is not clear whether Irgb4 is transcribed alone or only the second long exon of the Irgb5 because the amplification product of RT-PCR was specific to long coding exon of Irgb4 and in database, there is no Est available for Irgb4. Therefore, Irgb4 was considered as an alternative splicing form of Irgb5. Rat Irgb13 and Irgb14 represents same structure with Irg2-Irgb1 and Irgb5-Irgb4 tandems, therefore Irgb14 and Irgb13 are considered to be transcribed as tandem. However, there is no Est was reported in public rat databases either for tandem or for individual Irgb13 and 14.

Furthermore, new type of tandem gene formation has been identified in rat, which apparently contains rat Irgb10, Irgm2, Irgm3, encoding the three GTPases together on a single transcript, AY321344 (Fig 10 and Appendix Table 3). These genes are adjacent and have the

same polarity on BAC AC097938.6 localized to chromosome 10 in rat. This corresponds to the homologous Irgm2, 3 and Irgb10 in the same order as on the mouse chromosome 11 (see above). Alignment of the triple gene with individual rat Irgb10, Irgm2 and Irgm3 is shown in appendix fig. 1. After five times splicing by having short peptides which is unrelated to p47 GTPases, the triple gene starts with a GKS like GTPase characteristics-Irgb10 sequence with a classical myristoylation signal MGxxxS. After coding the whole N-terminal, G domain and C-terminal region of rat Irgb10, the triple tandem splices into one of the GMS type GTPase (rat Irgm3). Finally, the tandem ends with another GMS type GTPase (rat Irgm2) which is linked to the previous GTPases by three splicing with short peptides. End of the third gene codes classical C-Terminal sequence of GMS type GTPase (Irgm2) and splice into short sequences, which has a putative stop codon and is unrelated to any known p47 GTPase. Analysis of all splicing acceptors and donors together with the structure of the triple gene is shown in appendix table 3. Since Ests for rat Irgm3 and rat Irgm2, not splice into triple tandem, have been reported (Ests for rat Irgm2 CO388297, CB544546, CO566274 and for rat Irgm3 CK841941,CK841941). It is reasonable to consider the triple tandem formation simply as an alternative splicing form of rat Irgm3 and Irgm2. However, no Est was detected for Irgb10 alone indicating that rat Irgb10 is only transcribed as first long exon of the triple tandem AY321344.

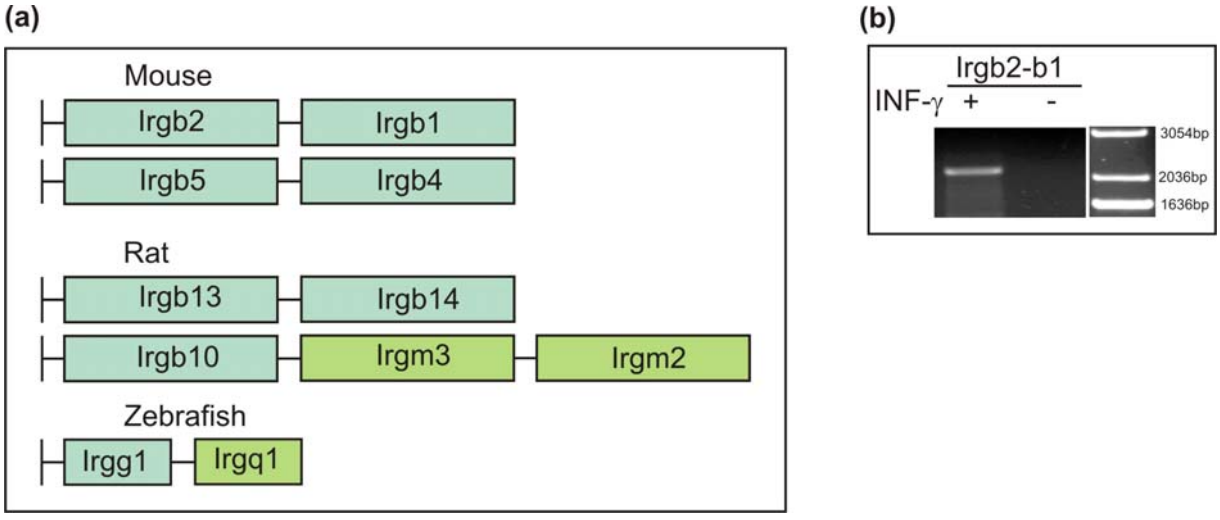


Figure 10. Triple and tandem gene formations in p47 GTPase family.
(a) All possible higher structure formations detected in mouse, rat and zebrafish in p47 GTPases are illustrated. Light blue shading indicates the GKS type GTPase, light green coloring indicates the GMS type or a Quasi GTPase. **(b)** An RT-PCR experiment showing interferon inducibility of mouse Irgb2-Irgb1 tandem. L929 cells were induced with 200 u/ml interferon γ for 24 hour (+) and uninduced (-).

Another tandem formation was detected in the zebrafish database forming irgg1-irgq1 which is located on chromosome 16. Like mouse tandems, irgg1and irgq1 are adjacent with the same polarity, with irgg1 positioned upstream of irgq1. There are two Ests available in databases,

BQ481364 and BQ481122. The tandem starts with irgg1 and splice into irgq1, however both GTPases code for the N-terminus and G-domain of p47 GTPases and irgq1 is considered to be a quasi GTPase which has valine (V) instead of lysine (K) in G1 motif (see below, Fig 10a and Maria Leptin personal communication)

III.I.3.The coding sequences of the p47 GTPases

A multiple alignment of the predicted translation products of the coding sequences of the 21 intact mouse p47 GTPase genes is shown in fig 7. Superimposed on the alignment is the known secondary structure of IIGP1 derived from the recently determined high-resolution crystal structure (Ghosh et al., 2004). The full alignment confirms a number of major features already apparent from the previously published alignment of six family members fig 5. The proteins are largely co-linear, with minor insertions or deletions. The GTP binding domain is rather strongly conserved in all proteins, with key elements for nucleotide binding being highly conserved. The previously noted abnormal methionine in the G1 motif (GMS instead of GKS) is found only in the three GMS proteins previously described. The many new genes described here are all of the conventional GKS type. Outside the nucleotide binding sites there is considerable sequence variation, especially in the C-terminal region, interspersed between highly conserved features common to the entire family. From the crystal structure of IIGP1, it can be tentatively predicted that the most highly divergent regions in the alignment correspond to extended loops between helical regions which vary in length. However confirmation for this interpretation depends on further structural information for other members of this family. A majority of the proteins, including all chromosome 18 gene products and some of chromosome 11 gene products Irgb10, Irgb2, Irgb5, Irgb9 carry the N-terminal myristoylation signal MGxxxS. It has been documented that IIGP1 is indeed myristoylated in cells, and, as expected, favors binding of the protein to membranes (Uthaiah, 2002), (Martens et al., 2004). It is therefore predictable that the putative myristoyl motifs of the other gene products may be active. No other membrane attachment sequences or lipid modification motifs are apparent elsewhere in the sequences, despite the documented attachment of several of these proteins to membranes. Several of the new gene products have C-terminal extensions up to about 65 residues compared with the canonical IIGP1 sequence. This is the case for the group of Irgb2, Irgb5, Irgb7ψ, Irgb9 proteins as well as for Irgc. However, C-terminal extension of Irgb2 and Irgb5 are largely excluded from the tandem sequences because of the splicing on the glutamine (Q) a.a residue (indicated as blue in Fig. 4).

III.I.4.Identification of interferon response elements and characterization of the putative promoter of mouse p47 GTPase genes.

The p47 GTPases are regulated by interferon gamma (Boehm et al., 1998). Therefore, it is essential to know the signature of the promoter elements used in the upregulation of these GTPases at the transcriptional level. The basis for interferon-inducible expression of the p47 GTPases has been investigated in a reporter assay only for Irgd (IRG47) (Gilly et al., 1996). In this study, Gilly et al., identified a classical ISRE sequence, upstream of the putative transcription start point. My analysis explored the generality of this observation, not just for the five other previously defined p47 genes but also for all the known transcribed p47 genes (Boehm et al., 1998). Fig 11 summarizes the essential findings superimposed on the genomic structure. The analysis indicated that there is another putative promoter region exists for irgd, in addition to that found by Gilly. It is also identified that there exist two promoter regions for IIGP1. Both promoters are used apparently in all tissues in which IIGP1 is expressed except the liver. In liver basal activity of Irga6(p2) is significantly higher than the Irga6(p1) ((Parvanova, 2005) and Jia Zeng personal communication). All known transcribed p47 genes possess the interferon inducible signature motifs, ISRE and GAS elements in characteristic clusters. Both the putative promoters of Irga6 and Irgd have intact interferon-inducible elements. Interestingly, infection with *L. monocytogenes* experiments revealed that Irga6(p1) driven expression is strongly upregulated while Irga6(p2) showed no or slight level of increase in upregulation in liver, spleen and lung (Parvanova, 2005). The positions relative to the putative transcription start site and the sequences as well as orientations of these elements are itemized in table 3. No systematic differences were apparent between the interferon inducible elements of any of the p47 genes except for Irgc. A more detailed search was done 10 kb upstream of the putative transcription start, also failed to reveal either clustered or isolated ISRE or GAS elements in the putative promoter region.

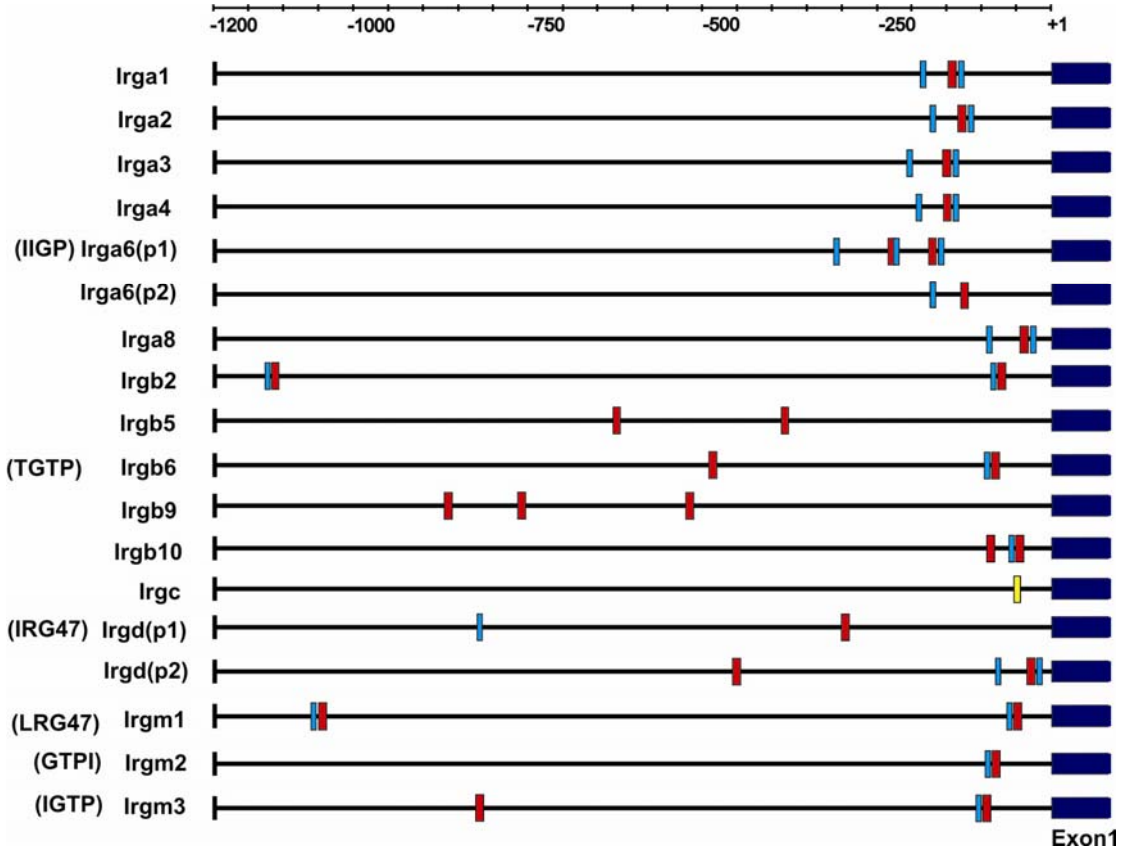


Figure 11. Promoter structure of mouse Irg GTPases.
Interferon response elements in the promoter regions of mouse Irg genes. γ-Activated sequences (GAS; pale blue blocks) and interferon-stimulated response element (ISRE; red blocks) sequences were identified in the promoters shown in panel a (also see Additional data file 7). Dark blue blocks downstream of each promoter represent the most 5' exon. The yellow block identifies a putative Sox1 transcription factor binding site in the proximal promoter region of Irgc. The scale bar is measured in base pairs from the first base of the 5' exon. Please note that Irga1 and Irga2 is the same promoter.

These data strongly indicated that interferon response elements for 14 uncharacterized p47 genes, and all except Irgc might therefore be inducible by interferon. To validate the importance of the identified interferon response elements, RT-PCR analysis was carried out. L929 cells were either stimulated or not stimulated with interferon γ (200 U/ml) for 24 hours and the results of induction were analyzed by RT-PCR. Of the 14 new p47 genes, eight of them showed clearly inducible transcription (Fig. 12a). As anticipated from the promoter analysis, Irgc showed no induction in fibroblasts, and in mice infected with *Listeria monocytogenes* (Christophe Rohde personal communication). Interestingly, there was lack of interferon-inducible transcription of Irga5 even though it shows perfect interferon-inducible upstream elements correctly positioned relative to the putative transcriptional start.

Gene name	Distance	GAS	Distance	ISRE
Irga1	-133	GTTCTTGGAA	-148	AGTTTCACTTTCCT(+)
	-188	CTTCTTTGAA		
Irga2	-119	GTTCTTGGAA	-134	AGTTTCACTTTCCT(+)
	-174	CTTCTTTGAA		
Irga3	-142	CTTCTTTGAA	-156	TGTTTCACTTTCAT(+)
	-207	TTTCTGCCAA		
Irga4	-141	CTTCTTTGAA	-156	TGTTTCACTTTCAT(+)
	-184	GTTTCTGGAA		
*Irga6(p1)**	-162	CTTCTTTGAA	-176	TGTTTCACTTTCAT(+)
	-226	TTTCTTGCAA	-235	CCTTTCTCTTTCTG(+)
	-312	GTTCCATTAA		
*Irga6(p2)**	-170	CTTCTTAGAA	-130	AGTTTCACTTTCCT(+)
Irga8	-30	CTTCTTTGAA	-45	GGTTTCACTTTCAT(+)
	-93	TTTCTGCCAA		
Irgb2	-87	TTTCCAGGAA	-77	AGAAAGTGAAACCT
Irgb4			-381	AGAAAGAGAAAGAC
			-627	TCAAAGAGAAAGTT
Irgb6	-96	TTTCCAGGAA	-86	CGAAACCGAAACCT
Irgb9			-223	AGAAAGAGAAAGAA
			-562	TCAAAGAGAAAGTT
			-665	TCAAAGAGAAAGAC
Irgb10	-61	ATTACTGGAA	-47	ACTTTCAGTTTCAC(+)
			-93	GCTTTCAGTTTCT (+)
*Irgd(p1)**	-821	TTTCTGTGAA	-301	ACTTTCTCTTTGAA(+)
*Irgd(p2)**	-21	TTTCCTGCAA	-35	AGTTTCACTTTTGT(+)
	-80	TTTCCTGGAA		
Irgm1	-64	TTTCAAGAAA	-54	AGAAACCGAAACTG
	-1061	TTTCCGGTAA	-1050	AGAAAGAGAAAGCC
Irgm2	-95	TTTCCAGGAA	-85	TGAAACTGAAAGCT
Irgm3	-108	TTTCTAGGAA	-98	TGAAACTGAAAGCT
			-825	TGAAAATGAAAGAC

Table 3: ISRE (Interferon stimulated Response Element) and GAS (Gamma activated sequences) elements of mouse Irg family genes.
Values in the distance column denote the position of ISRE and GAS element relative to the putative transcription start site. Black and gray shading indicates optimal and sub-optimal binding sites respectively. *(p1) alternative upstream promoter, (p2) alternative downstream promoter. ISRE and GAS elements marked as (+) have the same orientation relative to the putative transcription start site.

No additional elements such as an NFkB site which is frequently associated with the ISRE/GAS motifs were found. However the ISRE and GAS sites described in table 3 showed internal variation suggesting that they were not recently derived from a common ancestor. The relative positions of the GAS and ISRE elements varied from promoter to promoter and moreover both sites were not consistently present in all elements and the relative orientations of both components were variable.

Furthermore, to compare the number of fold induction of p47 GTPases by IFN γ, real time PCR was carried out. Classical p47 GTPases like Irga6 (IIGP1) and Irgm1 (LRG47), which have been characterized by in vivo and in vitro methods, were selected. Irgm1 shows 50 and 23 fold inducible transcription upon stimulation by IFN γ and β respectively. In contrast, Irga6 showed 215 and 23 fold induction by IFN γ and β respectively.

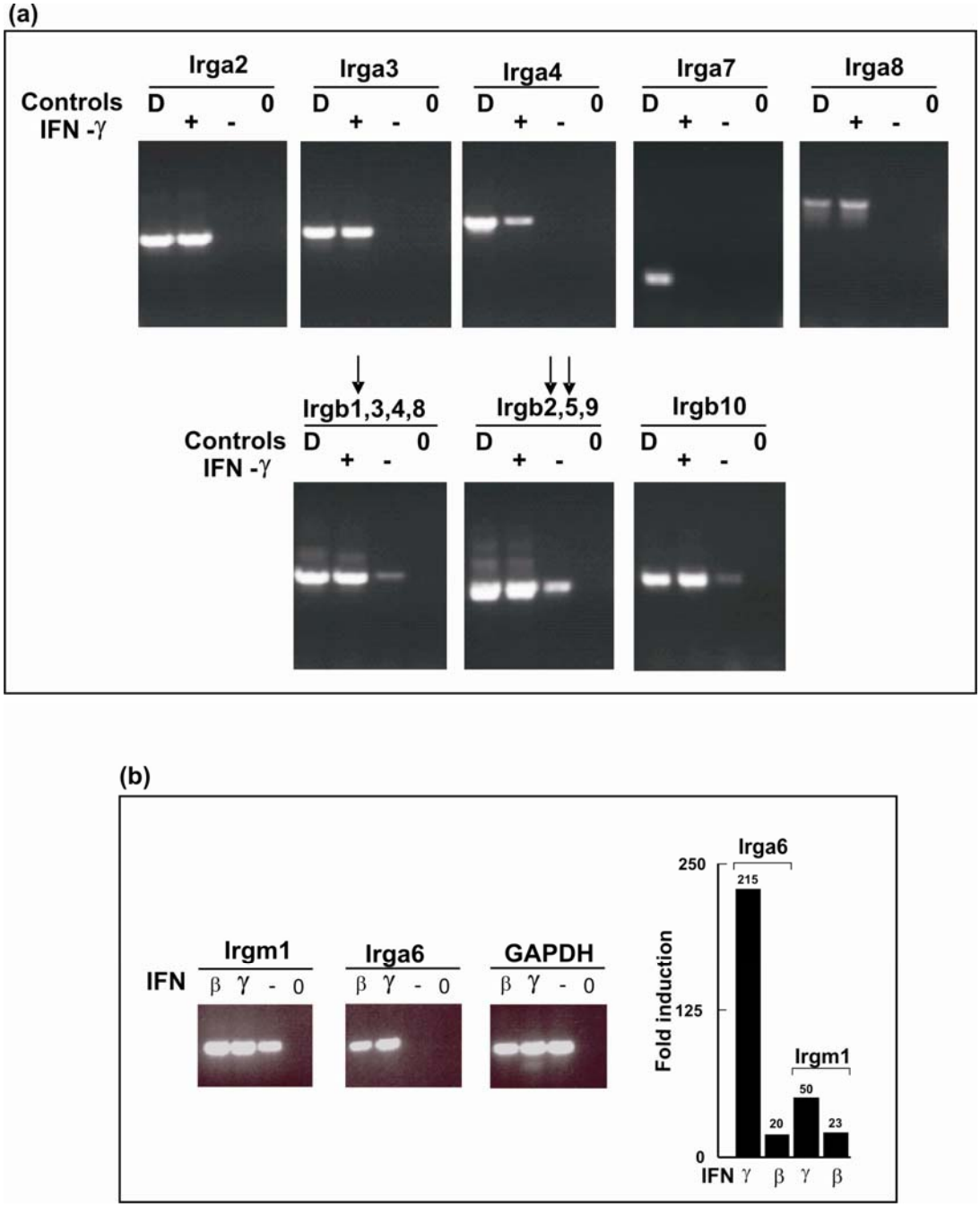


Figure 12. Interferon responsiveness of mouse and human p47 (IRG) GTPase.
(a) IFN γ inducibility of eight newly identified *Irg* genes. Induction was performed for 24 hours with IFN γ in L929 fibroblasts and was detected by RT-PCR. D refers to a positive control genomic DNA template; O refers to a negative control of the same genomic template after DNAse1 treatment; and + and – refer to RT-PCR on DNAse1-treated RNA templates from IFN-γ-induced and IFN-γ-noninduced cells, respectively. The sibling genes of the Irgb series could not be individually amplified because of their close sequence similarity. The identities of the amplified genes responding to interferon induction, indicated by vertical arrows, were subsequently established by sequencing of multiple clones from the PCR product. **(b)** Real-Time PCR analysis of the induction of Irga6 and Irgm1 in L929 fibroblasts induced for 24 hours with IFN-γ or β (also see (Boehm et al., 1998)). Demonstration of Interferon γ and β induction of Irga6 and Irgm1 in L929 fibroblasts, GAPDH was used as positive control (left). O refers to a negative control of the RT-PCR. γ, β and – refer to RT-PCR on DNAse1-treated RNA templates from IFN-γ/IFN-β induced and noninduced L929 cells, respectively. The detected induction ratio for Irga6 and Irgm1 by real-time PCR are illustrated (right). Numbers on the top of the box indicate the exact value of fold induction. Real-Time PCR was normalized using GAPDH. PGEMT-Easy containing ORF of Irga6 and Irgm1 was used as a reference for detection of the copy number of cDNA.

The reason for detected difference between Irga6 and Irgm1 in induction is probably due to chromosomal distribution of *Irg genes* (Fig 12b). It is detected that the promoters of the *Irg* genes, localized to the chromosome 11, have generally higher basal level of activity than the promoters of the *Irg* genes localized to chromosome 18 (Fig 12 and see above).

III.I.5.Identification of interferon response elements and characterization of the putative promoter of fish p47 GTPase genes.

The identified fish and dog p47 GTPases were analyzed for either interferon inducibility or existence of ISRE and GAS elements. Among the seven identified full-length dog p47 GTPases, four of them showed clearly inducible transcription upon IFN-γ stimulation in cell culture (see below). Preliminary analysis indicates that indeed ISRE and GAS elements are exist in the promoter regions of fish p47 GTPases. This is true especially in the case of irge3 and irge4 of which expression are derived from the same promoter like Irga1 and Irga2 (Fig 13 and 11). The fish p47 GTPases, irge3 and irge4 have perfect ISRE elements in the putative promoter region. The putative promoter region is identified using the Ests AW233145, CK142408 and analyzed according to supplementary analysis described in material and methods (Table 4). Identified putative promoter region of irge3 and irge4 is probably the representative of other irge like p47 GTPases because it has microsatellite repeats in its promoter region and the microsatellite spreads through *Danio rerio* BAC sequence (AL935330). The pattern of distribution of microsatellites is consistent with the distribution of fish *irge* genes which is located on the same BAC indicating that multiple genomic duplication events were responsible for generation of new *irg* genes with their promoter regions. Further analysis of promoter region of other fish p47 genes revealed that indeed p47 GTPases in fish have ISRE and GAS sites in their promoter region (Maria Leptin personal communication).

These properties strongly suggested that the association of the interferon-inducible elements with the p47 GTPase genes is old and their sequences are retained in position subsequently and are maintained in a working order by natural selection for a considerable period of time against the disruptive forces of spontaneous genome evolution.

Gene name	Species	Distance	ISRE	Reference
s HLA-A3	*H. sapiens*		AGAAA-AGAAACT	(Friedman and Stark, 1985)
s 2',5' AS	*H. sapiens*	-88	AGAAA-CGAAACC	(Benech et al., 1987)
as2',5' AS	*H. sapiens*	-140	GGAAACTGAAACT	(Floyd-Smith et al., 1999)
s Isg 20	*H. sapiens*	-39	AGAAACTGAAACA	(Gongora et al., 2000)
s Isg 15	*H. sapiens*	-95	GGAAACCGAAACT	(Reich et al., 1987)
s Isg 54	*H. sapiens*	-91	GGAAAGTGAAACC	(Reich et al., 1987)
s IFNα1	*H. sapiens*	-73	AGAAATGGAAACT	(Ryals et al., 1985)
s PKR	*H. sapiens*		GGAAAACGAAACT	(Kuhen and Samuel, 1997)
as MxA	*H. sapiens*	-91	AGAAA-CGAAACC	(Chang et al., 1991)
s PKR	*M. musculus*		GGAAAACGAAACA	(Tanaka and Samuel, 1994)
as Mx1	*M. musculus*	-120	AGAAA-CGAAACT	(Hug et al., 1988)
as Mx1	*G. gallus*	-50	AGAAA-CGAAACT	(Schumacher et al., 1994)
s Mx1	*O. mykiss*	-88	TGAAAGTGAAACA	(Collet and Secombes, 2001)
s Mx1	*D. rerio*		AGAAA-TGAAACT	(Altmann et al., 2004)
as Irga6(p1)*	*M. musculus*	-176	TGAAAGTGAAACA	Present study
as Irga6(p2)*	*M. musculus*	-130	GGAAAGTGAAACT	Present study
s Irgb6	*M. musculus*	-86	CGAAACCGAAACC	Present study
s Irgm1	*M. musculus*	-54	AGAAACCGAAACT	Present study
s Irgm2	*M. musculus*	-85	TGAAACTGAAAGC	Present study
s Irgm3	*M. musculus*	-98	TGAAACTGAAAGC	Present study
as Irge3, irge4	*D. rerio*	-76	GGAAA-CGAAACT	Present study

Table 4: Comparison of ISRE (Interferon Stimulated Response Element) elements representing Irg family genes with the known ISRE elements of other IFN inducible genes.
Values in the distance column denote the position of ISRE element relative to the putative transcription start site. Black and gray shading indicates optimal and suboptimal binding site respectively. *(p1) alternative upstream promoter, (p2) alternative downstream promoter. ISRE and GAS elements marked as (as) antisense, (s) sense. In the presence of two or more ISRE element in the respective promoter region, the one that is closest to the transcription start site has been used. The table was originally prepared by (Collet and Secombes, 2001) and updated using recent reports.

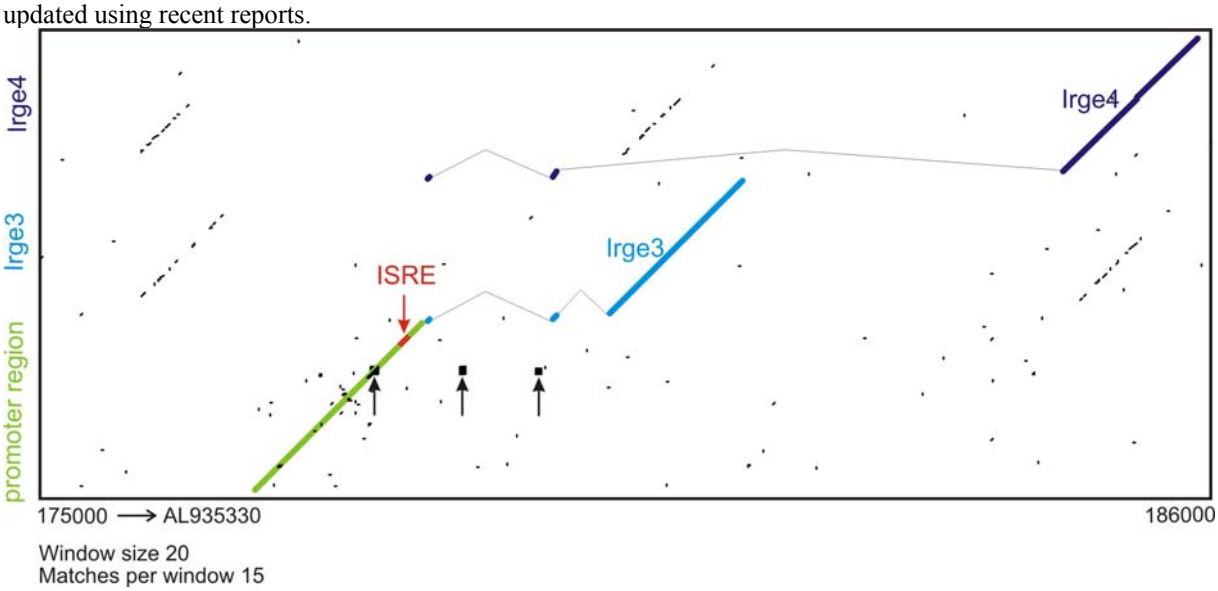


Figure 13. Promoter and genomic structure of irge3 and irge4
Dot plot matrix analysis using irge genes (*irge3* vertical-represented by light blue and *irge4* vertical represented by dark blue) and Danio BAC sequence (AL936330) matrix covers the 11000bp (from 175000 to 186000) of AL936330. Putative promoter region was identified using the 5'Ests (AW233145 for Irge3 and CK142408 for Irge4) (highlighted in green). Identified ISRE sequence on the promoter region of irge3/4 was indicated with red arrow. Black arrows indicate the microsatellites repeats.

III.I.6.The p47 GTPases in other rodents

Using a combination of screens on available databases, evolutionary analysis of p47 GTPases was extended to other rodentia species. Analysis was carried out using bioinformaticc approaches for the Czech II mouse strain (*Mus musculus musculus*) and Rat (*Rattus norvegicus*).

Either blastn or tblastn searches in NCBI blast server were used yielding 82 Ests from Czech II mouse showing significant homology to p47 GTPase. All collected Ests were used to generate contigs. A total of ten contigs was assembled using the supplementary analysis described in Material and Methods. Further search analysis was performed to confirm contigs association to p47 GTPases and putative full length transcript was extracted whenever possible. Nucleotide sequences of these genes were edited to get putative open reading frames, and were aligned (Fig. 14a and see below). 10 representative of the p47 GTPase family were recovered from the Czech II Est database, with a complete ORF. Irgm1 could only be partially constructed from collected Ests by contig generation and alignment of all p47 GTPase in Czech II mouse is shown in appendix fig 2. The phylogenetic analysis showed that there are sequence variations indicating recent diversification. In the phylogeny the branch containing Irga9 and Irga10 represent recently duplicated version of p47 GTPase family in Czech II mouse (Fig 14a). Irga8 is encoding full length p47 GTPase in Czech II mouse whereas it is truncated in C57BL/6 mice (see above). It would be of interest to elucidate whether there are any patterns of polymorphism of p47 GTPases in Czech II mouse leading to diversification of the family members by positive selection.

The rat genome was also screened for analyzing p47 GTPase homologue genes using the available database for Norway Rat (*Rattus norvegicus*). Fifteen p47 genes were recovered from the rat genome of which two are incapable of coding full length p47 GTPases (see above). Therefore it is concluded that *Irga14* and *Irga16* are pseudogenes since they do not encode full length p47 GTPases and have accumulated multiple null mutations. In contrast previously reported pseudogenes in mouse Irga5 and truncated p47 GTPase Irgb10, which is encoding only the G-domain of classical p47 GTPases, are encoding functional full length p47 GTPases in rat (Fig 14b). Multiple alignments of mouse p47 GTPase with their rat homolog show that every feature of p47 GTPases are also present in rat (Appendix Fig 3). Topology of the phylogeny of rat p47 GTPases generated together with mouse homologs reveals that diversity of p47 GTPases is protected here in rat as well. Each branch of phylogeny has at least one member of the rat p47 GTPase. Additionally, the branch containing Irga11, Irga12

and Irga13 (IIGP1), further diversified in the rat genome (Fig 14b) suggesting that diversification of the Irga genes probably is expanded by recent genomic duplication (see above).

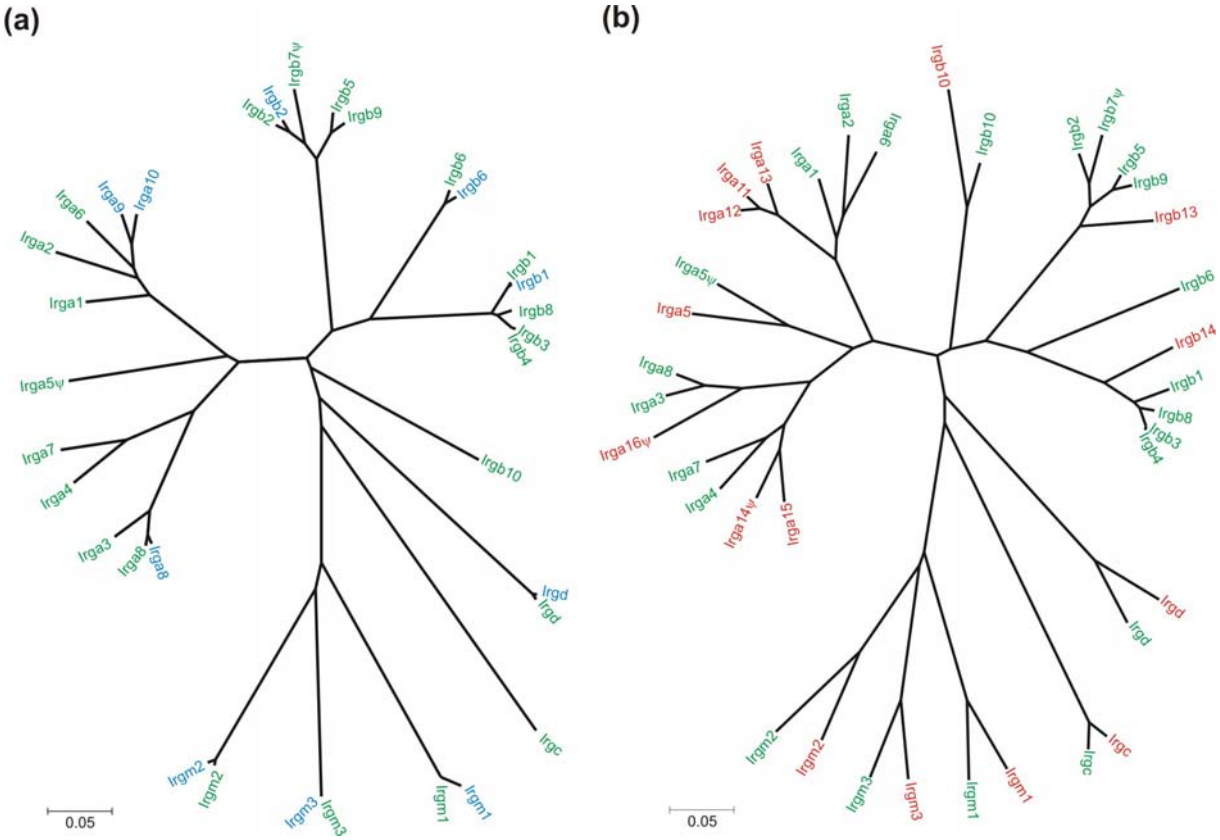


Figure 14. Phylogeny of other rodents (Czech II and rat) with C57BL/6 mice Irg GTPases
(a) Phylogeny of Czech II (blue) and C57BL/6 mice (green). **(b)** Phylogeny of rat (red) and C57BL/6 mice (green). For both phylogenic tree constructions, Nj tree based on nucleotide sequences was generated by using Mega3.1. The nucleotide and amino acid sequences themselves can be obtained in the p47 (IRG) GTPase database from our laboratory website (http://www.genetik.uni-koeln.de/groups/Howard/index.html).

Ests were identified for all rat p47 GTPases described above, except for Irgb14 and Irgb13 showing that family is indeed actively transcribed in rat as well. As mentioned above, triple gene formation was detected in the rat genome at transcriptional level. There is a special mRNA (AY321344) exist in the rat database encoding triple p47 GTPases in tandem which was reported as liver specific regeneration gene (see above and appendix table 3).

III.I.7.The p47 GTPase genes of the human genome.

An extensive analysis of the human genome databases was initiated to identify the p47 GTPase gene family members in humans. By both transcriptome and genome analysis only two sequences, both transcribed, corresponding to p47 GTPases were found, on chromosome 19, and chromosome 5 respectively. Analysis of these two sequences showed that, *IRGC* is closely homologous at both nucleotide and amino acid level, to the isolated mouse gene, *Irgc*.

The second sequence, *IRGM*, encodes a G-domain of p47 GTPases, which begins downstream of the typical start sites in the mouse p47 GTPases and terminates in the region of the α-helixH in IIGP1. 3‘of this point all recognizable homology at nucleotide or amino acid level in all five reading frames was lost. By a number of criteria, including the defining methionine, the *IRGM* transcribed gene fragments is a human homologue of the GMS subfamily of p47 GTPases (Fig 17). By exploring the human and mouse synteny maps, it was possible to locate the syntenic cluster containing mouse *Irgc*, accurately to an identical syntenic cluster on human chromosome 19. The human IRGC shows more than 90% identity at the amino acid level and more than 85% at the nucleotide level with the syntenic human gene. Thus, we concluded that *IRGC* gene in human is a true orthologue of the *irgc* gene in mouse.

Using various syntenic loci, it was possible to map unambiguously the region in the human genome corresponding to both mouse p47 GTPase clusters to the proximal long arm of human chromosome 5. The mouse chromosome 11 cluster, itself divided by a 10 Mb gap, is also divided in its syntenic relation to the human chromosome. The region corresponding to the 10 genes from Irgd (IRG-47) at one end Irgm1 (LRG-47) at the other is accurately located in a 30 kb interval between the two human marker loci HINT1 and TRIM7. The mouse chromosome 18 p47 GTPase cluster maps immediately centromeric to the human marker gene DCTN4 (Fig. 15). The synteny results strongly suggest that the interferon-inducible p47 GTPases were formerly encoded in a single cluster ancestral to the human chromosome 5 region. This ancestral block was subsequently broken down in the mouse lineage into two clusters located on chromosomes 11 and 18 respectively, while the p47 genes in the chromosome 5 cluster in the human species were progressively lost until the only trace of their former existence is the unique GMS fragment (see below).

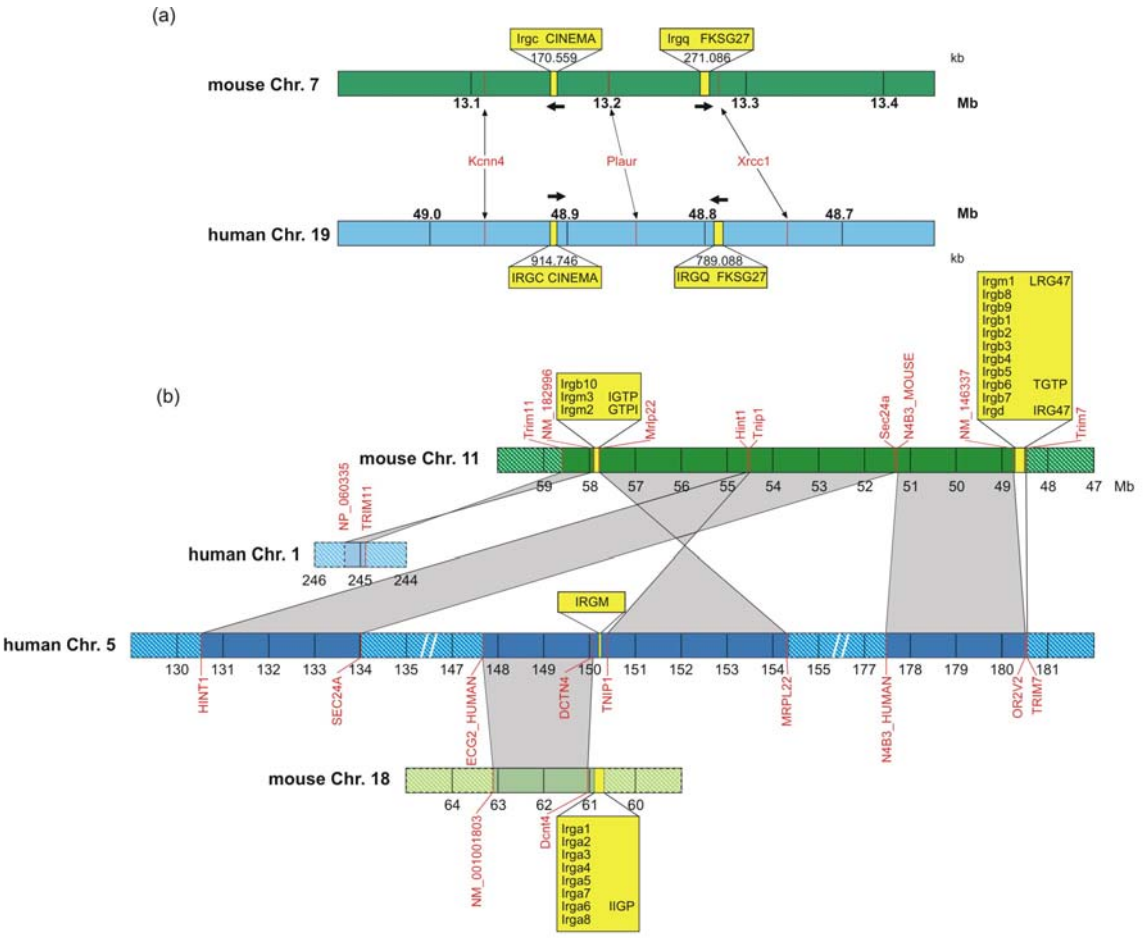


Figure 15. Synteny relationships between the human and mouse *IRG* genes
a. Synteny between mouse chromosome 7 and human chromosome 19 in the region of the *IRGC* and *IRGQ* genes. The figures indicate distances from the centromere in megabases. The locations of three further syntenic markers are given. Gene orientation is given by black arrows. **b.** Complex synteny relationship between human chromosome 5 and mouse chromosomes 11 and 18 in the regions containing the mouse *Irg* genes. Figures indicate distances from the centromere in megabases. The locations of *IRG* genes are shown in the yellow panels. Positions of diagnostic syntenic markers are also indicated. Syntenic blocks are given in full color, the rest is shaded. (courtesy of Julia Hunn)

III.I.8.The p47 GTPase genes of the dog genome.

Is the mouse (Order Rodentia) or the human (Order Primata) the exception? *IRG* genes in a third order of mammals, the Carnivora was screened. Totally, nine *IRG* genes from the public genome database of the dog, *Canis familiaris*, were recovered, (Fig 17, 18 and Appendix Fig. 4). Of these, one (AACN010088820) is a pseudogene by a number of criteria, another is clearly a dog *IRGC*, while the partial sequence (AACN010048557) is novel but most closely related to *IRGC*. The remainder assort into segments of the phylogeny already established for the interferon-inducible mouse *IRG* genes. Both GMS and GKS genes are represented and are inducible by interferon in dog MDCK II epithelial cells (Fig 16). The three dog GMS genes

seem to have diversified independently from the mouse GMS genes as represented in main vertebrate phylogeny (Fig 18). As in man and mouse, dog *IRGC* gene was not induced by IFN-γ. At least in dog, therefore, the absence of the interferon inducible p47 GTPase Ests in the databases can reflect a tighter control of transcription than in mouse, and the same argument could also be used for the other mammalian groups. Overall, the *IRG* gene status of dog clearly resembles that of mouse rather than that of human.

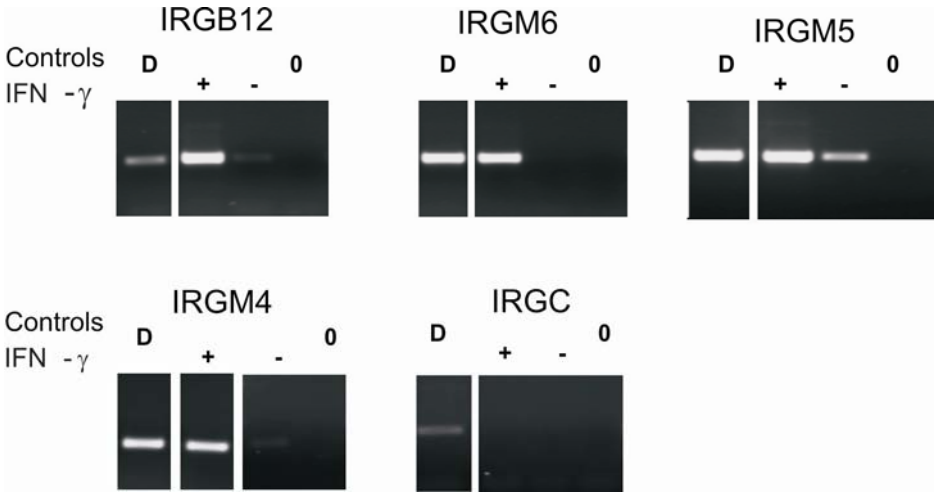


Figure 16. Inducibility of Dog *(Canis familiaris)* GTPases
Epithelial MDCK II cells were induced (+) or not induced (-) with 10 ng/ml dog interferon for 24 hours, D refers to 30 ng of genomic DNA as positive control, 0 refers to no DNA as negative control.

III.I.9.The p47 GTPase family in other vertebrates.

The public databases (ENSEMBLE and NCBI) for homologues of the p47 GTPases was screened in other taxonomic groups. Among the other mammals, p47 GTPase like genes have been identified in pig, hamster, and cow (Appendix Table 4). Examination of these sequences reveals that they have indeed characteristic features of p47 GTPases. Both pig and dog have Cinema (IRGC) as well as a Cinema like p47 GTPase. The degree of divergence of this second sequence from pig and dog Cinema (IRGC) suggests a relatively old duplication rather than a recent event. When mammalian lineage is considered, these results suggest that the absence of inducible p47 GTPases seen in man might be an unique case outside the murine rodents.

The p47 GTPases are present in several non-mammalian vertebrates (Appendix Table 1). There is p47 GTPase-like sequence available for Xenopus (*Xenopus tropicalis*). However, no p47 GTPase gene was detected for chicken (*Gallus gallus*) (Appendix Table 4). In addition, the completion of the two of the ray finned fish genomes has allowed us for a detailed analysis in these fish (Zebrafish, fugu, Tetraodon). The alignment given in fig. 17 shows conclusively that these are p47 GTPase genes, with all the characteristic sequence

features identified in the mammalian representatives present. Fish p47 GTPases fall in to two clades (f and e) in vertebrate phylogeny (Fig. 18) showing that diversification of the p47 GTPase family is probably expanded by an early genomic duplication event (Hoegg and Meyer, 2005) (Christoffels et al., 2004). Including quasi GTPases, 14 intact members of the p47 GTPase family were detected in zebrafish. However, the family was represented only with 2 members in Fugu and Tetraodon, respectively. No members of the GMS subfamily are present in these fish genomes. Exceptionally, the Tetraodon, fugu and zebrafish (only for Irgf) genes appear to be divided by a short intron positioned as indicated by the blue in Fig. 17. This is inferred from the alignment of the sequences with mammalian sequences, the presence of stops in all reading frames in the putative introns except fugu, the positioning of perfect splice donor and acceptor sites and available Ests in databases (CA589084 for Fugu irgf5). They show no significantly greater similarity to the highly conserved mouse CINEMA (Irgc) gene than to the variety of interferon-inducible genes. Thus, it can be concluded that fish p47 GTPase family, its own evolutionary trajectory probably related to diversification after species-specific multiple genomic duplications resulting in different complexity of p47 GTPases in Fugu, Tetraodon and Danio.

III.I.10.The p47 GTPase genes in invertabrates.

It is possible to identify p47 GTPase-like genes outside the vertebrates. Although no homologue of p47 GTPases were detected in Drosophila, the results of the recent database search show that *C. elegans* has p47 GTPase like proteins; (C46E1.3) which is encoded as tandem and additional single gene (W09C5.2) (Appendix Table 1). It could be argued from the alignment that W09C52 is much closer to p47 GTPase then C46E1.3 especially in N-terminus (Appendix Fig 5). However, validity of these genes as a member of p47 GTPase remains to be answered. Phylogenetic and bioinformatic analysis is not enogh to link these genes to family of p47 GTPases. Biochemical or structural studies are necessary to clarify the validity of these genes as a member of p47 GTPases.

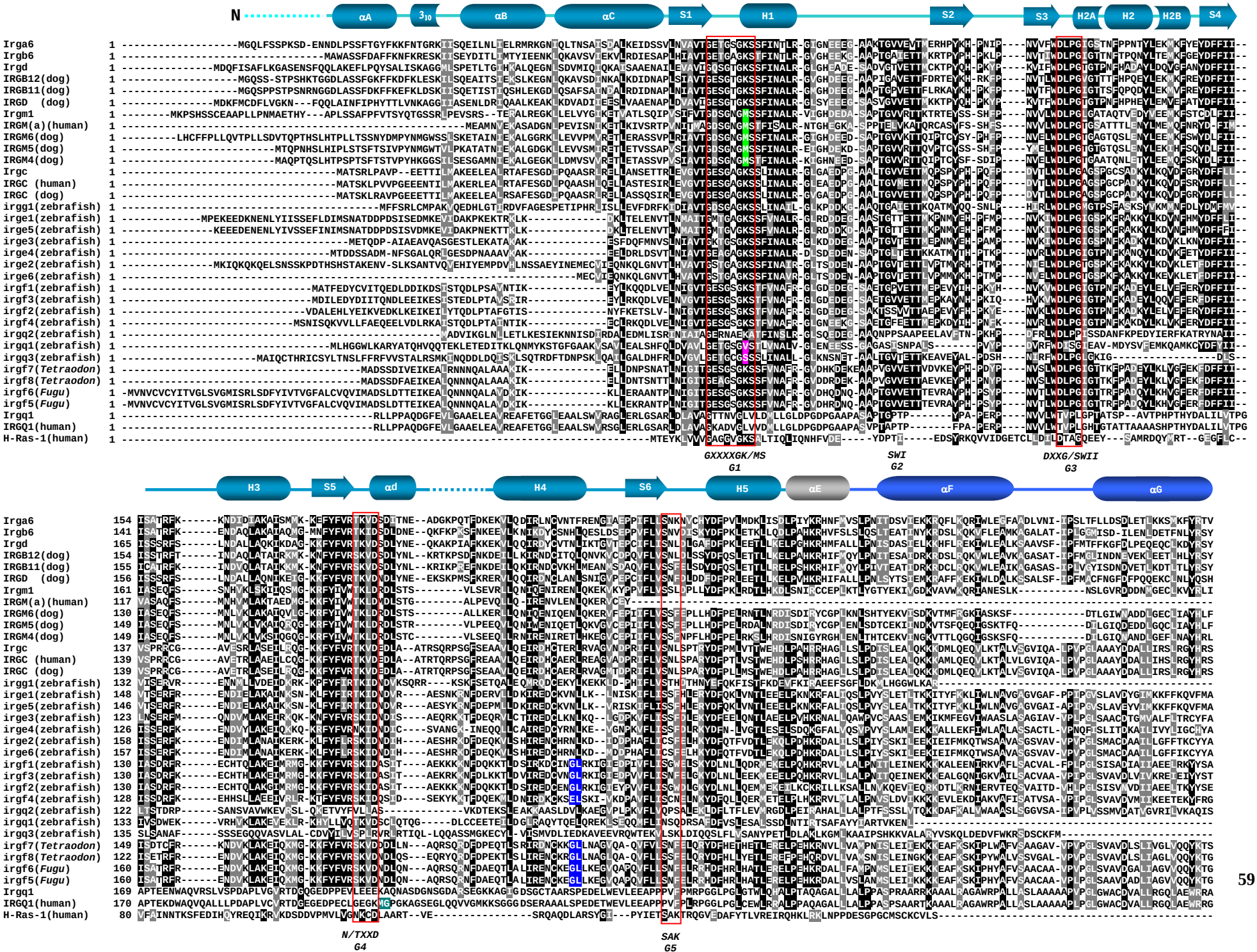
A series of 45-50 kDa GTPases of unknown function are recognizable in a number of cyanobacterial species, including common pathogens which show a plausible homology to the vertebrate GTPases in the G-domain. The G-domains of these enzymes are located within the protein at roughly the same position as in the p47 GTPases as a general characteristics of p47 GTPases. These observations raise the possibility that the vertebrate p47 GTPases may have been horizontally acquired from a microbial genome although it must immediately be conceded that no homology can be discerned outside the G domain however secondary

structure predictions analysis indicates that bacterial p47 GTPase like proteins have similar secondary structure to IIGP1 (Jonathan C. Howard personal communication).

III.I.11.IRG homologues with divergent nucleotide-binding regions: the quasi-GTPases

The mouse, human, xenopus and zebrafish genomes encode proteins homologous to the *IRG* GTPases but radically modified in the GTP-binding site. These modified GTPases, which are named here as "quasi IRG" proteins, thus IRGQ, have characteristic features of p47 GTPases. Human and mice contain a single *IRGQ* gene closely linked to IRGC. The zebrafish genome contains three *IRG* homologues with more or less modified GTP-binding motifs (*irgq1-q3)*, (Fig 17 and Fig 18). The homology of the fish irgq genes to *IRG* genes is stronger than that of human and mouse IRGQ genes but their function as GTPases is doubtful. *irgq1* is clustered on a single BAC clone with 4 apparently normal *irge* genes and immediately downstream of a truncated p47 gene, *irgg,* with which *irgq1* is transcribed as the C-terminal half of a tandem transcript (Maria Leptin personal comunication). Thus the hypothetical protein product would be a C-terminally truncated p47 GTPase, linked at its C-terminus to a similarly truncated p47 homologue probably without GTPase function (see above).

IRGQ sequences reveal their phylogenetic relationship to the IRG proteins, but are nevertheless more or less radically modified, primarily in the nucleotide binding site. In view of the substantial divergence between the *IRGQ* genes and functional p47 GTPases, it was unexpected not to find close homologues of the *Danio irgq* sequences in either the *Fugu* or *Tetraodon* genomes. The evolution and diversity of the *Danio irgq* genes is apparently linked to the evolution and diversity of the p47 GTPase family.


N
αA
3₁₀
αB
αC
S1
H1
S2
S3
H2A
H2
H2B
S4
GXXXXGK/MS
G1
SWI
G2
DXXG/SWII
G3
H3
S5
αd
H4
S6
H5
αE
αF
αG
N/TXXD
G4
SAK
G5
Irga6
Irgb6
Irgd
IRGB12(dog)
IRGB11(dog)
IRGD (dog)
Irgm1
IRGM(a)(human)
IRGM6(dog)
IRGM5(dog)
IRGM4(dog)
Irgc
IRGC (human)
IRGC (dog)
irgg1(zebrafish)
irge1(zebrafish)
irge5(zebrafish)
irge3(zebrafish)
irge4(zebrafish)
irge2(zebrafish)
irge6(zebrafish)
irgf1(zebrafish)
irgf3(zebrafish)
irgf2(zebrafish)
irgf4(zebrafish)
irgq2(zebrafish)
irgq1(zebrafish)
irgq3(zebrafish)
irgf7(Tetraodon)
irgf8(Tetraodon)
irgf6(Fugu)
irgf5(Fugu)
Irgq1
IRGQ1(human)
H-Ras-1(human)

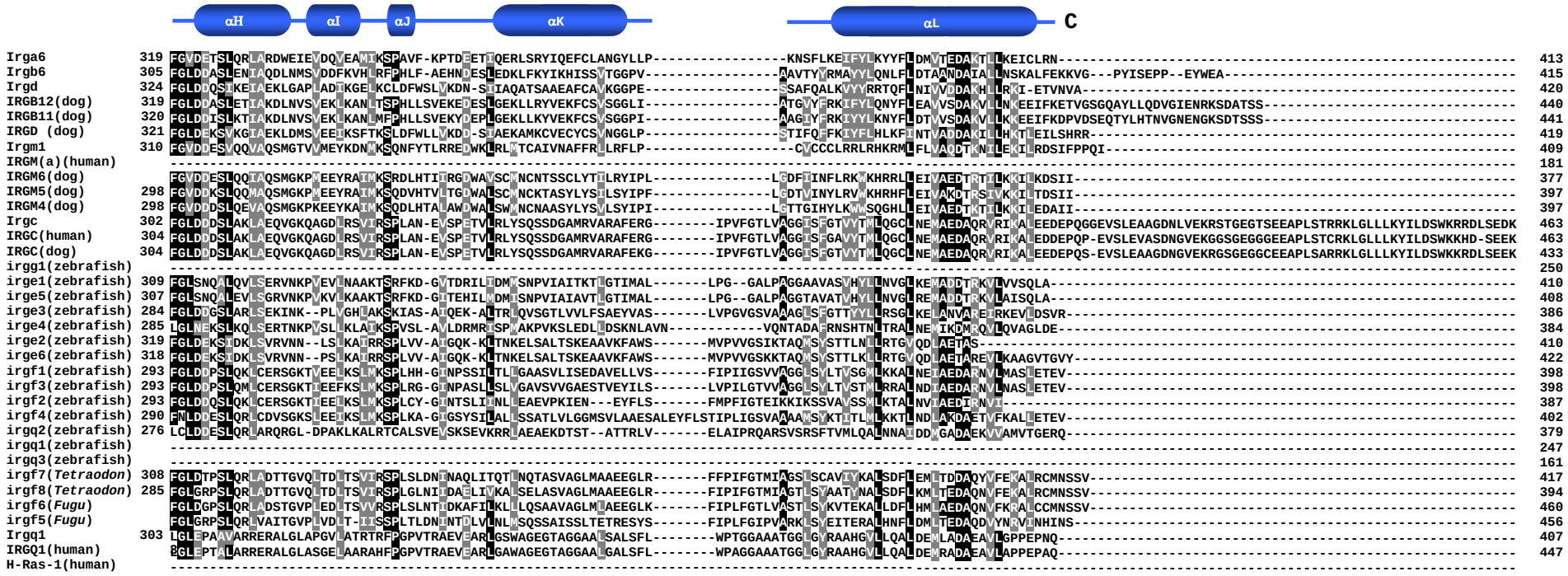


Figure 17. Extended alignment of the vertebrate IRG proteins.
Individual sequences are given in full and are labeled as in Figure 18. Unusual residues in the G1 motif are highlighted (M of the GMS proteins in green and two deviant residues in the zebrafish irgq sequences in pink). The essential structural relationship between IRG genes and quasi-IRG genes is apparent in the alignment despite the modified G-domains. For mouse and human IRGQ the long carboxylterminal coding exons that contain the p47 homology were used for the alignment. In human IRGQ the sequence NPKGESLKNAGGGGLENALSKGREKCSAGSQKAGSGEGP was removed from the alignment between positions 210 and 211 (highlighted in turquoise) to prevent extensive gap formation. The position of the intron present in pufferfish and zebrafish irgf genes is indicated by two adjacent residues highlighted in blue. Canonical GTPase motifs are indicated by red boxes. The nucleotide and amino acid sequences themselves can be obtained in the p47 (IRG) GTPase database from our laboratory website (http://www.genetik.uni-koeln.de/groups/Howard/index.htm).

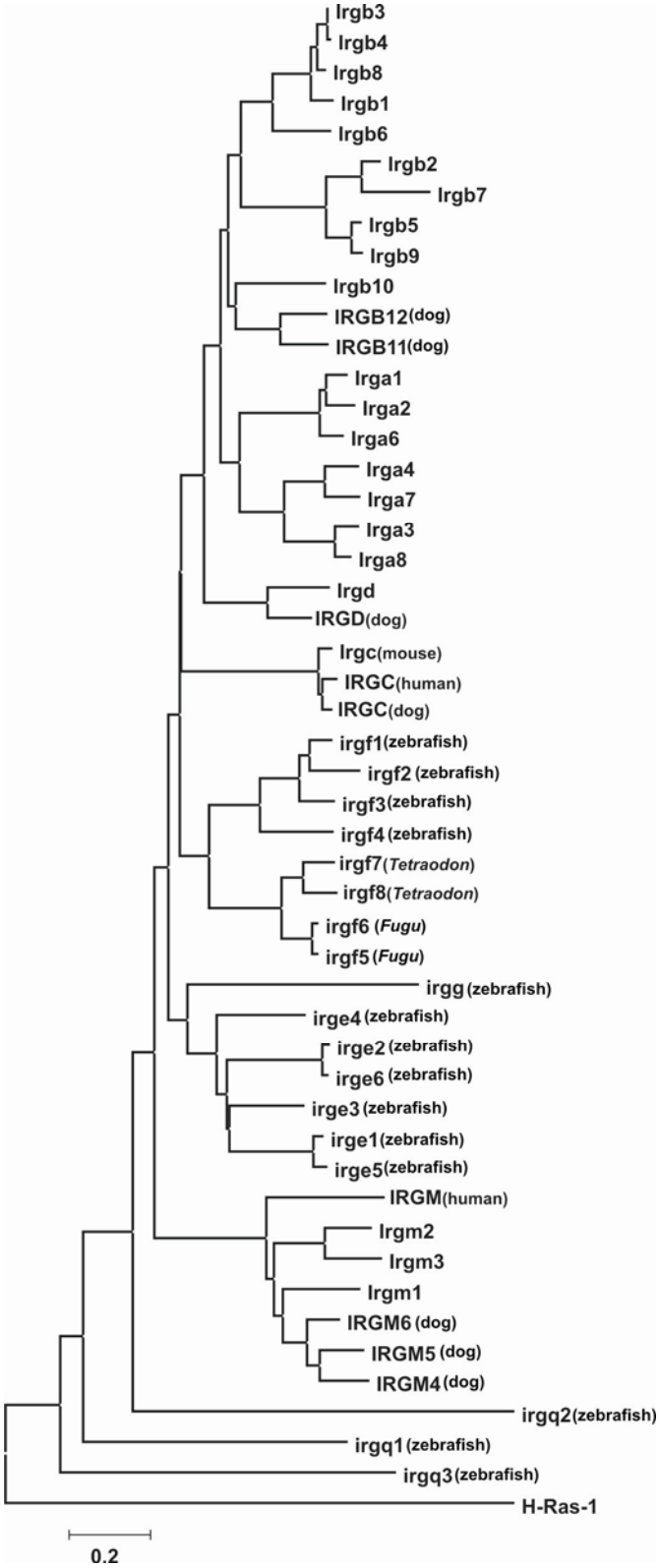


Figure 18. Extended phylogeny of the G domains of IRG and related proteins.
The phylogeny relates all of the IRG sequences described in this report and reveals the distinct clades on which the nomenclatural fine structure is based. All except the mouse sequences are labeled with the species of origin. Dog IRG sequences are found in the B, C, D and M clades, and human sequences only in clades C and M. The mouse and human quasi-IRG proteins, IRGQ (FKSG27), could not be included in the phylogeny because they are so deviant in the G-domain

III.I.12.Positive selection in the family of the p47 GTPases

The p47 GTPase family is a resistance mechanism to fight against intracellular pathogens. It is expected that genes involved in immunity will evolve faster to coadapt under the selective pressure generated by the pathogens which usually have fast evolving capacity. If the pairwise alignment of given sequences, which are closely related to each other, has more non-synonymous substitutions per site than synonymous substitutions (Ka/Ks>1), these genes are considered as fast evolving genes. It was a crucial question to answer whether the p47 GTPases are fast evolving genes or not? Therefore, the codon based selection test was employed for the estimation of synonymous (*Ks*) and non-synonymous (*Ka*) substitutions per site in the protein coding region of the p47 GTPases (Fig 19). The p47 GTPases within mouse or between mouse and rat were aligned in pairwise manner. The aligned sequences were edited to obtain correct ORFs, whenever possible. The protein coding sequences, which have correct pairwise alignments, were used for the estimation of synonymous (*Ks*) and non-synonymous (*Ka*) substitutions based on the methods established by (Comeron, 1999). All of the estimated Ka/Ks values are shown in figure 19. The selection test within the mouse was performed using the genes *Irga4, Irga7, Irga3, Irga8, Irga2, Irga6, Irgb1, Irgb4, Irgb2, Irgb5*. For the selection test between mouse and rat, the p47 genes *Irgm1 Irgm2, Irgm3, Irgd, Irgc, Irgb1, Irgb4, Irgb13, Irgb14, Irga4, Irga6, Irga15, Irga13* were used. Because these genes were only the genes have an intact full length pairwise alignment. Four (Irga4, Irga7, Irga8 and Irga3) of the p47 GTPases within the mouse were detected to be positively selected especially in the C-terminal region. Further analyzes were performed using the pairwise alignments between rat and mouse (Fig. 19b), Two (Irga4 (Irg15 in rat) and Irgm2) of the p47 GTPases were detected to be positively selected. It is worth noting that Irgc was detected to be evolving under purifying selection which is unique among the members of the p47 GTPase family (Fig. 19b). Micro-evolutionary analysis at the population level is necessary to reveal whether p47 GTPases are indeed fast evolving genes. If so, it will be of particular interest to know, which region of the individual protein preferentially positively selected and to which extant positive selection maintained the members of the p47 GTPases.

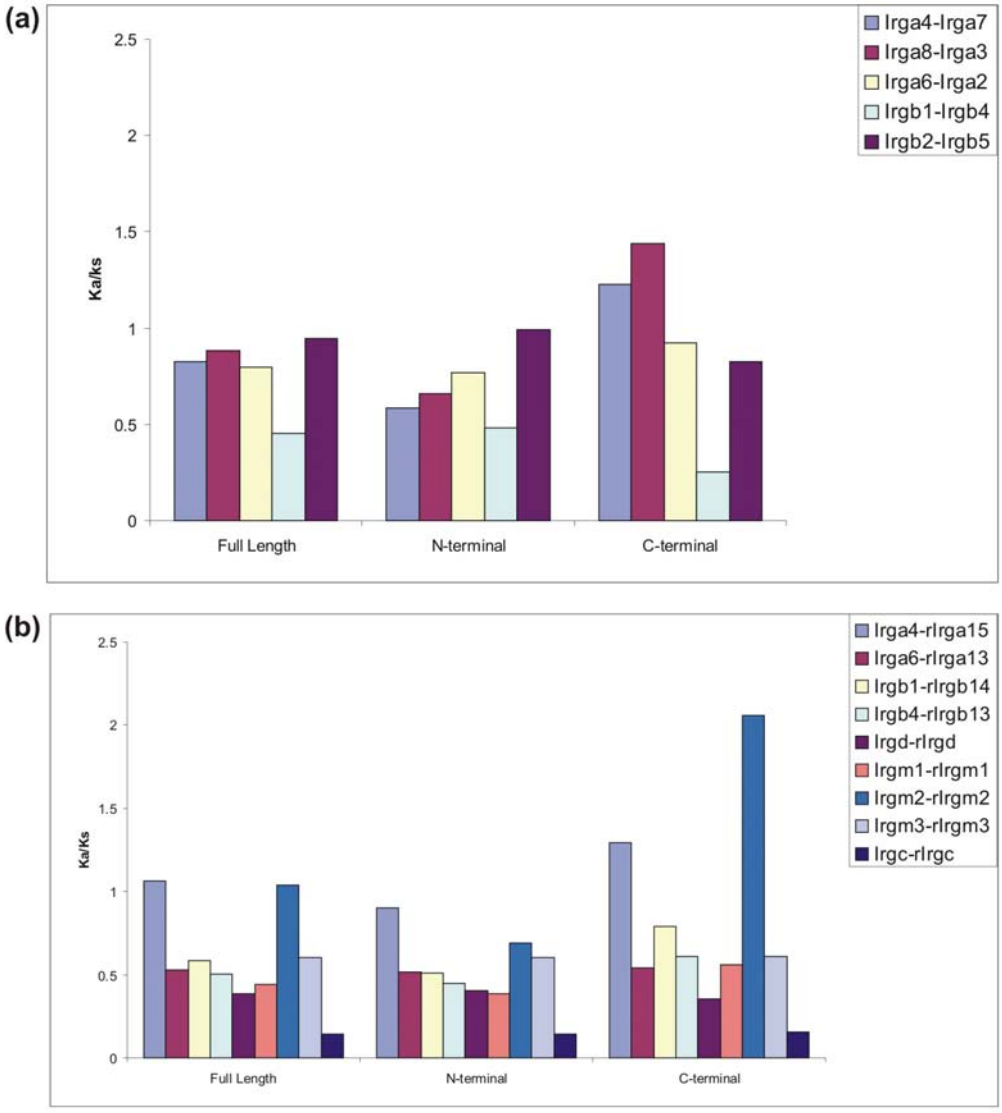


Figure 19. Codon based selection test for p47 GTPases
(a) Codon based selection test was performed using the ORF (Full length), N-terminal (1-275 a.a) and C- terminal (275-end of the respective sequence) region of close family members of p47 GTPases within the mouse. Deletions and insertions were removed to align sequences properly. **(b)** Codon based selection test was performed using the ORF (Full length), N-terminal (1-275 a.a) and C- terminal (275-end of the respective sequence) region of mouse and rat p47 GTPases. Deletions and insertions were removed to align sequences properly. Analysis was performed using the program K-Estimator 6.0 (Comeron, 1999) with multiple hits correction method Kimura-2 parameter.

III.II.RESULTS II

III.II.1.The human GMS fragment (IRGM).

IRGM is transcribed in unstimulated human tissue culture lines, Hela and GS293 cells with no increase after interferon induction (Fig. 20a). Polyadenylated transcripts of *IRGM* occur with five 3' splicing isoforms extending more than 30 kb 3' of the long coding exon. By a combination of Est and genomic database analysis, and 5'-3' RACE PCR from the coding region of the human GMS fragment, it was possible to clone different transcripts containing the human GMS fragment (Fig. 20b). The identity of amplified fragments with the human GMS fragment was confirmed by sequencing. Three Ests can be found in the public databases (BC038360, BC038359 and BI764111), and comparison of these Ests with amplified splice variants revealed that all of the Ests for IRGM in the databases are identical to the 3'splice variant, IRGMc. The IRGM coding ORF is located in long transcripts downsteam of a long putatively untranslated exon from the adjacent 5' genomic region, and upstream of a 3' region containing one or more exons derived from regions far 3' of the GMS fragment. As it is typical for the p47 GTPases (see above), the entire GMS ORF is encoded on a single exon. The shortest form of transcript, IRGM(a), reads through the splice-site immediately downstream of the ORF and terminates behind a polyadenylation signal sequence at the 5' end of the intron. The longer transcripts splice out of this region to one or more exons more than 30 kb downstream. In all cases, the transcripts are polyadenylated.

The transcript of the human GMS fragment thus has a highly unusual structure with its extended 5' untranslated region of more than 1000 nucleotides, and especially the presence of one or more exon-intron boundaries downstream of the putative termination codon in three of its five splice forms which are expected to lead to rapid RNA degradation via nonsense-mediated decay (Ohnishi et al., 2003; Singh and Lykke-Andersen, 2003; Wilkinson, 2005).

The 5' untranslated region of the GMS fragment transcripts is similar to the U5 region of an endogenous retroviral element (ERV9) repetitive element. The promoter region corresponds to the ERV9 U3 long terminal repeats (LTR) without interferon response elements. The difference in the expression level in different cell lines was consistent with the expectation of classical transcription profile of ERV9 promoters (Ling et al., 2002) (Fig. 20a) As noted above, transcripts (IRGM (b) and (c)) are easily detectable in unstimulated tissue culture cells and total RNA obtained from human tissues (Liver, Brain, Testis) (Fig. 20a). However, RT-PCR experiments using several different human cell lines (HepG2, Thp1, SW480, Primary fibroblasts

HS-27) failed to detect an induction of *IRGM* by IFN-γ (data not shown). The human lymphoblastoid cell line, T2, showed a 2-3 fold induction of IRGM after interferon treatment.

At the protein level the shortest isoform of IRGM, IRGM(a), is shorter than a canonical G-domain, due to truncation in the middle of β-strand six just before the G5 sequence motif which interacts with the guanine base of the bound nucleotide (Fig. 17 and Fig. 20b). The longer isoforms are terminated by short sequence extensions unrelated to any known GTPase domains.

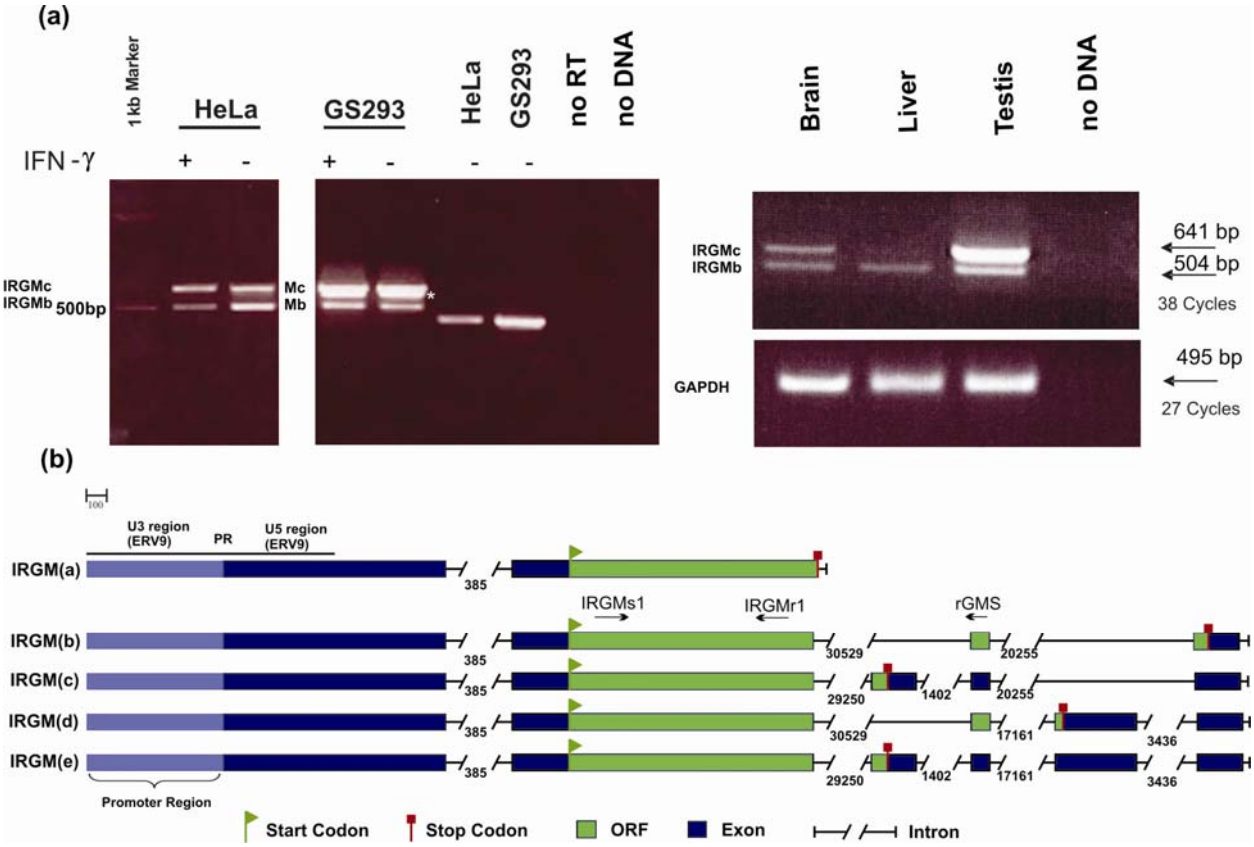


Figure 20. Structure and expression of the human IRGM gene.
(a) (left panels) RT-PCR analysis of the expression of IRGM in HeLa and GS293 cells**.** IRGM(b) and (c) splice variants were amplified simultaneously by the same primer pair (IRGMs1-rGMS). A different downstream primer (IRGMs1-r1), internal to all the 3' splice forms was used to show differences in the overall expression level of IRGM in the two cell lines. No RT a cDNA preparation without reverse transcriptase. The band immediately below the IRGMc band in GS293 cell material, indicated with an asterisk, is a nonspecific band amplified only in this cell line. The band was sequenced and is unrelated to IRGM. (right panel) Analysis of IRGM expression in human brain, liver and testis. GAPDH was used as a control. **(b)** Five splice forms of the IRGM gene have been identified, as indicated: IRGM(a)-IRGM(e). The promoter and 5'-untranslated regions of the gene are associated with an ERV9 retroviral LTR. Scale-bar is given in base pairs.

III.II.2.Purification and analysis of recombinant IRGM(a) protein

To characterize IRGM(a) protein biochemically, recombinant IRGM(a) protein N-terminally fused to MBP (Maltose binding protein) was expressed and purified from *E. coli*. Due to the in efficient digestion of the fused protein with enterokinase, thrombin digestion site was introduced

immediately after the enterokinase digestion site, integrated protease digestion site to pMALp2E vector, just before the putative start codon of IRGM(a) protein (Fig. 21a). MBP-IRGM(a) fusion protein was recovered in a soluble form (Fig. 21b) and subjected to further purification using gel filtration and anion exchange chromatography. Gel filtration experiments showed that the protein of interest eluted in the void volume of the column suggesting a high molecular complex or aggregation of the protein. Further analysis was performed using the dynamic light scattering to detect exact molecular mass of the complex. Dynamic light scattering analysis showed that IRGM(a) protein forms a complex of about 14000 kD (Fig. 22). Due to the consistent impurity problem faced during the purification experiments, a GTPase deficient IRGM(a) protein was generated by mutational exchange of serine (S) to asparagine (N) at position 47 corresponding to G1 motif, which is known to be essential for GTPase activity (GxxxxGMS to GxxxxGMN) (Taylor et al., 1996). This mutant MBP-IRGM(a) protein was expressed under the same conditions as wild type protein.

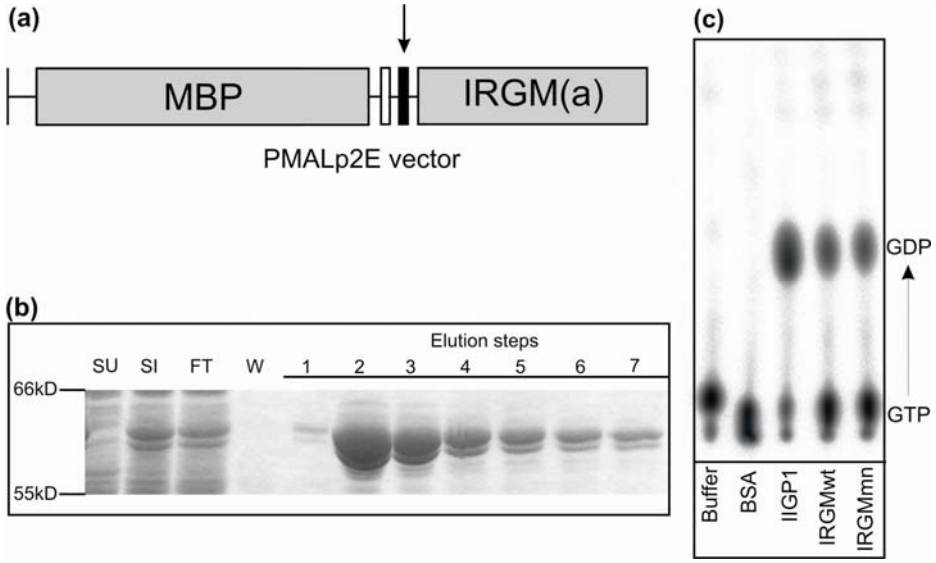


Figure 21. Purification and analysis of recombinant IRGM(a) protein.
(a) Schematic representation of expression construct of IRGM(a) protein. IRGM(a) protein was fused to C-terminus of MBP (Maltose binding protein) in open reading frame using commercially available prokaryotic expression vector pMALp2E . Arrow indicates the position of introduced thrombin digestion site by PCR (black box) after the enterokinase digestion site (white box). **(b)** Purified recombinant MBP-IRGM(a) protein using amylase resin. Supernatant obtained after centrifugation of 50000g for 30min at 4°C of lysate obtained from *E. coli* (NB42) which is induced with 100 μM IPTG (SI) for 15 hours at 18 °C or not induced (SU). The supernatant (SI) was loaded on 1G amylase resin column. Flow thorough (FT) was collected and washed (w) 10 column volumes. Proteins specifically bound to the coloumn was eluted with 10 μM Maltose containing elution buffer (elution steps 1-7). Purified protein used to raise an antiserum**(c)** GTPase hydrolysis assay of recombinant MBP+IRGM(a) protein. 80μM of MBP+IRGM(a) wild type (IRGMwt) and mutated (IRGMnm) incubated in the same condition for 2 hours at 37 °C in B1 buffer. IIGP1 (80μM) was used as positive control. BSA (80μM) and Buffer alone were included as negative controls.

Thin layer chromatography (PLC) experiments based on radioactively labeled GTPase assay showed that both mutated and wild type fusion proteins exhibit indistinguishable GTPase

activity (Fig. 21c). It can therefore be concluded that GTPase activity is the result of non-specific protein contamination. Recently, same biochemical properties were observed with the mouse GMS type Irg protein Irgm2, (GTPI) which is expressed in *E. coli* N-terminally fused to GST, (Robert Finking, personal communication). Therefore, it is probable that expression of GMS type p47 GTPases in a prokaryotic system is problematic and other expression systems have to be tested.

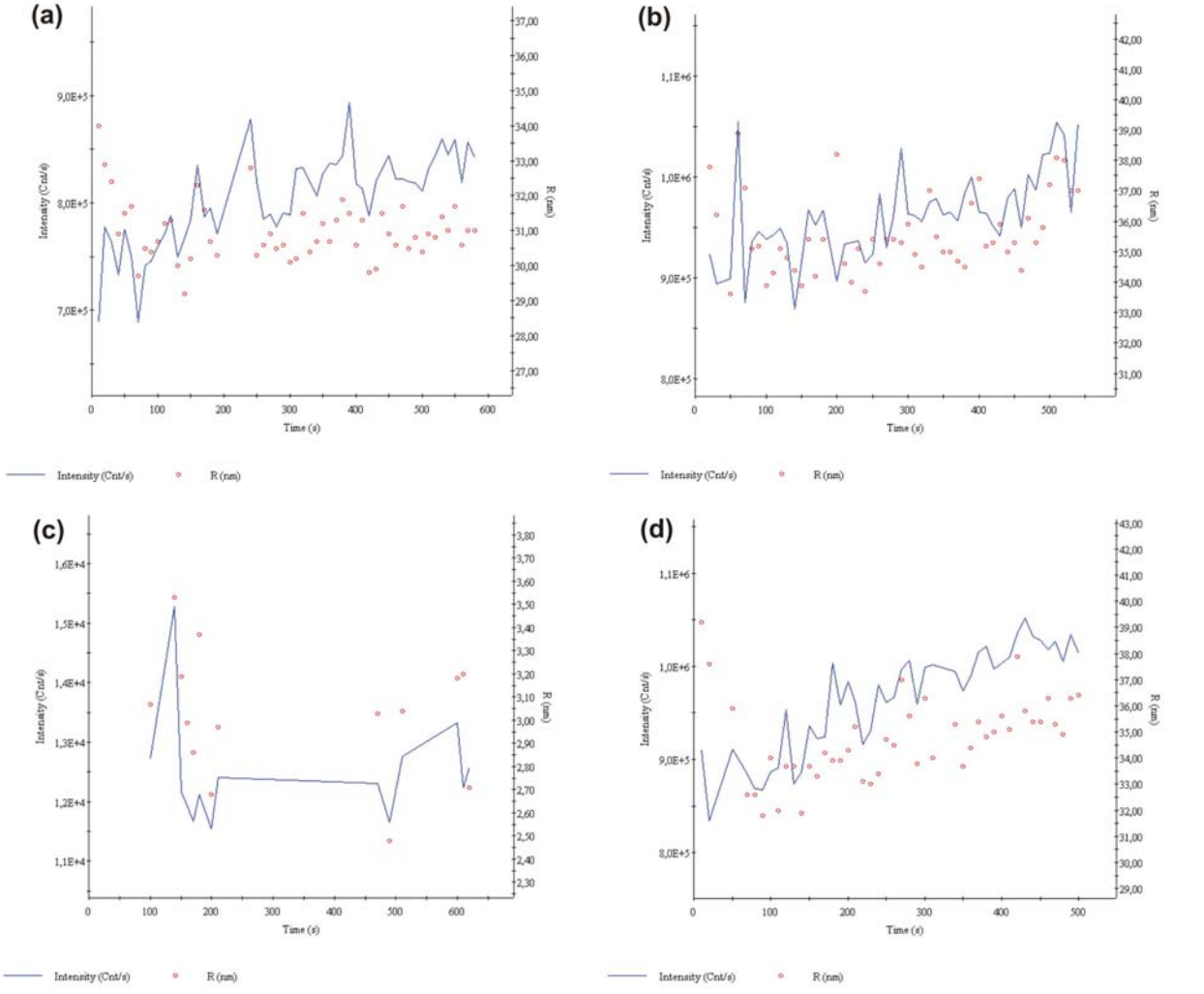


Figure 22. Dynamic light scattering (DLS) of MBP-IRGM fusion protein complex
(a) Dynamic light scattering (DLS) analysis of mutated MBP-IRGM(a) fusion protein in the presence of GTP average number of R (hydrodynamic radius) is 31nm which is approximately equal to 14000kD protein mass (b) DLS analysis of wt MBP-IRGM(a) fusion protein in the presence of GTP average number of R is 36nm which is approximately equal to 14100kD protein mass (c) DLS analysis of IIGP1 in the absence of GTP average number of R is 3.1nm which is approximately equal to 47kD protein mass (Uthaiah, 2002). (c) DLS analysis of MBP-IRGM(a) fusion protein in the presence of GTP average number of R 36nm which is approximately equal to 14500kD protein mass. 80 µM from each protein in the presence or absence of GTP in B1 buffer (50mM Tris/HCl, 5mM MgCl2, 2mM DTT, PH: 7.4) in final volume 70 µl was kept on ice. 10 µl of 100mM GTP was added and mixed very quickly by pipetting. The mixture immediately was transferred to spectrophotometer cuvette and placed into the Dynamic Light Spectrophotometer (Dynapro, protein solutions) at 37°C.

III.II.3.Immunofluorescence analysis of IRGM(a)-ctag1 and IRGM(b)

Immunofluorescence analysis was carried out to examine the intracellular distribution of IRGM(a) in human cell lines. IRGM(a) tagged with ctag1 (see Material and Methods) was transiently expressed in HeLa and GS293 cells. Under these conditions IRGM(a) exhibited two types of formation; aggregated and soluble (Fig. 23a). The ratio between aggregated and soluble form was varied between experiments. 90% of the aggregated form of IRGM(a) protein tagged with ctag1 has unexplained nuclei disruption (data not shown). Other spliced form of IRGM protein, IRGM(b) was cloned in mammalian expression vector pGW1H and was transiently expressed to confirm that aggregate formation of IRGM protein is not specific to differential splicing form. IRGM(b) was detected using the rabbit antiserum, raised against recombinant human IRGM produced in *E. coli,* (see material methods and below) (Fig 23b). Our analysis was consistent with previous observation that there were two types of formation and of those formed aggregate, 90% has disrupted nucleus formation. This is probably due to an experimental artifact, which could be linked to the general problems of overexpression of proteins however other p47 GTPases (IIGP1, LRG47, IGTP) expressed under identical conditions did not cause nuclei disruption. Finally, endogenous expression of IRGM protein was analyzed by immunofluorescence. Analysis using the human cell lines HeLa, GS293 and T2 cells revealed that there are no specific signals to antibody used (α4181) in detectable level. There is only background Golgi staining which is also observed with preimmune antiserum (indicated with white arrows in Fig. 23c). Our immunofluorescence analysis is consistent with the results obtained by immunoblotting (see below). Thus, it is evident that specific signal for endogenous IRGM protein can not be detected under these experimental conditions.

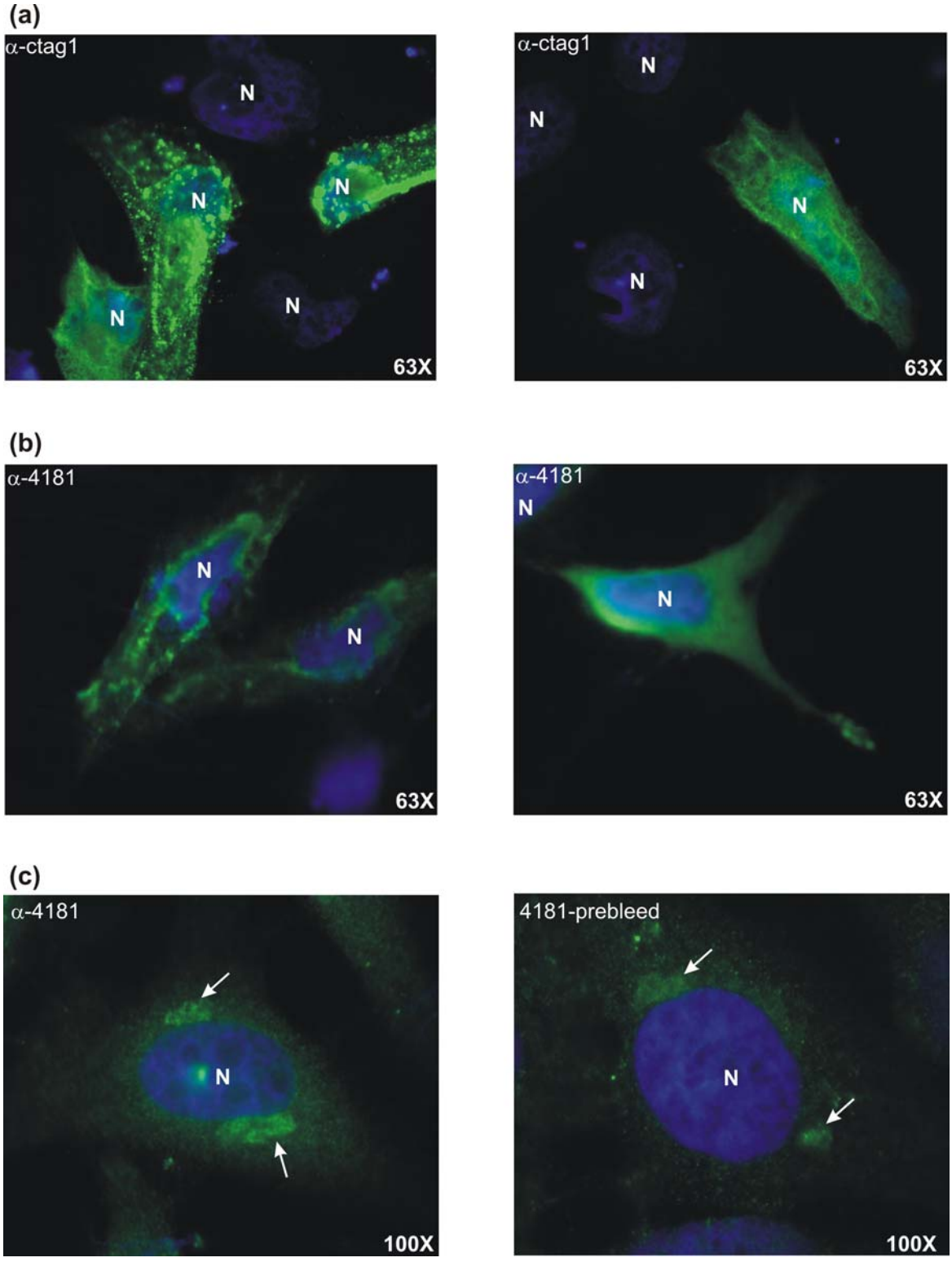


Figure 23. Immunoflourescence analysis of IRGM protein
(a) IRGM(a) protein c-terminally tagged with ctag1 is transiently expressed in HeLa cells using the antibody (αctag1) of a dilution 1:5000 ratio showing aggregated (left) and soluble expression (right). Dapi is used to label nuclei and indicated with N. Images were taken using the 63X objective (630) **(b)** IRGM(b) protein is transiently expressed in Hela cells using the antibody (α4181) in 1-500 ratio showing aggregated (left) and soluble expression (right). Dapi is used to label nuclei and indicated with N. Images were taken using the 63X objective (630) **(c)** Screening of endogeneous IRGM protein in Hela cells (left). The results were crosschecked using 4181-prebleed serum (right). Background Golgi staining is indicated with white arrows. Dapi is used to label nuclei and indicated with N. Images were taken using the 100X objective (1000). For handling all images, Zeiss Axioplan II microscope equipped with cooled CCD camera and metamorph software (4.5) are used.

III.II.4.Westernblot analysis of IRGM protein

A rabbit antiserum (α4181), raised against recombinant human IRGM(a) produced in *E. coli* (see material and methods) could not detect specific signal for endogenous IRGM protein from extracts prepared using human Hela, GS293, T2, Thp1, HepG2 cell lines (Fig. 24a-b). Additionally, no IRGM protein could be detected after induction by interferon (Fig. 24c) suggesting that IRGM protein is not translated *in vivo* and is not induced by interferon in cultured cell lines under these experimental conditions.

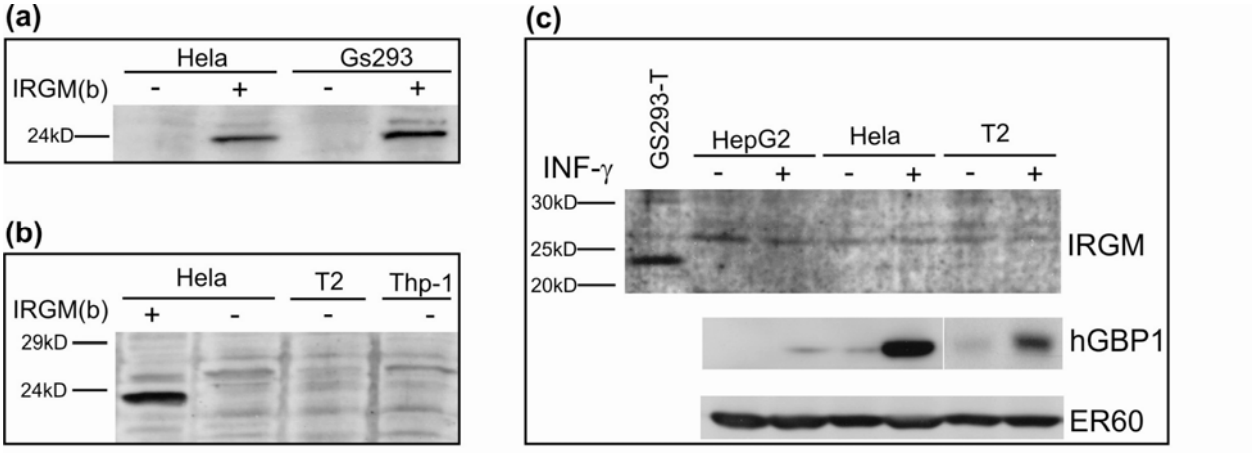


Figure 24. Endogeneous expression of IRGM protein in human cell lines
(a) Eukaryotic expression vector pGW1H containing of *IRGM(b)* transfected (+) or not transfected (-) to Hela and GS293 cells, respectively. 24 hours post transfection, cells were harvested and lysed in 1X SDS protein loading buffer. IRGM protein was detected by immunoblotting with antibody (α4181) at 1:500 dilution. **(b)** Endogeneous expression of IRGM protein in Hela, T2 and Thp1 cells. (+) and (-) indicate transfected or untransfected cells, respectively. Cells were harvested and lysed in 1X SDS protein loading buffer. IRGM protein was detected by immunoblotting with antibody (α4181) in 1-500 dilution. **(c)** Interferon induction experiments of IRGM protein in human cell lines (Hela, HepG2 and T2 cells). Cells were induced (+) or uninduced (-) for 48 hours with 200 u/ml IFN-γ and protein extracts were prepared by cell lysis (2% TritonX 100 in PBS with protease inhibitor) for 2 hours on ice. *IRGM(b)* transfected GS293 cells were used as positive control (GS293-T). hGBP1 was used as positive control for interferon induction and gel loading was assessed by immunoblotting with ER60 specific antibodies. Detection of IRGM protein on nitracellulase membrane was performed using the antibody (α4181) in 1-500 dilution.

IV.DISCUSSION

Adaptation of an organism to different environments is the main cause of organismal diversity but there are many types of adaptation mechanisms. Host-pathogen interaction is possibly the strongest adaptation mechanism that leads different species to coevolve (Haldane, 1949), (Summers et al., 2003). Coevolution of the species is maintained by two-way biochemical interactions leading to responses in both pathogen and host cell (Galan and Bliska, 1996). Responses against pathogens by the host cell use complex signaling pathways and require involvement of different types of regulation and induction of specific regulators such as cytokines. One of the cytokines involved in resistance against pathogens is interferon γ, which is known to be one of the most important regulators of immunity. Interferon γ is responsible for induction of more than 800 genes (Boehm et al., 1997). It has been suggested that in mouse, the interferon γ-induced resistance activity against protozoa and pathogenic bacteria is mainly mediated via the p47 GTPase family in a cell autonomous manner (Taylor, 2004).

The p47 GTPase family may be one of the most important resistance factors in the mouse (Taylor, 2004). The evolutionary analysis of p47 GTPases led us to the following unexpected conclusions; Firstly, the family of p47 GTPases has 23 members in the mouse. It is also shown that a minimum estimate of the number of potentially functional p47 GTPases in mouse is not just six, as previously described (Boehm et al., 1998), but rather 20. Strikingly, the resistance mechanism of p47 GTPases appears to be completely absent from the human lineage. Secondly, this mechanism might make use of hetero-dimer, trimer or even higher oligomer formation. Thirdly, members of the p47 GTPase family appear to be fast evolving genes. However, only 5 (4 in C-terminus) genes of the mouse family members could be shown to have been under positive selection. Finally, different numbers 23, 15, 7, 18, 2, and 2 of relatively differentiated members of p47 GTPases are present in mouse, rat, dog, Zebrafish, Fugu and Tetraodon, respectively. Variable numbers of p47 GTPases in different species may reflect a co-adaptation process in order to generate diversity in the resistance mechanisms acting on pathogens which are usually known to be fast evolving.

All the above conclusions and implications of the results obtained in this study will be discussed in detail in the following sections.

IV.1.p47 GTPases are completely absent from the human lineage

Including with the previously published six inducible p47 GTPases, I have reported here that the family of p47 GTPases in the mouse is encoded by 23 genes sharing many common properties at their N-terminus (first three α-helices), C-terminus (last seven α-helices), and G-domain (near to the N-terminus with six β sheets and five α-helices) (Fig. 7, 9 and 11). Out of the 23 identified genes, two are likely to be pseudogenes. Of the remaining 21 genes, 14 were shown to be interferon-γ-inducible and database analysis indicates that the 6 of the remainig 7 p47 GTPase gene are also functional with respect to the promoter and transcript structure. One gene, Irgc, was shown to be not containing any interferon response element in its promoter region and this study provides evidence that it is not involved in immunity (see below). As the interferon-γ-inducibility is generally indicative of an immune function of the respective gene, the presence of 14 interferon-γ-inducible members of the p47 GTPase family argues for a remarkable significance of these genes for immunity.Indeed several members this family have been shown to be essential for the resistance of mice against diverse pathogens, including *T. gondii, L. monocytogenes, M. tuberclosis* (Taylor, 2004). On the other hand, the human genome encodes only one representative p47 GTPase-like gene, which seems not to be involved in immunity and one expressed fragment, which encodes only the G-domain of a GMS-like GTPase with the promoter region containing an endogeneous retroviral element (ERV9) (see below). It is very well known that pathogens that are able to infect mice have at least one close relative, which is infectious for human. The number of pathogens, against which p47 GTPases are involved in resistance are listed in Table 1. It is therefore necessary to ask, “Why do humans lack such a strong resistance mechanism?”

It is expected and known that there are differences between the human and the murine immune system, some of which are listed in table 5. It is known from previous reports (Mestas and Hughes, 2004) that none of the known immune mechanisms drastically differs between mouse and man. At least one representative of each resistance mechanism which is present in mouse is also present in man. This can be explained in the context of coadaptation of host with their specific pathogens. However, to link the absence of the entire resistance mechanism mediated by p47 GTPases in humans to a classical coadaptation of host and pathogen is not a satisfactory explanation. It can be suggested that the mechanism disappeared from the human lineage because of the integration of a retroviral element into the promoter region of the GMS-like GTPase, *IRGM*. As discussed below, the loss of one central member of the p47 GTPase family may have implications for the proper function of the whole resistance mechanism

mediated by this family. Thus, changing the expression profile of the putative *IRGM* gene may have resulted in a non-functional system. In the human lineage, another strategy must have been present to eliminate intracellular pathogens. Therefore, the disappearance of p47 GTPase family from the human genome might be either because of an accident or classical host-pathogen coadaptation.

	Mouse	**Human**	**Notes**	**Reference**
TLR2 expression on PBL	Low (induced on many cells including T cells)	Constitutive (but not on T cells)	Binds lipopeptides	(Rehli, 2002)
TLR3	Expressed on DC, Mac. Induced by LPS.	Expressed by DC. No LPS induction	Binds dsRNA	(Rehli, 2002)
TLR9	Expressed on all myeloid cells, plasmocytoid DC and B cells	Expressed only on B cells, plasmocytoid DC and NK cells	Binds CpG	(Lund et al., 2003)
TLR10	Absent	Present		(Roach et al., 2005)
Sialic acid Neu5GC expression	Widespread	Absent	Binds pathogens	(Varki, 2001)
Leukocytes defensins	Absent	Present		(Risso, 2000)
Paneth cell defensins	At least 20	Two		(Ouellette and Selsted, 1996)
Macrophage NO	Induced by IFN-γ	Induced by IFN-α/β, IL-4^+ anti CD23		(Weinberg, 1998)
CD4 on Macrophages	Absent	Present		(Crocker et al., 1987)
NK inhibitory receptor for MHC1	Ly49 (family except Ly49D and H)	KIR		(Lanier, 1998)
FcαRI	Absent	Present		(Monteiro and Van De Winkel, 2003)
*TLR11, TLR12 and TLR13	Present	Absent	Recognize profilin like molecules from the protozoan parasite *T. gondii* and uropathogenic *E. Coli*	(Roach et al., 2005) (Yarovinsky et al., 2005)
* The family of p47 GTPases	20 functional genes	Absent	A mechanism required for resistance against vacuolar pathogens (Table 1)	(Bekpen et al., 2005)

Table 5. Summary of known differences between mouse and human innate immunity.
Some of the different genes or gene family involve in innate immunity were summarized. Original table containing all the differences known in adaptive and innate immunity were prepared by (Mestas and Hughes, 2004) and * updated by using recent reports.

As mentioned above, it is concluded that humans have only one full-length p47 GTPase, IRGC, which is homologous to mouse IRG proteins. Irgc is the single p47 GTPase located on mouse chromosome 7, hence, showing a different chromosomal location from the other chromosome 11

and 18 groups. Human IRGC, which is located on chromosome 19 is syntenic to mouse Irgc. This protein displays a high degree of homology (more than 90%) and is orthologous between mouse, dog and human. Codon based selection analysis revealed that the *Irgc* gene is evolving under purifying selection (Fig. 19), thereby following the characteristic evolutionary behavior of housekeeping genes rather than immunity related genes. Notably, it is expressed only in testis. Furthermore, inducibility experiments were carried out with interferon γ and β. There was no detectable level of interferon-induced transcription of Irgc in both human or mouse cells. In addition, analysis of different tissues derived from mice infected with the pathogen *Listeria monocytogenesis,* which is known to cause massive interferon-dependent induction of classical p47 GTPases (Boehm et al., 1998), showed no up-regulation of Irgc (Christophe Rohde personal communication). Considering the evolutionary behavior, gene structure and functional analysis, the *IRGC* gene is very unlikely to be a representative of the p47 GTPase family in the context of human cell autonomous immunity.

The other p47 GTPase like protein IRGM is not considered to be a functional gene because it does not encode a full-length p47 GTPase and because no protein product could be detected by immunoblotting and immunofluorescence analysis under all experimental conditions tested. Currently, there is no explanation why the *IRGM* gene is not translated. ERV9 involved regulation of transcription is very well known and reviewed by Lower et al., 1996. The promoter and transcriptional structure of the *IRGM* gene is very similar to ZNF80 gene which encodes a putative zinc finger protein (Di Cristofano et al., 1995) and it is reported that ERV9 LTR regulates the transcription of β-globin gene via locus control region (LCRs) (Routledge and Proudfoot, 2002). Furthermore, ERV9 LTR is located, in the antisense orientation, in the second intron of the axin gene, which contains eleven exons and spans 58kb on chromosome 16. It has been shown that ERV9 LTR also has an effect on the transcription of the axin gene (Ling et al., 2002).

ERV9 LTR driven expression is highly effective especially in embryonic, hemotapoietic cells. The various kind and different number of transcripts expression driven by LTR was detected in adrenal gland and testis (Ling et al., 2002; Svensson et al., 2001). The difference in the level of *IRGM* gene expression in HeLa and GS293 cells was also observed. Approximately, 100-fold higher expression was detected in the embryonic kidney cell line, GS293 than in HeLa cells. RT-PCR analysis using cDNA, synthesized from human brain, liver and testis total RNA showed that IRGM has highest expression in testis (Fig. 20). It was of interest to elucidate whether the IRGM promoter region also possesses the capability of interferon inducibility or not. Bioinformatic screening analysis for an interferon response element, using 10kB upstream of the

transcription start site, in the promoter region of IRGM, revealed no potential ISRE or GAS site. To confirm that this gene is not regulated by interferons, interferon-inducibility experiments were carried out using different cell types (see above) and resulted in a failure to up-regulate the *IRGM* gene or protein by interferon γ and β after 24 hour. Thus, an alternative interferon-inducible promoter comprising ISRE or GAS elements out side the analyzed 10kB promoter region, is unlikely to be present. Immunoblotting and immunofluorescence analysis using an antiserum raised against recombinant IRGM protein failed to show its presence in human cells despite the presence of transcript. It is therefore concluded that *IRGM* is an expressed pseudogene. However, there is no doubt that the IRGM protein sequence is closely homologous to the mouse *Irgm* genes. Irgm1 (LRG47) is up to now the most effective resistance gene among all mouse *Irg* genes (Table 1) (MacMicking et al., 2003). The *IRGM* gene in Chimp (*Pan troglodytes*) a close relative of human, has promoter and genomic structure similar to that of the human gene. This clearly indicates that the structure of IRGM is a common feature in primates at least for Hominini tribe. Therefore, it can be suggested that humans had the mechanism of p47 GTPases and probably lost the entire mechanism during the course of primate evolution leading to the human lineage (see Fig. 25) while the mechanism was retained by the other vertebrates (Fig. 17) as dogs rodents and fish.

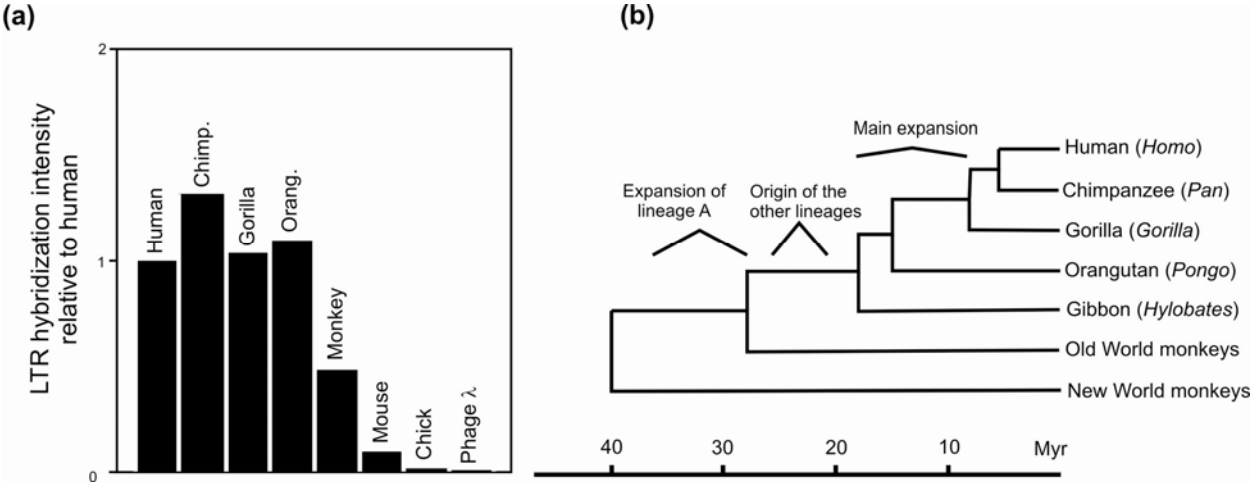


Figure 25. ERV9 element in primates and evolutionary history.
(a)ERV9 LTRs are present both in the higher and lower primates (Ling et al., 2002).Copy numbers of ERV9 LTRs in primates and non-primates relative to the haplaoid copy numbers in human detected by northern blot. (b)Inferred evolutionary history of ERV9 elements superimposed on a phylogenetic tree of primate evolution (Costas and Naveira, 2000).Estimations of ERV9 transpositional ages are based on average divergences of members of each subfamily from their respective consensus sequences.

Maybe the explanation for the loss of the p47 GTPase family in the human lineage, is the disruption of the promoter region of the *IRGM* gene by the ERV9 LTR. Such an event would be predicted to lead to a complete change in the expression profile. The interferon inducible

promoter would become converted to a constitutive promoter which is unresponsive to interferons. The promoter region of IRGM like other ERV9 derived promoters contains GATA (Shivdasani and Orkin, 1996), CCAAT (Tenen et al., 1997) (Yamanaka et al., 1997), and CCACC (Miller and Bieker, 1993) motifs and is potentially capable of binding to cognate transcription factors expressed in embryonic and hematopoietic cells. ERV9 is an endogenous retroviral element belonging to a family containing at least 14 different subfamilies and is specific to primates. Probable appearance of ERV9 was calculated to be as early as 40 million years ago and the main expansion in primates was observed 15 million years ago (Costas and Naveira, 2000) (Ling et al., 2002). Therefore, it can be assumed that the disruption of the promoter region of human IRGM in the ancestor of primate lineage took place during the expansion period of the retroviral element within the primate lineages (see Fig. 25b).

However, the question remains why the the whole p47 GTPase family should disappear when only one gene is damaged by a retroviral integration. Recent studies indicate that the family of p47 GTPases is functionally interdependent. This is supported by the observation that the GMS proteins (Irgm1, Irgm2 and Irgm3) are required for the function of the murine GKS type Irg proteins (Irga6, Irgb6 and Irgd). Namely, the transfected individual GKS proteins form unexplained aggregate structures whose behaviour differs drastically from the intracellular behavior of the interferon-induced endogenous GKS type p47 GTPases. In culture cell lines using a transient eukaryotic expression system, co-transfection of GMS proteins together with GKS proteins results in re-localization of the transfected GKS proteins similar to that of endogenous GKS proteins. Similarly and more importantly, the GMS type Irg proteins are required for the transfer of GKS proteins to the toxoplasma containing vacuole in cultures cells (Julia Hunn and Nina Schroeder unpublished results). Additionally, some Irg proteins appear to be transcribed as unusual tandem genes (Irgb2-1, Irgb5-b4) or as a triplex, which contains two GMS proteins (rat irgm2, rat Irgm3) and one GKS protein (rat Irgb10) (see below). Furthermore, recent functional genetic analyses by targeted gene knock out experiments suggest a unique importance of Irgm1 protein among other p47 GTPases. The Irgm1 (LRG47) appears to be required for all the p47-dependent resistances yet tested, while the other p47s appear to be required only individual resistances. Therefore, it can be suggested that the GMS genes are the key players for p47 GTPases in mouse, meaning that the system itself probably is working in a layer like structure or combinations (see below). This interdependent mosaic behavior of p47 GTPases is also observed in the intracellular localization of the family members. Irgm2, Irgm3, Irga6, Irgb6 and Irgd localize to the taxoplasma containing vacuole, (Martens S, 2005). Irgm1 is localized to the Mycobacterium containing vacuole (MacMicking et al., 2003). However, It can

be suggested that the other *Irg* genes probably also require GMS genes to be fully functional. Therefore, altering the expression profile of one of the most important family member by ERV9 is the possible reason why human lost the entire family of p47 GTPases. After the disruption of the promoter region of human IRGM by the retroviral element, the p47 GTPase family may have lost much of its advantage and may indeed have caused enough costs to be eliminated. In this context, it is worth reiterating that loss of normal gene function driven by endogenous retroviral element integration associated with several type of disease and cancer formation is well known and reviewed in detail by Lower et al., 1996. On the other hand, fitness costs of resistance genes is a very well described phenomenon in immunity (Tian et al., 2003), (Rigby et al., 2002), (Burdon and Thrall, 2003). As mentioned before, one resistance gene in Arabidopsis, *RPM1*, has a significant cost of fitness. Both resistance and susceptibility alleles frequently occur together within natural populations. The evolution of the interferon-inducible resistance gene, *Mx1*, which is required for resistance against influenza virus A and B, *Mx1*, is maintained by balancing selection in the nature (Jin et al., 1998b; Staeheli et al., 1988). However, cost of fitness for the Mx1 gene, have not been reported. Resistance genes in the mouse are generally regulated by cytokines such as interferon γ and β. Fitness cost related with resistance genes is probably the reason why the transcription of inducible large GTPases is controlled by interferons. Higher fitness cost of p47 GTPases might be also responsible for the loss of the mechanism from human lineage. It is hard to prove whether the ERV9 or fitness cost of p47 GTPases was primarily responsible for the disappearance of the family from human lineages. Perhaps, the best explanation will be that the combination of both was the reason why humans do not have the mechanism of p47 GTPases. However, as pointed out in results, the family of p47 GTPases has also appearently disappeared from other groups like birds and Xenopus suggesting that these genes evolved with the mechanism of immune response, under different selection pressures (coevolution) leading to their disappearance from some of the main branches of the eukaryotes.

It is an important and un-answered question what replaces p47 GTPases function in man? All the innate immune mechanisms such as nitric oxide and oxygen radicals (Fang, 2004; Nathan and Shiloh, 2000), purinergic receptors (Lammas et al., 1997), tryptophan depletion (Pfefferkorn, 1984; Robinson et al., 2003), cation depletion (Schaible and Kaufmann, 2004), authophagy (Gutierrez et al., 2004) and TLRs (Roach et al., 2005), are present in the mouse. It is possible that one or more of the mechanisms listed above filled the gap left by loss of the p47 GTPases in man. This is consistent with the observation that a mouse oviduct cell line expresses interferon inducible iNOS (inducible nitric oxide synthase) however does not express IDO (2'3' indolamine deoxygenase) upon treatment by interferon whereas in Hela cells IDO expression can

be induced by interferon treatment and is responsible for a remarkable level of resistance against *Chlamydia* species causing disease in humans (Nelson et al., 2005).

It is of course possible that an unrelated and so far unidentified molecular machine in the primates performs the resistance mechanism of p47 GTPases. In fact, preliminary screening revealed that a primate specific gene family called *Morpheus*, which has similar evolutionary behavior like p47 GTPases, exists with unknown function. It has been shown by Johnson M. E. et al., (Johnson et al., 2001) that fifteen distinct copies of duplicated segments were present on chromosome 16 of human and transcripts were identified for six of the 15 genomic copies. Similar to p47 GTPases, the number of the duplication segments is variable within the primate lineages as 9, 17, 15, 25-30 which is specific to primate species in orangutans, gorillas, human, and chimpanzee respectively. Codon based selection analysis revealed that they are relatively fast evolving genes, therefore, their function are expected to be related either with immunity or reproduction. It will be of interest to see whether this family or undiscovered families specific to the human genome are responsible for the mechanism of p47 GTPases in man.

The mouse is a model organism used as an experimental model for human diseases for many decades, however in this study, a clear distinction between the mouse and the human immune system in the sense of cell autonomous immunity is discovered. In the light of the data presented here, scientists should consider the differences in cell autonomous immunity between man and mouse when they carry out experiments to analyze the immune response against intracellular pathogens.

IV.2.Evolution of p47 GTPases

Evolution of p47 GTPases can be explained in two ways. Firstly, the family of p47 GTPases is evolving by increasing or decreasing the number of the genes probably because of the negative selection pressure by the pathogens. This leads to a increased or decreased diversity within p47 GTPases among different species. Secondly, the members of p47 GTPases itself are relatively fast evolving genes (see below and Fig. 19).

All the vertebrates analyzed so far fugu, danio, mouse, dog, cow, pig, amphibian have at least one copy of a p47 GTPase-like gene or a set of p47 GTPases whereas in plants, so far no p47 GTPaselike gene was detected. It is clear from fig. 18 and 26 that variations in the number of the p47 GTPases among different species are generated via gene duplications, which can arise through polyploidization, non-homologous recombination, or retrotransposition. The plausible duplication scenario for p47 GTPases in mouse is depicted in fig. 27. Gene duplications are

considered to be a mechanism to increase the diversity in immunity-related genes or gene families, (Wagner, 2002) (Kondrashov et al., 2002) (Leister, 2004). This has e.g. been shown for the 2'-5' Oas family that is crucial in the interferon induced antiviral response (Kumar et al., 2000) (Mashimo et al., 2003).

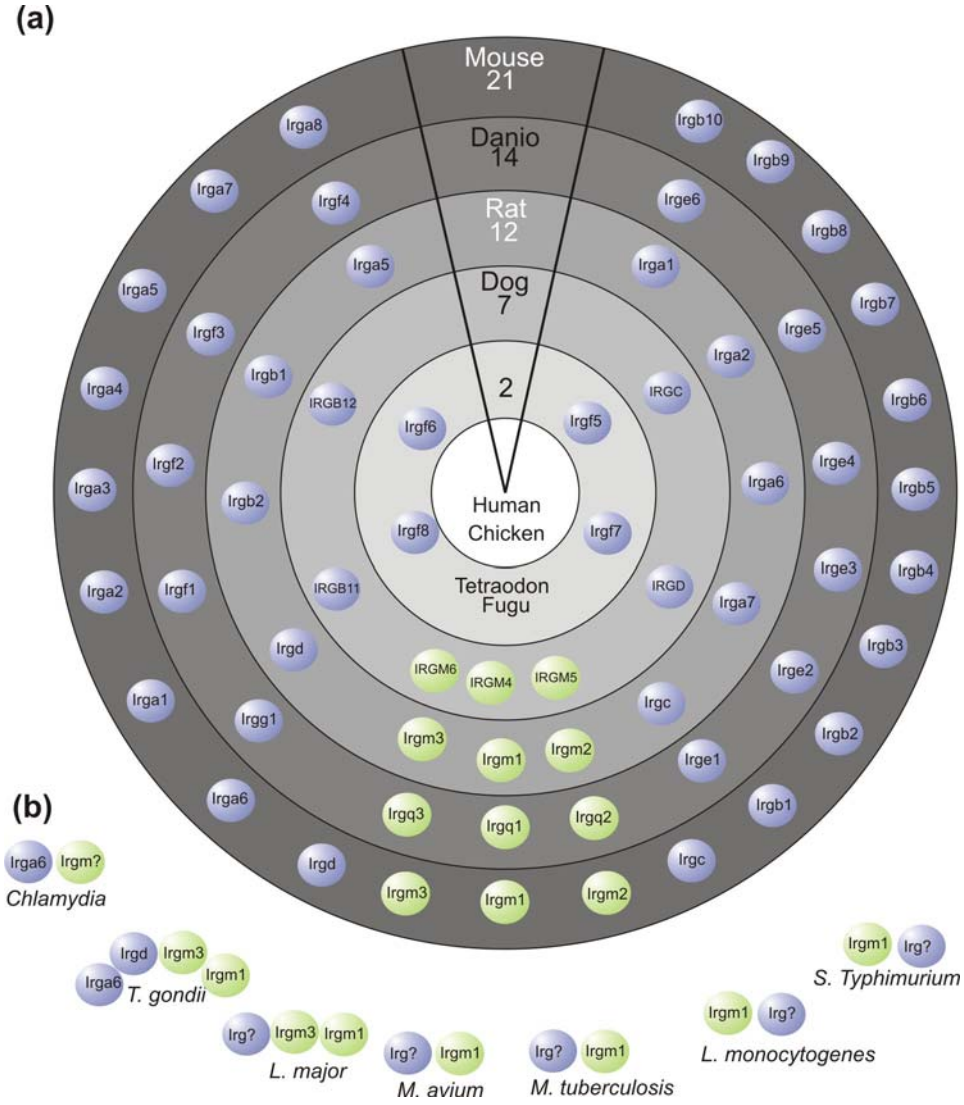


Figure 26. Schematic representation of diversity of IRG proteins in vertebrates
(a) Observed diversity of p47 GTPases within vertebrates specific to species that is correlated with the diversity of parasites of which p47 GTPases used for. When the diversity of the pathogen increase, the diversity of p47GTPases increase (Intensity of the black color and number of the IRG proteins). Pseudogenes were not included. GMS or Quasi type GTPases depicted in green color. Please note that over all diversity for rat and dog summarized here based on the search performed on available public databases which were not completed yet. By the accumulation of data the picture can be changed. **(b)** The plausible combinations of p47 GTPases acting as resistance factor against pathogens were summarized. The question mark indicates the expected GKS or GMS type p47 GTPases waiting to be functionally analyzed on specified pathogen. For more information, please see (Table 1)

When the gene duplication occurs, the duplicated gene becomes redundant and free of selection (Kondrashov et al., 2002). Therefore most of the duplicated genes are predicted to become lost due to accumulation of deleterious mutations or subjected to directional positive selection because they are now free from the obligation of purifying selection (Wagner, 1998)

(Wagner, 2002). Only two of the 23 p47 GTPases in mice were classified as pseudogenes in C57BL/6 mice by a criterion not being able to encode a full-length protein because of the accumulation of null mutations resulting in generation of stop codons in the primary ORF. However, it appears that in the family of p47 GTPases, many genes stay intact after duplication.

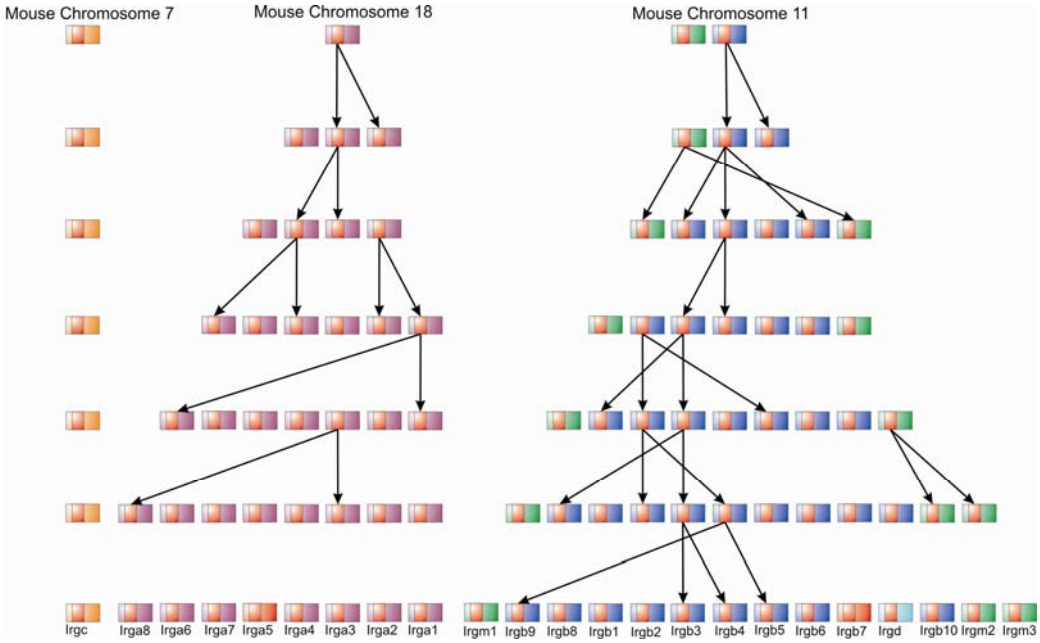


Figure 27. Duplication scenario for mouse Irg proteins
Possible duplication scenario of 23 mouse Irg proteins were illustrated by using phylogenetic analysis and chromosomal distribution of p47 GTPases as reference. Arrows indicates the predicted duplication events giving rise to new gene formation. Irga proteins, which are located on mouse chromosome 18, were illustrated in purple color. Irgb, Irgm and Irgd proteins which are located on mouse chromosome 11 were illustrated in dark blue, green and light blue color respectively. Irgc, which is located on chromosome 7, was illustrated in orange color. Pseudogenes are colored in red.

Diversity of p47 GTPases acting on pathogens was probably generated by increasing the number of p47 GTPases in different species. This will result in different numbers of p47 GTPases among different organisms which may reflect evolutionary coadaptation by direct host-pathogen interactions. So far, dog, mouse, rat, fugu, tetraodon, and zebrafish show very different distribution of p47 GTPases (Fig 26). Furthermore, two of the reported pseudogenes Irgb5 and truncated p47 GTPase Irgb10 in mouse encode full length functional p47 GTPases in rat whereas two of the highly degraded pseuodogenes in rat Irga14 and Irga16, are functional genes in mouse. Even in the Czech II mouse (*M. musculus musculus*), which is a very close relative of C57BL/6 (*M. musculus domesticus*), different subsets of p47 GTPases can be detected. For instance, in C57BL/6 mice Irga8 is a pseudogene since there is an insertion of adenine base at the position 614 corresponding to 204th amino acid while Czech II mouse has a full length gene closely homologue to Irga8 (Fig 7).

Generation of diversity by somatic recombination and gene conversion are very well known mechanism in adaptation to fight against pathogens (Martinsohn et al., 1999),(Flajnik, 2004), (Summers et al., 2003). Pancer et al., (Pancer et al., 2004) showed that LRRs (leucine reach repeats), which are a characteristic feature of innate immune recognition receptors, could be used for generation of diversity. This suggests that genes involved in innate immunity can indeed be used in generation of diversity. Therefore, it could be suggested that host pathogen coadaptation is the primary defining force for the fate of the duplicated p47 GTPase gene or genes to determine whether they will decay to pseudogenes or evolve into new functional genes and act as another tool for diversity to fight against pathogens.

Apart from generating diversity by genomic duplication, the individual p47 GTPases are subjected to positive selection. Five of the p47 GTPases were detected to be under positive selection especially in their C-terminus, suggesting recent coadaptation (Fig. 19). It is known that the so called αK helix in the C-terminal region of Irgm1 and Irgm2 is important for the proper intracellular localization of the respective protein (Martens et al., 2004). Recently, Kaiser et al., (Kaiser et al., 2004) identified for the first time interaction of the IIGP1 protein with Hook3, which is a microtubule motor binding protein and involved in cellular trafficking. The interaction occurs via the last α helix (αL) of C- terminus of IIGP1, which has some homology to other p47 GTPases within the family. Two fugu p47 GTPases, Irgf6 and Irgf5, highly differentiated in their C-terminus, however, preserve the classical properties of the C-terminal region of the p47 GTPases family whereas both genes are almost completely identical in N-terminus and G-domain. Therefore, it is possible that the C-terminus of p47 GTPases is generally important for localization as well as interaction with other proteins, perhaps even direct interaction with the proteins or molecules from pathogens.

IV.3.Oligomeric structures in p47 GTPases family

It is reported here that four of the mouse p47 GTPases were found to be transcribed as tandem genes, Irgb5 together with Irgb4 and Irgb2 together with Irgb1. The Irgb2-b1 tandem can be amplified by RT-PCR on cDNA synthesized by using mRNA extracted from L929 cell line and was shown to be inducible by interferon γ. In rat, there is a transcript encoding a triple p47 GTPase, comprising sequences equivalent to rat Irgb10-Irgm3-Irgm2. Rat Irgb13-Irgb14 has a genomic structure and splicing pattern similar to that of mouse Irgb2-b1 suggesting that rat Irgb13-Irgb14 can be transcribed as a tandem gene (Fig. 9 and 10a). Moreover, Zebrafish has one tandem pair containing irgg with quasi GTPase irgq.

Dynamin, dynamin like GTPases, Mx, and GBP are GTPases, which can form GTP dependent oligomers and this oligomerization is required for the function. Since the family of p47 GTPases shares biochemical properties with the dynamin like GTPases, it was expected that p47 GTPases are functional by formation of higher molecular structure in vivo, however formation of tandems at the transcriptional level was unexpected. Similar to classical dynamin like GTPases, Irga6 can form oligomers in vitro and formation of oligomers is stimulated by GTP binding (Uthaiah et al., 2003). Furthermore, Irga6 (IIGP1) forms dimers as determined by crystal structure and site-directed mutational analysis of the dimer-interface showed that N-terminal interaction is essential for dimerization (Ghosh et al., 2004). However, the dimer observed in the crystal structure shows N-terminus to N-terminus interaction of Irga6. The tandems and trimer are encoding head to tail genes and especially, the crystal structure of the rat trimer is completely unpredictable.

Why does a mouse need 20 functional genes of which transcription is tightly regulated by interferon γ and use this repertoire in a non-redundant way? The p47 GTPase family is massively induced by IFN-γ after 24 hours and the calculated induction ratio ranges from no or very low level up to 215 fold for IIGP1 and 50 for LRG47 at transcriptional level. The calculated number of protein molecules of IIGP1 per cell in L929 cells induced with 200 U/ml interferon γ after 24 hour is approximately $2x10^6$ (Jia Zeng unpublished results). If we make a rough calculation and assume that all the p47 GTPases are inducible (130 fold in approximate average) and are translated and active, the number of p47 GTPases in the cell within first 6 hours will increase from almost zero to 14 which is in total 14X130 = 1820 fold more p47 GTPases in the cell. However, the existence of transcription of tandem and triple Irgs, suggests that the functional unit of p47 GTPases might be dimmers, trimers or even higher oligomers. This view is also supported by yeast two hybrid assay (analysis of protein-protein interaction in *S. cerevisiae*) (Kaiser, 2005).

If we imagine that p47 GTPases would function as dimers and the position of the invidual p47 GTPases within the dimer is omitted (for example A-B=B-A, A-C=C-A), then the total number of the different dimers would be 91 possible combinations by using 14 individual p47 GTPase. If the functional unit is a trimer, then the total number will be 364 different combinations. Such diversity could be a big advantage for an organism to fight against pathogens (Fig. 26). In reality, combinations of p47 GTPases might be different. However, each pathogen has its way of infection and requires different niche to survive within the cell. The functional unit (combination) of p47 GTPases is probably required for resistance against specific pathogens. In

fact, it is known that TLRs can form homo and hetero-dimers in different combinations and so they can recognize different PAMPs. For example, TLR2 can form heterodimers with TLR1 or TLR6. A consequence of this cooperation is an increased repertoire of ligand specificities (Beutler, 2004) (Janeway and Medzhitov, 2002). It can be suggested from genetic evidence in mouse, fish, and rat that formation of higher molecular structures in p47 GTPases might naturally occur in a way that hetero-dimers, trimers, tetramers or even higher oligomers may form. Perhaps, it is an advantage for an organism to transcribe two or three genes in one unit rather than transcribing them separately and arranging them to interact post translationally. In fact, it is known that functionally related genes, at least in immune system are often genetically linked. For example, The *TAP1* and *LMP2* genes are transcribed from a shared bidirectional promoter containing an IFN response factor element that confers IFN-γ inducibility (Wright et al., 1995), (Dovhey et al., 2000).

Classical p47 GTPases TGTP (Irgb6), IIGP1(Irga6), IRG47 (Irgd), and IGTP (Irgm3) localize at the parasitophorous vacuole upon infection by *Toxoplasma gondii* (Martens et al., 2005 and Sascha Martens unpublished results). The pathogen containing vacuole is probably the place where functional oligomers form and one can easily imagine how many combinations of p47 GTPases are available at the same time on the phagosome.

IV.4.Origin of p47 GTPases

The p47 GTPase family might have been originated from cyanobacteria by horizontal gene transfer, and evidence for this assumption is that firstly all ORFs of classical p47 GTPases are encoded on a single exon which is characteristic for a gene of prokaryotic origin, though certainly not diagnostic. Secondly, there are GTPases present in bacteria (especially in cyanobacteria) with significant homology to the G-domain of p47 GTPases. Secondary structure prediction analysis reveals that they are related with p47 GTPases (Jonathan C. Howard personal communication). It will be of interest to elucidate whether cyanobacterial p47 GTPase-like genes possess a crystal structure similar to IIGP1.

V.APPENDIX

V.1.Appendix Table 1. List of all IRG gene family members and related genes

(Please note that detailed descriptions of the most of the genes presented in the table was prepared by Julia Hunn)

Gene name	Genesymbol/ID	Synonyms	Genomic sequences /Accession no.	cDNA or EST sequence Accession numbers	Notes
Mouse					
Irga1	Irga1 MGI:1795294 MGI:1653512 New gene		AC132320 AC102225	BI658674 (NMRI, 5'EST, nearly 100%) BG915086 (NMRI, 5'EST; not 100%)	
Irga2	Irga2 MGI:915200 MGI:1257137 MGI:1257136 New gene		AC132320 AC102225 XM_140378	AA968296 (C57BL/6, 5'EST, 100%, not full length) AA968378 (C57BL/6, 3'EST, 100%, not full length)	Inducible by IFN-γ.
Irga3	Irga3 New gene		AC132320 XM_140379 (C57BL/6J)	BY751179 (NOD, EST, not 100%, 610bp)	Inducible by IFN-γ.
Irga4	Irga4 New gene		AC132320) XM_140380 (Irgb4/Irgb5 tandem)	BY750970 (NOD, EST, nearly 100%, 700 bp) BU696309 (C57BL/6, EST, nearly 100%, 530 bp)	Inducible by IFN-γ.
Irga5Ψ	Irga5 New gene		AC132320	None	A transcript is inducible by IFN-γ but the coding sequence of the gene is disrupted repeatedly.
Irga6	Irga6 MGI:1926259 MGI:2147195 MGI:2147350	IIGP, IIGP1, Iigp1	AC135638	AJ007971 (C57BL/6, 100% correct) AF194871 (C57BL/6, also NM_021792, 100%) BC004649 (C57BL/6, cDNA 100%, 2330bp)	(Boehm et al., 1998) (MGI:1889878); (Zerrahn et al., 2002) Inducible by IFN-γ.
Irga7	Irga7 New gene		NT_039674 (C57BL/6J, Chr.18 genomic contig, 73.9 Mb) XM_487533 (C57BL/6, 100%)	None known	
Irga8	Irga8 MGI:953940 (C57BL/6) MGI:2384767 MGI:1489193 (CZECHII)	MGC:28198 BC023105	AC135638	BC023105 (CZECHII cDNA, = NM_145357, not 100%, full length) BB637466 (C57BL/6J, 5'EST, not 100%, not full length) BF163606 (CZECHII, not 100%, not full length) BE198503 (C57BL/6, 3'EST, 100%, not full length)	In C57BL/6 a non-canonical guanine after bp 849 in BC023105 (= aa 204) puts the sequence out of frame just before Helix H4; the reading frame is complete in BC023105 (CZECHII, *Mus musculus musculus*). Inducible by IFN-γ.

	New gene			BE198089 (C57BL/6, 3'EST, 100%, not full length) BX520309 (C57BL/6, 3'EST, 100%, not full length	
Irgb1	Irgb1 MGI:1519766 New gene		AL645849	BC022776 tandem Irgb2/ Irgb1 (CZECHII, not 100%, protein: Q8R5D8) BF144722 (CZECHII, EST, not 100%, starts with 3' end of Irgb2)	The *Irgb2/Irgb1* gene pair is almost certainly transcribed in tandem. The protein has not yet been described. Inducible by IFN-γ.
Irgb2	Irgb2 MGI:1518599 New gene		AL645849	BC022776 tandem Irgb2/ Irgb1 (CZECHII, not 100%, protein: Q8R5D8) BF144934 (CZECHII, 5' Irgb2 cDNA, not 100%) BY735436 (from cell line RCB-0527 Jyg-MC(B), strain unknown, 5' Irgb2, not 100%)	See note above, *Irgb1*.
Irgb3	Irgb3 MGI:1553791 (FVB/N) New gene		AL627237 AL669850 (unordered) AF060196 (129/SvJ, genomic, 1 bp difference, ATG(Irgb3)= bp 1353; Stop = bp 2659)	BF539106 (FVB/N, 3'EST, not 100%)	The genomic sequence of *Irgb3* is followed after 950 bp by a retroposon corresponding to the proteasome regulator PA28b (MGI:1331589). The presence or absence of this retroposon unambiguously distinguishes *Irgb3* from *Irgb4*.
Irgb4	Irgb4 MGI:1795392 MGI:3041173 New gene	9930111J21Rik	AL627237 AL669850 (unordered)	BC066104 (C57BL/6, Irgb5/Irgb4 tandem, 100%) BI655221 (NMRI, EST, not 100%)	See note above for *Irgb3*. *Irgb4* is probably normally expressed as a distinct 3' exon in a tandem transcript downstream of *Irgb5*.
Irgb5	Irgb5 MGI:3041173 MGI:2401562 New gene	9930111J21Rik	AL627237 AL645688 AL669850 (unordered)	BC066104 (C57BL/6, Irgb5/Irgb4 tandem; not 100% at 5' end) AK037088 (C57BL/6, cDNA, = NM_173434, 100%, unknown 5' end) (protein = BAC29698= Q8CB10)	*Irgb5* is probably normally expressed as a separate 5' exon in a tandem transcript upstream of *Irgb4*. However AK037088 does not splice into *Irgb4*. Thus *Irgb5* can exist as a single p47 unit or as a tandem with *Irgb4*. The reference number MGI:2401562 refers to several ESTs belonging to *Irgb5* and *Irgb9*. Inducible by IFN-γ.
Irgb6	Irgb6 MGI:98734 MGD-MRK-15077	TGTP, Mg21, Gtp2	AL627237 AL645688 AL669850 (unordered)	L38444 (C57BL/6, 100%) NM_011579 (NOD, 2 aa difference) U15636 (C.D2-Idh-1/Pep-3, 2 aa difference) BC085259 (NMRI, cDNA, 100%) BC034256 (CECHII, cDNA, not 100%)	(Carlow et al., 1998; Lafuse et al., 1995) Inducible by IFN-γ.
Irgb7Ψ	Irgb7 New gene		AL645688 AL669850 (unordered)	None known	Pseudogene: STOP codon before G-domain. Not inducible by IFN-γ, no known transcript.
Irgb8	Irgb8 MGI:1672892 New gene		AL645849	BG974191 (NMRI, 3' EST, not full length, not 100%,)	So similar to *Irgb1*, *b3* and *b4* that non-identical EST sequences are hard to disentangle.
Irgb9	Irgb9 MGI:2401562		AL645849 XM_204704 (C57BL/6, full	BB630182 (EST, short)	The reference number MGI:2401562 refers to several ESTs belonging to *Irgb5* and *Irgb9*.

	New gene		length, 100%)		
Irgb10	Irgb10 MGI:1282384		AL928857	AI122314 (C57BL/6, short EST, not 100%)	Short, terminates before end of G domain in S6. Inducible by IFN-γ.
Irgc	Irgc New gene	CINEMA	AC073810 (RP23-57J6) GENSCAN00000140134	BB615720 (C57BL/6 cDNA, 99%, 606 bp) 36 ESTs, none full length (e.g. CA464745 5'mRNA, 874bp, 100% except of first two bp)	An *Irgc*-related sequence has recently been named HGTP-47(MacMicking, 2004). This sequence (NP_950178=NM_199013= AK089224, NOD) contains 4 frameshifts relative to the C57BL/6 genomic sequence leading to a largely incorrect protein sequence. The reference numbers MGI:2685948 and MGI:2685320 both relate to this error sequence.
Irgd	Irgd MGI:99448 MGD-MRK-16217	IRG-47, IRG47, Ifi47, 47kDa, Iigp4	AL645688 AL669850 (unordered)	M63630 (B6D2F1, =NM_008330, 100% correct)	(Gilly and Wall, 1992). This is the first report of a p47 GTPase and has given its name (IRG-47) to the whole family. Inducible by IFN-γ.
Irgm1	Irgm1 MGI:107567 MGD-MRK-36139	LRG-47, LRG47, Ifi1, Iigp3	AL645849	U19119 (BALB/c, =NM_008326, 100% correct)	(Sorace et al., 1995); Two 5' splice variants exist. See notes human IRGM below. Inducible by IFN-γ.
Irgm2	Irgm2 MGI:1926262 MGI:2144195	GTPI Iigp2	AL928857	AJ007972 (C57BL/6; 100%) NM_019440 (CZECHII, = BC005419, not 100%)	(Boehm et al., 1998), MGI:1889878. Two 5' splice variants exist. Inducible by IFN-γ.
Irgm3	Irgm3 MGI:107729 MGD-MRK-36305 MGI:2144580	IGTP Igtp	AL928857	U53219 (C57BL/6, cDNA, 100%) NM_018738 (NOD, cDNA, not 100%)	(Taylor et al., 1996), MGI:82341 Inducible by IFN-γ.
Irgq	Irgq MGI:2667176 New gene	FKSG27	AC073810	AF322649 (C57BL/6, mRNA, = NM_153134)	
Human					
IRGC	UniGene Hs.515444 R30953_1 GeneID: 56269 New gene	CINEMA human IIGP5, cinema1	AC005622 HChr.19 cosmid	BC066939 (cDNA, 100%) NM_019612 (cDNA, 100%)	
IRGM	UniGene Hs.519680 New gene	human LRG-47-like protein (LRG47, LRG-	AC010441 Chr.5 XM_293893 (splice variant a,	BC038360 (splice variant c, 3'EST) BC038539 (short EST) BI764111 (short EST)	5 different 3' splice variants (a-e) (see main paper Bekpen *et al*, Fig. 8b). The orthology of *Irgm1* with human *IRGM* implied

	GeneID: 345611 MIM: 608212	47), IFI1	100%)	Sequences have been confirmed by RT-PCR (unpublished)	by use of the name *LRG47* or IFI1 for the human gene is incorrect. The use of LRG47 as a synonym or alias for human IRGM is therefore not recommended.
IRGQ	UniGene Hs.546476 GeneID: 126298	Homo sapiens FKSG27, Irgq1	AC006276	AF322648 (=NM_001007561 mRNA, 100%)	
Dog					
IRGB11	New gene		AACN010148430 AAEX1030324 AAEX1030325		
IRGB12	New gene		AACN01030937 AAEX1030324 AAEX1030325		Confirmed by RT-PCR but not sequenced. Inducible by IFN-γ.
IRGC	New gene	CINEMA	AACN010031536 AAEX01054272		
IRGD	New gene		AAEX01030325		
IRGM4	New gene		AAEX01059458		Confirmed by RT-PCR but not sequenced. Inducible by IFN-γ.
IRGM5	New gene		AACN010384735 AAEX01030325		Confirmed by RT-PCR but not sequenced. Inducible by IFN-γ.
IRGM6	New gene		AACN010300899 AAEX1030325		Confirmed by RT-PCR but not sequenced. Inducible by IFN-γ.
Fugu					
irgf5	New gene		Fugu_Sc2554 (Ensembl v3)		*irgf* genes of zebrafish, Fugu and Tetraodon have the long coding exon broken by an intron.
irgf6	New gene		Fugu_Sc2554 (Ensembl v3)	CA589084 (GI:25133662: 606 bp mRNA linear EST; hab53f04.y1 Fugu UT7 adult skin Takifugu rubripes cDNA clone) AL837863 (GI:21879801; 491 bp mRNA linear; F000A Takifugu rubripes cDNA clone F000A03aF7, mRNA sequence, skin)	See note above, *irgf5*
Tetraodon					
irgf7	New gene		SCAF112 (Ensembl v32, Jul 05)	GSTENT00000024001	*irgf* genes of zebrafish, Fugu and Tetraodon have the long coding exon broken by an intron.
irgf8	New gene		SCAF112 (Ensembl v32, Jul 05)	GSTENT00000023001	See note above, *irgf7*.
Zebrafish					
irge1	XP_693404		AL935330 (CH211-230C14)	BM316215 (3' EST)	Zebrafish *irge* genes have the long coding exon

	New gene		CR391937 (CH211-175G6)		unbroken by an intron, like the mammalian p47 genes XP_693404 (GI:68383735, 502 aa linear VRT 30-JUN-2005 predicted: similar to immunity-related GTPase family, cinema 1 [Danio rerio]. DBSOURCE REFSEQ: accession XM_688312.1 (Short N-terminus)
irge2	XP_693474 New gene		AL935330 (CH211-230C14) CR391937 (CH211-175G6)	None	See note above, *Irge1*. XP_693474 (GI:68383738, 352 aa linear VRT 30-JUN-2005 predicted: similar to immunity-related GTPase family, cinema 1 [Danio rerio]. DBSOURCE REFSEQ: accession XM_688382.1 (Short N-terminus)
irge3	New gene		AL935330 (CH211-230C14) CR391937 (CH211-175G6)	AW233145 (5' cDNA)	See note above, *Irge1*..
irge4	XP_693622 New gene		AL935330 (CH211-230C14) CR391937 (CH211-175G6)	CN501017 (5' EST) CK142408 (5' EST)	See note above, *Irge1*.. XP_693622 (GI:68383741, 385 aa linear VRT 30-JUN-2005 predicted: similar to immunity-related GTPase family, cinema 1 [Danio rerio]. DBSOURCE REFSEQ: accession XM_688530.1
Irge5	XM_681093 New gene		NW_635044 (GI:67045019; chr. 9 contig; bp 307225 308757)		See note above, *Irge1*. XM_681093 (GI:68365895, 1533 bp mRNA linear VRT 30-JUN-2005 predicted: Danio rerio similar to immunity-related GTPase family, cinema 1 (LOC557936), mRNA.
Irge6	XM_695163 New gene		NW_633868 (gi:67045754; chr. 18 contig; bp 5057602-5058696)		See note above, *Irge1*. XM_695163 (GI:68390584, 1095 bp mRNA linear VRT 30-JUN-2005 predicted: Danio rerio similar to immunity-related GTPase family, cinema 1 (LOC571560), mRNA.
irgf1	XP_700498 New gene		CR384077 DKEY-79I2	CN503005 (5' EST)	*irgf* genes of zebrafish, Fugu and Tetraodon have the long coding exon broken by an intron. XP_700498 (397 aa linear VRT 30-JUN-2005 predicted: similar to immunity-related GTPase family, cinema 1, partial [Danio rerio].
irgf2	New gene		CR384077 DKEY-79I2	None	See note above, *irgf1*.
irgf3	New gene		WGS traces zDH64-1061h13.q1k ZDH88-124d21.p1k	AL924569	See note above, *irgf1*.

			zfish35935-195b06.p1c		
irgf4	New gene		ENSDARG00000010545	None	See note above, *irgf1*.
irgg	New gene		AL935330 (CH211-230C14) CR391937 (CH211-175G6)	CA473205 (5' EST)	No intron in long coding exon. Short, terminates in Helix F. Probably the 5' end of a tandem with *irgq1*.
irgq1	New gene		AL935330 (CH211-230C14) CR391937 (CH211-175G6)	BQ481364 (5' EST) and BQ481122 (3' EST) from cDNA clone IMAGE:5899497. The 5' end of this clone is in the 3' end of irgg and reads into the 5' end of irgq1.	Short, terminates in helix F. Probably the 3' end of a tandem with *irgg*.
irgq2	XP_684591 New gene		BX072550 DKEY-245P1	BF938149 5' EST and BI880124 3'EST from cDNA clone IMAGE:4200886	XP_684591 (GI:68381188, 379 aa linear VRT 30-JUN-2005 predicted: similar to RGD1311107_predicted protein [Danio rerio]. DBSOURCE REFSEQ: accession XM_679499.1
irgq3	New gene		BX127973; SP6 end of BAC DKEY-279M7 Zv4_scaffold1709.9	None	
C. elegans					
C46E1.3	WP:CE34758 GI:3300129; CAE17750		AL008867.1 (GI:3217208, cosmid C46E1)	None	Predicted protein, tandem G domains.
W09C5.2	CAB63329.1 GI:6580259		Z82077 (GI:3873420, Cosmid W09C5)	None	Predicted protein.
Bacteria					
BAA10832	GI:1001345		BA000022.2 (GI:47118304, Synechocystis sp. PCC 6803)		Synechocystis sp . PCC 6803 Predicted protein.
BAA18140	GI:1653224		BA000022.2 (GI:47118304, Synechocystis sp. PCC 6803)		Synechocystis sp . PCC 6803 Predicted protein.
BAA18642	GI:1653731		BA000022.2 (GI:47118304, Synechocystis sp. PCC 6803)		Synechocystis sp . PCC 6803 Predicted protein.
BAC08557	GI:22294728		BA000039.2 (GI:47118315, T. elongatus BP-1)		Thermosynechococcus elongatus BP-1 Predicted protein.
BAC08842	GI:22295014		BA000039.2 (GI:47118315, T. elongatus BP-1)		Thermosynechococcus elongatus BP-1 Predicted protein.

V.2.Appendix Table 2. Splicing acceptors and donors for IRG proteins in mouse

Splicing junctions are indicated with blue. Genes alternatively spliced are indicated as (a) and (b). (2) indicates the splice variant specific to alternative promoter of respective gene.second promoter.

	First Splicing		Second Splicing	
Name of the Gene	Splicing Donor	Splicing Acceptor	Splicing Donor	Splicing Acceptor
Irgb1	GAGCAGGTGAGCTCA	TCTATTCAGCATCCT		
Irgb2	TGTGAGGTAAGGGT	CTCCCTTAGGTACCA		
Irgb5(a)	GGAACAGGTACCTGAA	TTTCTTTAGTACCATC		
Irgb5(b)	AGGACAAGGCAGGTAA	TTCTCTCCAGAGCACC	GGAACAGGTACCTGAA	CTTTCTTTAGTACCATC
Irgb6	CTGCTGAGGTAAGTGA	CTCCATTCAGCTTCTA		
Irgb7	CTGCTGAGGTAAGTGA	CCTCCCTTAGGTACCATT		
Irgd(a)	TCTTCACTGTGAGTACC	TCTCTGCTAGGGCTCAT	TGGATCTGGTGAGTGCG	CTTCTCACAGAGCTTCC
Irgd(b)	TGGATCTGGTGAGTGCG	AATTTTGCAGTAGTCTT	GAGGCCAGGTAGGCTG	CTTCTCACAGAGCTTCC
Irgd(c)	TGGATCTGGTGAGTGCG	CTTCTCACAGAGCTTCC		
Irgm1(a)	TGGATCAGGTAAGTAA	TATCTAATAGGGTTTGA	CCTGCGAGGTGGGGTAG	ACTCTTACAGGCTGCTC
Irgm1(b)	GGATCAGGTAAGTAAA	ACTCTTACAGGCTGCTC		
Irgb9	GGAACAGGTACCTGAG	CTCCCTTTAGTATCATC		
Irgm2(a)	TGAGCAGGTAGGTGAG	GTAATTTCAGGTTGCCC		
Irgm2(b)	TGAGCAGGTAGGTGAG	TTTAAGCCAGTTCTGGA	ATCCAGGGTGAGTCTT	GTAATTTCAGGTTGCCC
Irgm3	TGAGCAGGTAGGTGAG	TTTCTAACAGGTTCTGA	CTGGAAGGTGAGTTAG	TCTTCTGCAGACTTTTA
Irgb10	GGAGCTGGTGAGTGAG	TTCCCTCCAGTGTCCTG		
Irga1	GATTTCTGGTAACTCA	CTCCACACAGTGCAGCA	ATTTGTTGGTTTGTTT	TTTCTTCCAGTGGCTTT
Irga2	GATTTCTGGTAACTCA	TTTCTTCCAGTGCCTTT		
Irga3	AGTTTCTGGTAAGTGT	CTTCTTTCAGTGCCTTT		
Irga4(a)	AGTTTCTGGTAAGTGG	CTTCTTTCAGTAACCTT		
Irga6(a)	AGTTTCTGGTAAGTGG	TTTCTTCCAGTGCCTTT		
Irga6(b)	AGTTTCTGGTAAGTGG	TCAAAACAGGATTTCT	ATTAAAGGTAGGCTAT	TTTCTTCCAGTGCCTTT
Irga6 (2)	ATTTTCTGGTAACTCA	TTTCTTCCAGTGCCTTT		
Irga8(a)	AGTTTCTGGTAAGTGG	CTTCTTTCAGGGCCTTT		
Irga8(b)	AGCAGGGGTGGGTTCT	TCTGTTAAGATTTAAT	AGTTGAGGTATACCTA	CTTCTTTCAGGGCCTTT
Irgc	CATCTGAGGTAGGTTAG	TTCTCCTGCAGCCACT		

V.3.Appendix Table 3. Detailed analysis of Triple formation in Rat

AY321344 (mRNA amplified from rat liver). Blue highlighted seq. indicates the position of the splicing, yellow highlights show the charecteristics conserved a.a sequence for p47 GTPase.

```
   1  AGCTAGAAATGGGGAGAAAGAGAAAACAAACTTCTCGGCAATGAACAAGATCCTGTGT
   1  X  A  R  N  G  E  K  E  K  T  N  F  S  A  M  N  K  I  L  C
  61 ACAGCAAGGTTTGTGTGGGAACAAACACAGGACTGTCAAGAAGCACGCATGTTGACTGTG
  21  T  A  R  F  V  W  E  Q  T  Q  D  C  Q  E  A  R  M  L  T  V
 121 TTTAATTTAGTGCTGAAGAAACGGCCATGCCTGCGGCCCCCGTGGCTGCTGCCACCACTG
  41  F  N  L  V  L  K  K  R  P  C  L  R  P  P  W  L  L  P  P  L
 181 CTGCTCCTGCAGCTGCTTGCTGCTGCCCCAGCCAAGGCCGAAGCAAAGGAAGAGTCGGAG
  61  L  L  L  Q  L  L  A  A  A  P  A  K  A  E  A  K  E  E  S  E
 241 GAATCAGATGAGGACCAACCCACACTGCCTTACAACTTTGTAGTAAGCTGCGACTGTTGT
  81  E  S  D  E  D  Q  P  T  L  P  Y  N  F  V  V  S  C  D  C  C
 301 GTCTATGTGGCCGTGGGGATTGAACCTTGGACCCAGGGCAAGCTAGTGTCCTGTGCACTT
 101  V  Y  V  A  V  G  I  E  P  W  T  Q  G  K  L  V  S  C  A  L
 361 ACAGCCATGGGTCAGTCTTCTTCTAAACCTGATGCAAAGGCCCATAATATGGCCTCCAGC
 121  T  A  M  G  Q  S  S  S  K  P  D  A  K  A  H  N  M  A  S  S
 421 TTTAATGAGTTCTTCAAGAGTTTCAAAATGGAAAGTAAAATCCTTTCTGAGGAGACCATC
 141  F  N  E  F  F  K  S  F  K  M  E  S  K  I  L  S  E  E  T  I
 481 AATTCAATTCAGTCGTGTGTGCAAGAAGGAGACATACAGAAGGGAATTTCTATAATCAAT
 161  N  S  I  Q  S  C  V  Q  E  G  D  I  Q  K  G  I  S  I  I  N
 541 GCTGCCTTGGCAGACATTGAGAAGGCCCCCCTGAACATCGCAGTGACAGGGGAGACGGGG
 181  A  A  L  A  D  I  E  K  A  P  L  N  I  A  V  T  G  E  T  G
 601 GCAGGAAAGTCCACTTTCATTAATGCCCTGAGGGGAGTAGGGCATGAAGAGAGTGAGTCA
 201  A  G  K  S  T  F  I  N  A  L  R  G  V  G  H  E  E  S  E  S
 661 GCTAAGATTGGAGCAGTGGAGACAACCATGGATAAGTTCCCTAAGTTCCCTAACGTGACC
 221  A  K  I  G  A  V  E  T  T  M  D  K  F  P  K  F  P  N  V  T
 721 ATCTGGGACCTCCCTGGGGTCGGGACATGTAACTTCAAACCAGAAGAATATCTGAAGAAG
 241  I  W  D  L  P  G  V  G  T  C  N  F  K  P  E  E  Y  L  K  K
 781 CTGCGGTTCCAGGAGTATGACTTCTTCCTTATCATCTCAGCTACTCGCTTTAGAGAGAAT
 261  L  R  F  Q  E  Y  D  F  F  L  I  I  S  A  T  R  F  R  E  N
 841 GATGCCCAGCTGGCCAAAGCAATCAAAAAAATGAAAAAGAACTTCTATTTTGTTCGAACA
 281  D  A  Q  L  A  K  A  I  K  K  M  K  K  N  F  Y  F  V  R  T
 901 AAAATTGACAGTGATTTGTGGAATCAGAAGAAGTGTAAACCCAAGTCCTACAATAAGGAA
 301  K  I  D  S  D  L  W  N  Q  K  K  C  K  P  K  S  Y  N  K  E
 961 AAAATCCTGGAGGAAATTCGCAAAGACTGTGTGGAGAAGCTGCAGAACGCTCGGGTGGCC
 321  K  I  L  E  E  I  R  K  D  C  V  E  K  L  Q  N  A  R  V  A
1021 TCTGCTCGCGTCTTCTTAGTCTCCAGCGTTGAGGTAGCACAGTTTGACTTTCCTGAGCTG
 341  S  A  R  V  F  L  V  S  S  V  E  V  A  Q  F  D  F  P  E  L
1081 GAGTCCACCCTTTTGGAAGAGCTGCCAGCGCACAAGCGTCATGTCTTCATGCAGTGCCTC
 361  E  S  T  L  L  E  E  L  P  A  H  K  R  H  V  F  M  Q  C  L
1141 CCTAGCATTACCGAGAGGGCTATTGACCGCAGGAGAGATGCCCTGAGACAGAAGATCTGG
 381  P  S  I  T  E  R  A  I  D  R  R  R  D  A  L  R  Q  K  I  W
1201 TTGGAGGCTCTGAAGTATGGCGCGTCGGCCACCATCCCCATGATGTGTTTCTTCAATGAT
 401  L  E  A  L  K  Y  G  A  S  A  T  I  P  M  M  C  F  F  N  D
1261 GACATCGAGGAGCTTGAGAAGATCCTGACCCACTACAGGGGTAGCTTTGGGCTGGATGAC
 421  D  I  E  E  L  E  K  I  L  T  H  Y  R  G  S  F  G  L  D  D
1321 GAGTCGCTGAAAAACATGGCCAGTGAGTGGTCCATGTCTGTGGAGGAGCTGAAGTCCTTC
 441  E  S  L  K  N  M  A  S  E  W  S  M  S  V  E  E  L  K  S  F
1381 ATTAACTCACCCCATTTGCTGTCATGTGAGATGAATGAATCTGTGTCAGACAAGATGGTG
 461  I  N  S  P  H  L  L  S  C  E  M  N  E  S  V  S  D  K  M  V
1441 AAACCCTACCGGGCAGAGCTCTACCGGGTCACTATCCCCCAGCATAGAGCTGCCATCCAG
 481  K  P  Y  R  A  E  L  Y  R  V  T  I  P  Q  H  R  A  A  I  Q
1501 GATAGGACCTGGACAGGAGTGCAGAGAGTCACCTTTGTCCCAGGACAGCAGGAGACCAAG
 501  D  R  T  W  T  G  V  Q  R  V  T  F  V  P  G  Q  Q  E  T  K
1561 GAGGCAATTCCCTCAGAGCCACAGAAAGTCTCCATGTCACAGGGAGACAACTGGGGTGTT
 521  E  A  I  P  S  E  P  Q  K  V  S  M  S  Q  G  D  N  W  G  V
1621 TTTACCCCTTTCATAAACATGGCGAAACCTCTCAAGCCGCCATTGTTTAAATCCATCACT
 541  F  T  P  F  I  N  M  A  K  P  L  K  P  P  L  F  K  S  I  T
1681 GCTGGTGAGTCATCCTATAGCAGCCAGAACTCTTCTTCTCCAGAAGTCATTGAGAAGGTC
 561  A  G  E  S  S  Y  S  S  Q  N  S  S  S  P  E  V  I  E  K  V
1741 GGTAAGGCTGTGGCAGAGGGGGATTTACAGAAAGTGATATACACAGTCAAAGAGGAAATG
 581  G  K  A  V  A  E  G  D  L  Q  K  V  I  Y  T  V  K  E  E  M
1801 CAGAGTAAGTCTAGATACACGGTAAAAATCGCCGTGACTGGGGACTCTGGCAATGGCATG
 601  Q  S  K  S  R  Y  T  V  K  I  A  V  T  G  D  S  G  N  G  M
1861 TCATCTTTCGTCAACGCCCTTAGGCTCATTGGACATGAGGAGGAGGATTCAGCTCCCACT
 621  S  S  F  V  N  A  L  R  L  I  G  H  E  E  E  D  S  A  P  T
1921 GGGGTGGTGAGGACCACCCAGAAACCAGCCTGTTACTCCTCTTTCCACTTTCCCTATGTG
 641  G  V  V  R  T  T  Q  K  P  A  C  Y  S  S  F  H  F  P  Y  V
1981 GAGCTGTGGGACCTGCCTGGCACCGGGGTCACAGCCCAGAGCATGGAGAGCTACCTGGAT
 661  E  L  W  D  L  P  G  T  G  V  T  A  Q  S  M  E  S  Y  L  D
2041 GAGATGCAGTTCAGCGCATATGACCTTATCATCATCATTGCTTCTGAGCAGTTCAGCTCG
 681  E  M  Q  F  S  A  Y  D  L  I  I  I  I  A  S  E  Q  F  S  S
2101 AATCATGTGAAGCTGGCCGAAGCCATGCAGAGGATGAGAAAGAGGTTCTATGTCGTCTGG
 701  N  H  V  K  L  A  E  A  M  Q  R  M  R  K  R  F  Y  V  V  W
```

```
2161 ACCAAGCTGGACAGGGACATCAGCACAAGTACCTTCCCTGAACCCCAGCTCCTGCAGAGT
 721  T  K  L  D  R  D  I  S  T  S  T  F  P  E  P  Q  L  L  Q  S
2221 ATCCAAAAGAATATCAGGGAGAATCTCCAGAAGGCTCAGGTGAGGGACCCCCCCATATTT
 741  I  Q  K  N  I  R  E  N  L  Q  K  A  Q  V  R  D  P  P  I  F
2281 CTGGTCTCCTGCTTTAGTCCATCTTTTCACGACTTCCTAGACCTTAGAGAGACACTGCGA
 761  L  V  S  C  F  S  P  S  F  H  D  F  L  D  L  R  E  T  L  R
2341 AAAGACATCCACAACATCAGGTACAGAGATCCCTTAGAGACCCTTTCTCAAGTCTGCGAC
 781  K  D  I  H  N  I  R  Y  R  D  P  L  E  T  L  S  Q  V  C  D
2401 AAGTGCATCAACAATAAGGCGCTCTCTCTGAAGGAGGACCTGATGTTCACGAAACACCTG
 801  K  C  I  N  N  K  A  L  S  L  K  E  D  L  M  F  T  K  H  L
2461 GAGGCAGCTGTCAGCCCCCCGTATGATATTGCTGACCTGGAGAGGAGTCTGGACACCTAC
 821  E  A  A  V  S  P  P  Y  D  I  A  D  L  E  R  S  L  D  T  Y
2521 CAGAAGCTCTTTGGTGTGGATAATGAGTCACTTAGGAGGGTAGCTCAGAGTACAGGGAGA
 841  Q  K  L  F  G  V  D  N  E  S  L  R  R  V  A  Q  S  T  G  R
2581 CCAGAGATGAGCACCAGGGCCTTGCAGTTCCAGGACTTGATCAAGATGGACAGGAGACTG
 861  P  E  M  S  T  R  A  L  Q  F  Q  D  L  I  K  M  D  R  R  L
2641 AGGTTGATGATGTGTTTTGTCGTGAACATACTCCTCAGGGTTCTTGGAAGTCCATGGTGG
 881  R  L  M  M  C  F  V  V  N  I  L  L  R  V  L  G  S  P  W  W
2701 TTCGGCTTGTGGGATGTCGTTACCCGATACTTCAGACACCAGAGACAGAAGCGCATCATT
 901  F  G  L  W  D  V  V  T  R  Y  F  R  H  Q  R  Q  K  R  I  I
2761 GAAATAGTTGCTAAGAACACCAAGACCTCCTTGAGGAGAGCTCTAGAGGACTATACACTT
 921  E  I  V  A  K  N  T  K  T  S  L  R  R  A  L  E  D  Y  T  L
2821 CCTCCTGAAATCCTTTGTGAAGGCTGCCTGGAAGACAGCAGTTTTCTCCCGTTCACCTTT
 941  P  P  E  I  L  C  E  G  C  L  E  D  S  S  F  L  P  F  T  F
2881 GGAACACAGAGCCATGCCTACCTGGATGCTGCCATGCTTCTGGCCTTGCCTGAACCTCTG
 961  G  T  Q  S  H  A  Y  L  D  A  A  M  L  L  A  L  P  E  P  L
2941 AACCTGGGCCAAGAGGTGGGTGGCTTCCGGCAATTTCGTGCTGATAGCAGCAGGGGATTA
 981  N  L  G  Q  E  V  G  G  F  R  Q  F  R  A  D  S  S  R  G  L
3001 GAAAAAGAAACCCAGTTACTTGGGCTCTTTCTCTCACTGCCTGATAGAATTGGGTTAACT
1001  E  K  E  T  Q  L  L  G  L  F  L  S  L  P  D  R  I  G  L  T
3061 CTTTGTGAGCCAGTGAGAAAGGAGAAAAAAGGAAGCACACTGGGCTCCTGGGAGGAAGCC
1021  L  C  E  P  V  R  K  E  K  K  G  S  T  L  G  S  W  E  E  A
3121 ACTGACAATAAGCCTTCATCAGTGCTGGCCCCCGACTTGGGGAGGTTCTGGACCCCAGCC
1041  T  D  N  K  P  S  S  V  L  A  P  D  L  G  R  F  W  T  P  A
3181 GGCAAGACGGTGAAGTCTACATCCAGGGTTGCTCCATTGCTCACCAGCATGGAAGAAGCA
1061  G  K  T  V  K  S  T  S  R  V  A  P  L  L  T  S  M  E  E  A
3241 GTCGGGTTGCCCGAGGATAAACAGTTTGCATGCTTATCCGACGCTGTATTCATTTCCAAA
1081  V  G  L  P  E  D  K  Q  F  A  C  L  S  D  A  V  F  I  S  K
3301 GACAACAGTATTTTATCTGTAGAAGTCATCAAGAGTATTCAGGCTGCTGTGGCGGGAGGG
1101  D  N  S  I  L  S  V  E  V  I  K  S  I  Q  A  A  V  A  G  G
3361 AACGGGGTGGAAGTGGTCTCTATAGTTAAAGAGATTGTGCAGAAAGTATCCAGAACCACA
1121  N  G  V  E  V  V  S  I  V  K  E  I  V  Q  K  V  S  R  T  T
3421 ATGAAAATCGCTGTGACTGGGGACTCTGGCAATGGCATGTCATCTTTCGTCAACGCCCTT
1141  M  K  I  A  V  T  G  D  S  G  N  G  M  S  S  F  V  N  A  L
3481 AGGCTCATTGGACATGAGGAGGAGGATTCAGCTCCCACTGGGGTGGTGAGGACCACCCAG
1161  R  L  I  G  H  E  E  E  D  S  A  P  T  G  V  V  R  T  T  Q
3541 AAACCAGCCTGTTACTCCTCTTCCCACTTTCCCTATGTGGAGCTGTGGGACCTGCCTGGC
1181  K  P  A  C  Y  S  S  S  H  F  P  Y  V  E  L  W  D  L  P  G
3601 ATAGGGACCACAGCCCAGAGCATGGAGAGCTACCTGGATGAGATGCAGTTCAGCGCATAT
1201  I  G  T  T  A  Q  S  M  E  S  Y  L  D  E  M  Q  F  S  A  Y
3661 GACCTTATCATCATCATTGCTTCTGAGCAGTTCAGCTCGAATCATGTGAAGCTGGCCGAA
1221  D  L  I  I  I  I  A  S  E  Q  F  S  S  N  H  V  K  L  A  E
3721 GCCATGCAGAGGATGAGAAAGAAGTTCTATGTCGTCTGGACCAAGCTGGACAGGGACATC
1241  A  M  Q  R  M  R  K  K  F  Y  V  V  W  T  K  L  D  R  D  I
3781 AGCACAAGTACCTTCCCTGAACCCCAGCTCCTGCAGAGTATCCAAAAGAATATTAGGGAG
1261  S  T  S  T  F  P  E  P  Q  L  L  Q  S  I  Q  K  N  I  R  E
3841 AATCTCCAGAAGGGGAAAGTGAAGGAGCCCCCCATATTCCTGGTATCTATCATGAAGCCT
1281  N  L  Q  K  G  K  V  K  E  P  P  I  F  L  V  S  I  M  K  P
3901 TTATTACATGACTTCGAAAGGCTTAGGGAGACCCTACGGAAAGACCTCTCTGACATCAAG
1301  L  L  H  D  F  E  R  L  R  E  T  L  R  K  D  L  S  D  I  K
3961 TACCATGGTCTCTTAGAAACCCTTTACCAAATTTGTGAGAATACTATTAATGAGAGAGTA
1321  Y  H  G  L  L  E  T  L  Y  Q  I  C  E  N  T  I  N  E  R  V
4021 GAGTCCATTAAAAAGATCATAGATGAAAATAACCTACAAAGAGAGTTTGGAATCTTGACT
1341  E  S  I  K  K  I  I  D  E  N  N  L  Q  R  E  F  G  I  L  T
4081 CCAGACAACCTGACAGAGACTCGGAAAGTCTTCCAAGAAATCTTTGGTGTGGATGACCAA
1361  P  D  N  L  T  E  T  R  K  V  F  Q  E  I  F  G  V  D  D  Q
4141 TCTCTCAGCCAGGTGTCTCGGAGTATGGAAAAGCCAGATACACATTACAAGGCTAGCATA
1381  S  L  S  Q  V  S  R  S  M  E  K  P  D  T  H  Y  K  A  S  I
4201 GAGTCCCAGGAGATACAGGGGACCTCTGCTCCAGATCCTGGGAAGCGCACCCTCTTGGCT
1401  E  S  Q  E  I  Q  G  T  S  A  P  D  P  G  K  R  T  L  L  A
4261 TTCTCTGTGTTTACCCTTACGACACCAAACCCAGACACTATTGCGAATGCCAAGAAGTGC
1421  F  S  V  F  T  L  T  T  P  N  P  D  T  I  A  N  A  K  K  C
4321 ATGCTGACAGGAGCCTGAAATAAA
1441  M  L  T  G  A  *  N  K
```

V.4.Appendix Table 4. Accession numbers of p47 GTPases or p47 GTPases like sequences in vertebrates

Bos taurus
NCBI
gi|76641788|ref|XM_584684.2|
gi|76641786|ref|XM_868819.1|
gi|73586534|gb|BC102181.1|
gi|77736036|ref|NM_001034545.1|
gi|76641800|ref|XM_868949.1|

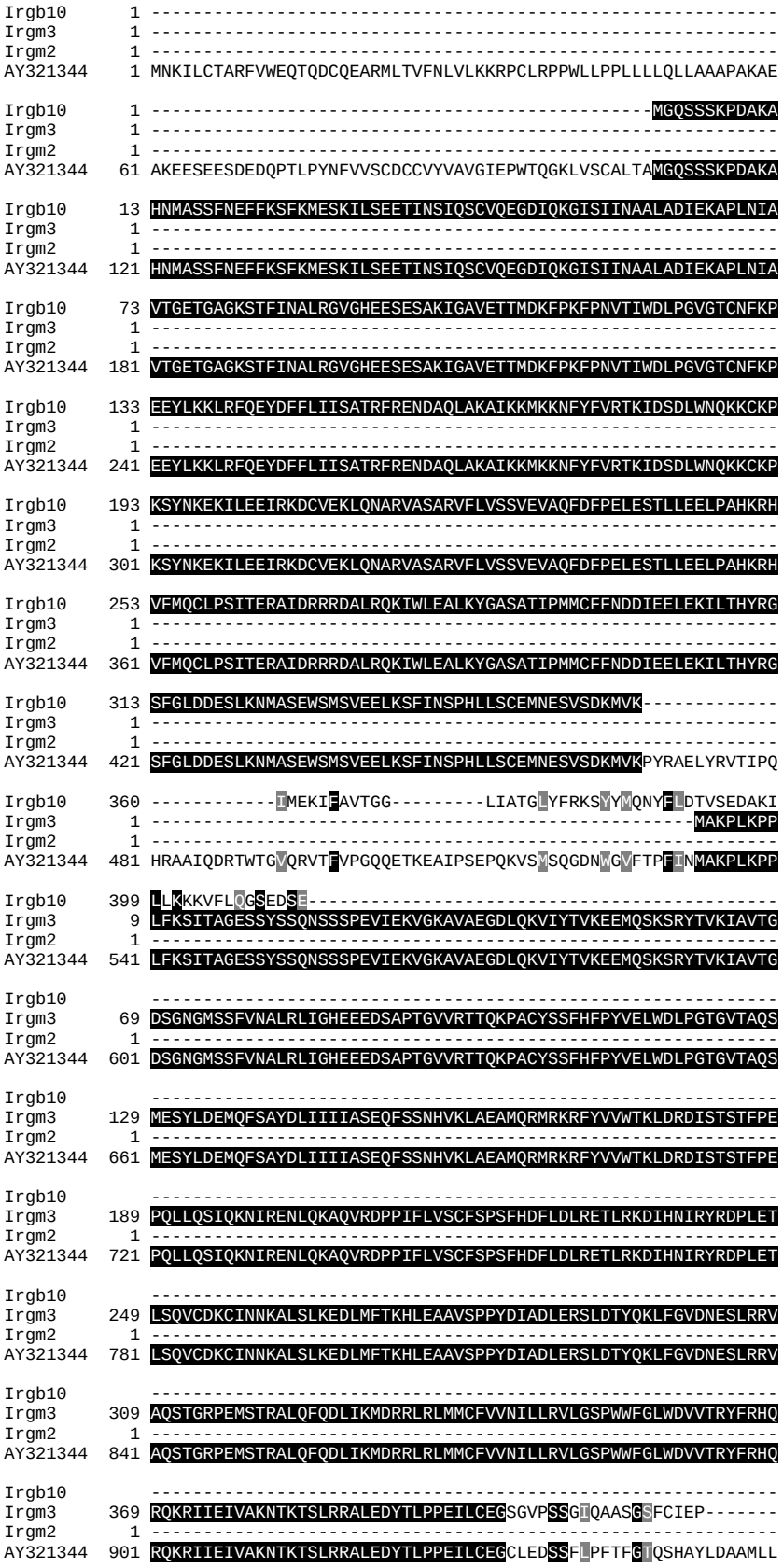
Gallus gallus
NCBI
gi|50749529|ref|XM_426495.1|
gi|46428515|emb|CR389870.1|
gi|46428036|emb|CR389391.1|

Ensemble
Contig2.1127
Contig2.1130
Contig40667.1
Contig2737.5
Contig2.1063
Contig2.1060
Contig42.364
Contig2.1057

Sus scrofa
AW435928
BI346828

Xenopus tropicalis
Scaffold_496

V.5.Appendix Figure 1. Multiple alignment of individual rat proteins (Irgb10,Irgm2, Irgm3) with rat tandem AY321344. Alignment was performed using BCM search launcher with default options and highlighted using Boxshade server version 3.21.



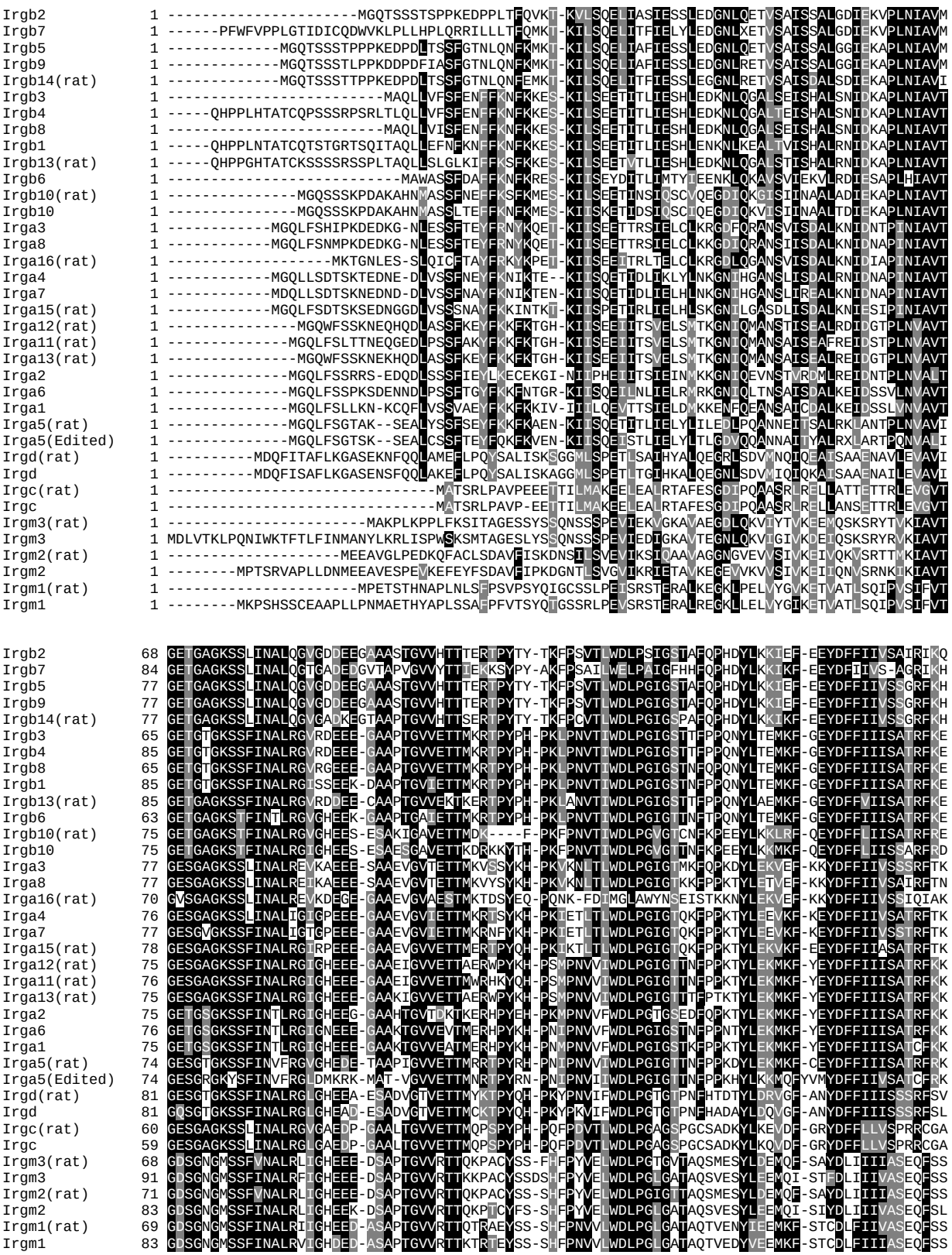
```
Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2       1 ------------------------------------------------------------
AY321344  961 ALPEPLNLGQEVGGFRQFRADSSRGLEKETQLLGLFLSLPDRIGLTLCEPVRKEKKGSTL

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2       1 ------------------------------------------MEEAVGLPEDKQFACLSD
AY321344 1021 GSWEEATDNKPSSVLAPDLGRFWTPAGKTVKSTSRVAPLLTSMEEAVGLPEDKQFACLSD

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2      19 AVFISKDNSILSVEVIKSIQAAVAGGNGVEVVSIVKEIVQKVSRTTMKIAVTGDSGNGMS
AY321344 1081 AVFISKDNSILSVEVIKSIQAAVAGGNGVEVVSIVKEIVQKVSRTTMKIAVTGDSGNGMS

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2      79 SFVNALRLIGHEEEDSAPTGVVRTTQKPACYSSSHFPYVELWDLPGIGTTAQSMESYLDE
AY321344 1141 SFVNALRLIGHEEEDSAPTGVVRTTQKPACYSSSHFPYVELWDLPGIGTTAQSMESYLDE

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2     139 MQFSAYDLIIIIASEQFSSNHVKLAEAMQRMRKKFYVVWTKLDRDISTSTFPEPQLLQSI
AY321344 1201 MQFSAYDLIIIIASEQFSSNHVKLAEAMQRMRKKFYVVWTKLDRDISTSTFPEPQLLQSI

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2     199 QKNIRENLQKGKVKEPPIFLVSIMKPLLHDFERLRETLRKDLSDIKYHGLLETLYQICEN
AY321344 1261 QKNIRENLQKGKVKEPPIFLVSIMKPLLHDFERLRETLRKDLSDIKYHGLLETLYQICEN

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2     259 TINERVESIKKIIDENNLQREFGILTPDNLTETRKVFQEIFGVDDQSLSQVSRSMEKPDT
AY321344 1321 TINERVESIKKIIDENNLQREFGILTPDNLTETRKVFQEIFGVDDQSLSQVSRSMEKPDT

Irgb10        ------------------------------------------------------------
Irgm3         ------------------------------------------------------------
Irgm2     319 HYKASIESQEIQGYQQDGWPLVWLHRPVIQFFSTGLDRVPCCFYSPHHRYTQQKGVLDET
AY321344 1381 HYKASIESQEIQGTSAPDPGKRTLLAFSVFTLITPNPDTIANAKKCMLTGA---------

Irgb10        ----------
Irgm3         ----------
Irgm2     379 AGKTKNFLWK
AY321344      ----------
```

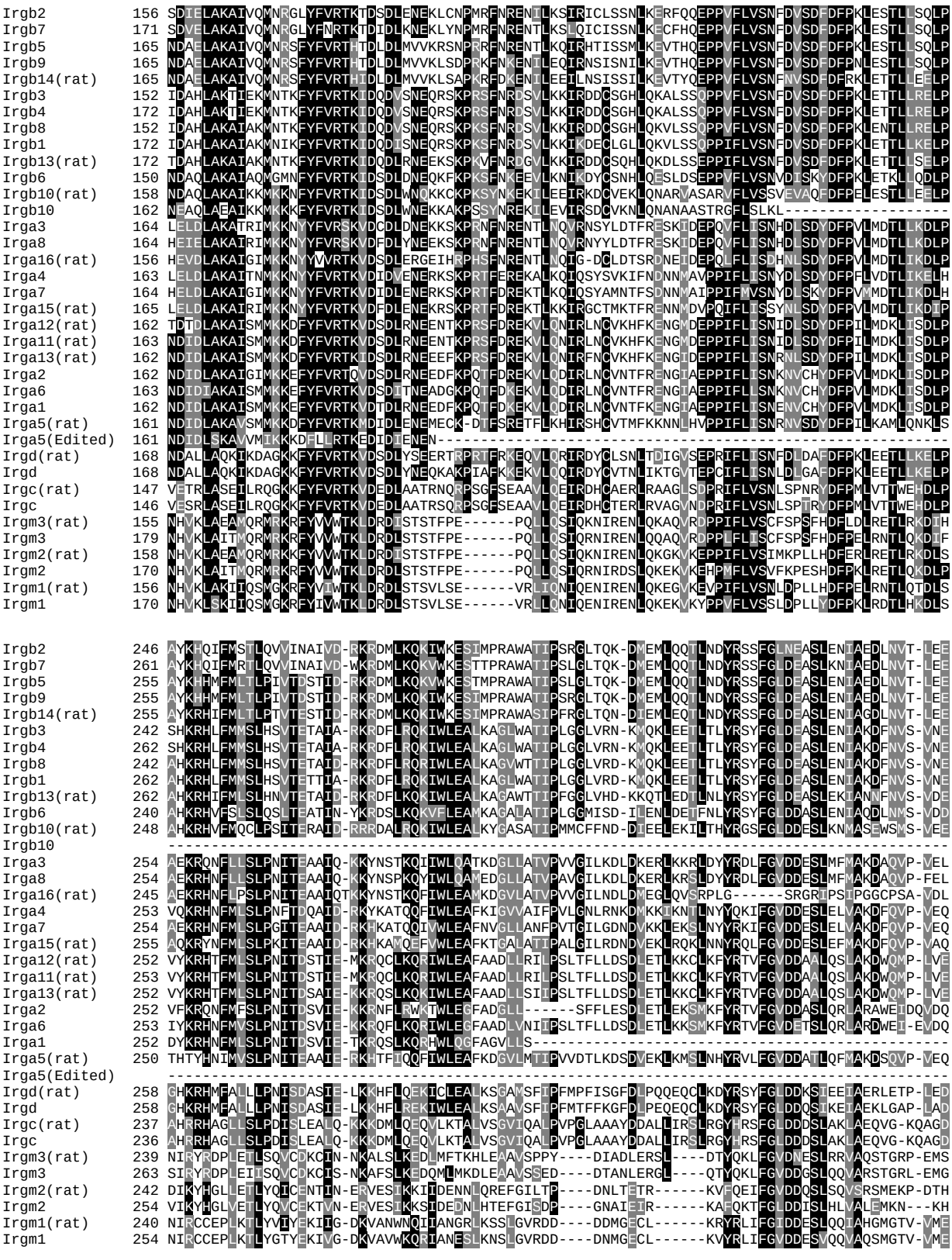
V.6.Appendix Figure 2. Multiple alignment of Czech II mouse p47 GTPases.

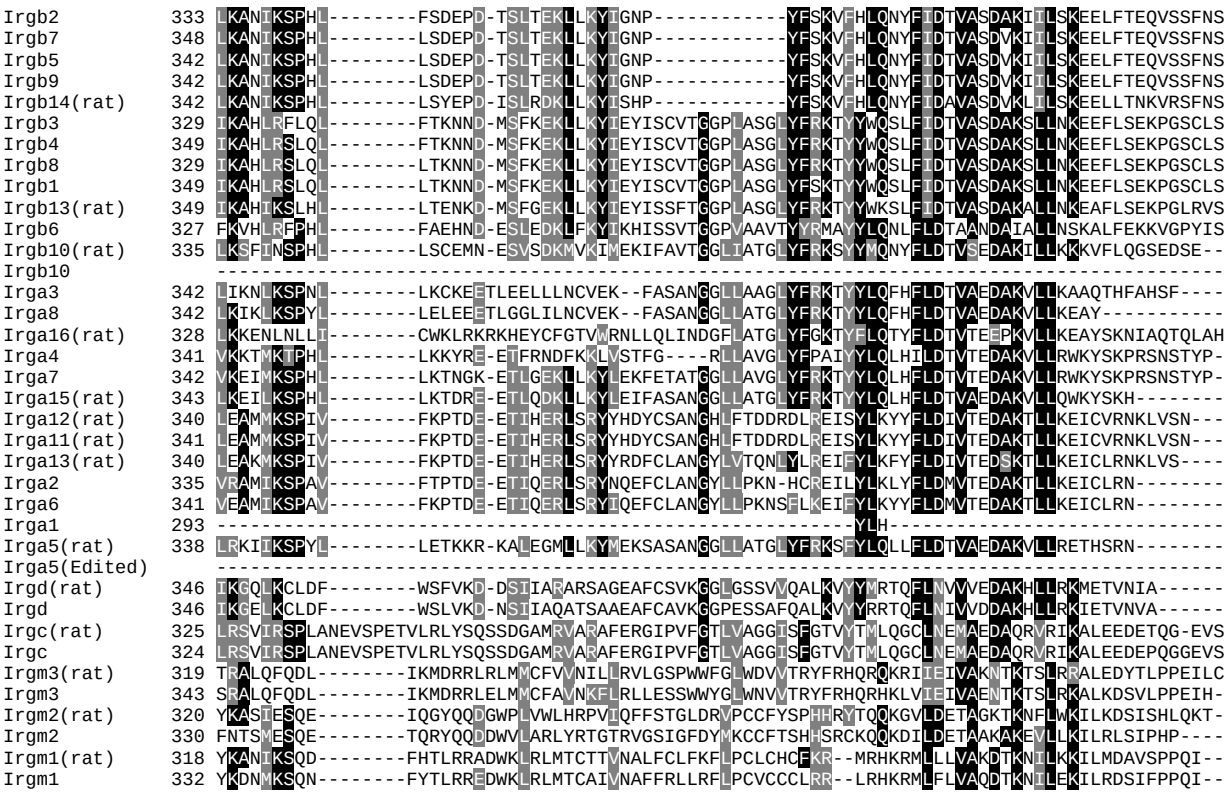
Alignment was performed using BCM search launcher with default options and highlighted using Boxshade server version 3.21. Irgm1 is excluded from alignment because it is partial sequence.

```
Irgb6     1 ----------------------------MAWASSFDAFFKNFKRESKIISEYDITLIMTYIEENKLQKAVSVIEKVLRDIEDAPLNIAVT
Irgb1     1 -------HPPLNTATCQTSTGRTSQITAQLLGFNFKNFFKSFKKESKILSEETILIESHLENKNLQGALTVISHALRNIDKAPLNIAVT
Irgb2     1 ----------------------MGQTSSSTSPPKEDPPLKFQVKTKVLSQELIASIESSLENGNLQETVSAISSALGDIEKAPLNIAVM
Irga9     1 --------------MGQLFS--SPQSEHQDLASSFTEYFKKFKTGNKIISQDIITLVELSLAKGNIQEANSAIRNALKEIDSTPLNVAVT
Irga10    1 --------------MGQLFS--SPQSEHQDLASSFTEYFKKFKMGNKIISQDIISLVELSVAKGNIQGANSAIKNALKEIDSTPLNVAVT
Irga8     1 --------------MGQLFSNMPKDEDKGNLESSFTEYFRNYKQETKIISEETTRSIELCLKKGDIQRANSIISDALKNIDNAPINIAVT
Irgm2     1 --------------------MEEAVESPEVKEFEYFSDAVFIPKDGNTLSVGVIKRIETAVKEGEIVKVVSIVKEIIQNVSRNKIKIAVT
Irgm3     1 MDLVTKLPQNIWKTFTLFINMANYLKRLISPWSKSMTAGESLYSSQNSSSPEVIEDIGKAVTEGNLQKVIGIVKDEIQSKSRYRVKIAVT
Irgd      1 ----------MDQFISAFLKSASENNFQQLAKEFLPQYSALISKAGGMLSPETLTGIHKALQEGNLSDVMSQIQKAISAAENAILEVAVI

Irgb6    63 GETGTGKSTFINALRGVGH-EEKDAAPTGAIETTMKRTPYPH-PKLPNVTIWDLPGIGSTTFTPQNYLTEMKFGEYDFFIIISATRFKEN
Irgb1    84 GETGTGKSSFINALRGISS-EEKDAAPTGVIETTMKRTPYPH-PKLPNVTIWDLPGIGSTNFPPQNYLTEMKFGEYDFFIIISATRFKEI
Irgb2    68 GETGTGKSSLINALQGVGDDEEGAAASTGIIHTTTERTPYTY-TKFPSVTLWDLPGIGSTAFQPHDYLKKIEFEEYDFFIIVSATRIRQS
Irga9    75 GESGSGKSSFINTLRGIGH-EEKGAAKTGVVEETMERHPYKH-PNMPNVVFWDLPGIGTQTFPPKTYLEKMKFYEYDFFIIISATRFKKN
Irga10   75 GESGSGKSSFINTLRGIGH-EEKGAAKTGVMEETMERHPYKH-PNMPNVVFWDLPGIGTTKFPPKTYLEKMKFYEYDFFIIISATRFKKN
Irga8    77 GESGAGKSSLINALREIKA-EEESAAEVGVTETTMKVYSYKH-PKVKNLTLWDLPGIGTKKFPPKTYLETVEFKKYDFFIIVSAIRFTNH
Irgm2    71 GDSGNGMSSFINALRLVGH-EEKDSAPTGVVRTTQKPTGYFS-SHFPYVELWDLPGLGATAQSVESYLEEMQISTYDLIIIVASEQFSSN
Irgm3    91 GDSGNGMSSFINALRFIGH-EEEDSAPTGVVRTTKKPAQYSSDSHFPYVELWDLPGLGATAQSVESYLEEMQISTFDLIIIVASEQFSSN
Irgd     81 GQSGTGKSSFINALRGLGH-EADESADVGTVETTMCKTPYQH-PKYPKVIFWDLPGTGTPNFHTDTYLDQVGFANYDFFIIISSSRFSLN

Irgb6   151 DAQLAKAIAQMGMNFYLVRTKIDSDLDNEQKFKPKSFNKEKLLKKIKDYCSNHLQESLYSEPLVFLVSNVDISKYDFPKLETKLLQDLPA
Irgb1   172 DAHLAKAIAKMNTKFYFVRTKIDQDVSNEQRSKPRSFNRDSVLKKIKDECLDLLQKVLSSQPPIFLVSNFDVSDFDFPKLETTLLKELPA
Irgb2   157 DIELAKAIVQMNRGLYFVKTKTDSDLENEKRVNPMRFNRENTLKRLRICISSNLKEWIQQEPPVFLVSNFDVSDFDFPKLESTLLSQLPA
Irga9   163 DIDLAKAISMMKKEFYFVRTKVDSDLNNEEDFKPQNFDREKVLQNIRLNCVNTFKENGIAEPQIFLVSNKNVCHYDFPVLIDRLISDLPL
Irga10  163 DIDLAKAISMMKKEFYFVRTKVDSDLNNEEDFKPQNFDREKVLQNIRLNCVNNFKENGIAEPPIFLVSNKNVCHYDFPVLMDKLISDLPV
Irga8   165 EIELAKAIRIMKKNYYFVRSKVDFDLYNEEKSKPRNFNRENTLNQVRNYYLDTFRESKIDEPQVFLISNHDLSDYDFPVLMDTLLKDLPA
Irgm2   159 HVKLAITVQRMRKRFYVVWTKLDRDLSTS------TFPEPQLLQSIQKNIRDSLQKEKVKEHPMFLVSVFKPESHDFPKLRETLQKDLPV
Irgm3   180 HVKLAITVQRMRKRFYVVWTKLDRDLSTS------TFPEPQLLQSIQRNIRENLQQAQVRDPPLFLISCFSPSFHDFPELRNTLQKDIFS
Irgd    169 DALLAQKIKDAGKKFYFVRTKVDSDLYNEQKAKPIAFKKEKVLCQIRDYCVTNLIKTGVTEPRIFLISNLDLGTFDFPELEETLLKELPG

Irgb6   241 HKRHVFSLSLQSLTEATINCKRDSLKQKVFLEAVKAGVLATIPLGGMIS-DILENLDETFNLYRSYFGLDDASLENVAKDLNVSVDDFKV
Irgb1   262 HKRHLFMMSLHSVTETTIARKRDFLRQKIWLEALKAGVWATIPLGGLVR-DKMQKLEEITLYRSYFGLDEASLENIAKDFNVSVNEIKA
Irgb2   247 YKHQIFMSTLQVVINAIVDRKRDMLKQKIWKESIMPRAWATIPSRELTQ-KDMEMLQQTLNDYRSSFGLDEASLENIAEDLNVTLEELKA
Irga9   253 YKRQNFMLSLPNITESAIEKKQQFLKQSIWLECFAPDLLSIKPSLAFIN-SDLETLEKSMKFYRSVFGVDEASLKSLATAWKIPVDQVEA
Irga10  253 YKRHNFMLSLPNITESAIEKKRQFLKQRIWLECFAADLMSIIPTLTFILDSDLETLKKSMKFYRTVFGVDEASLKSLATAWKIPVDQVEA
Irga8   255 EKRHNFLLSLPNITEAAIQKKYNSPKQYIWLQAMEDGLLATVPAVGILKDLDKERLKRSLDYYRDLFGVDDESLMFMAKDAQVPFELLKI
Irgm2   243 IKYIGLVEILYQVCEKTVNERVEFIKKS----------IDEDNLHTDFGISDPGNAIEIRKAFQKTFGLDDISLHLVSLEMKNKHFNTSM
Irgm3   264 IRYFDPLEIISQVCDKCISNKAFSLKEDQ------M--LMKDLEAAVSSEDDTANLERGLQTYQKLFGVDDGSLQQVARSTG--QTGDGL
Irgd    259 HKRHMFALLLPNISDASIELKKHFLREKIWLEALKSAAVSFIPFMTFFKGFDLPEQEQCLKDYRSYFGLDDQSIKEIAEKLGAPLADIKG

Irgb6   330 HLRFPHLFSEHNDESLEDKLFKYIKHISSVTGGPVAAVTYYRMAYYLQNLFLDTAANDAIALLNSKALFEKKVGPYISEPPEYWEA----
Irgb1   351 HLRSLQLLTKNNDMSFKEKLLKYIEYISCVTGGPLASGLYFRKTYYWQSLFIDTVASDAKSLLNKEEFLSEKPGSCLSDLPEYWETGMEL
Irgb2   336 NIKSPHLFSDEPDTSLTEKLLKYIG-------NP-----YFSKVFHLQNYFIDTVASDVKIILSKEELFTEQ------------------
Irga9   342 MMKSPAVFKPTDEETIQERLSRYVLEICFANRYLFRKNLFLRNVFYLKYYFLDMVTEDAKTLLKEIYLRNKLPSN---------------
Irga10  343 MMKSPAVFKPTDEETIQERLSRYVREFCLANGFLVTKNHYLREIFYLKYYFLDMVTEDAKTLLKEICLRNKLLSN---------------
Irga8   345 KLKSPYLLELEEETLGGLILN-CVEKFASANGGLLATGLYFRKTYYLQFHFLDTVAEDAKVLLKEAY----------------------
Irgm2   323 ESQETQRYQQDDWVLARLYRTGTRVGSIGFDYMKCCFSSHHSRCKQQKDILDETAAKAKEVLLKNPETLHSSSLGRFRQLESSPQPCTPC
Irgm3   344 QGSAVPGLDQD-------------GQETGADDVFCREQVPQGSRKLWVVWLVETLSPATSDTRGTSSSLK------------------
Irgd    349 ELKCLDFWSLVKDNSIIAQATSAAEAFCAVKGGPESSAFQALKVYYRRTQFLNIVVDDAKHLLRKIETVNVA------------------

Irgb6       ----------------------
Irgb1   441 ----------------------
Irgb2       ----------------------
Irga9       ----------------------
Irga10      ----------------------
Irga8       ----------------------
Irgm2   413 VPPVEITPLSPTPWGPSSLRLLG
Irgm3       ----------------------
Irgd        ----------------------
```

V.7.Appendix Figure 3. Multiple alignment of rat and mouse p47 GTPases.
Alignment was performed using BCM search launcher with default options and highlighted using Boxshade server version 3.21. rat Irga14 is excluded from alignment because it is highly degraded and was diffucult to align.

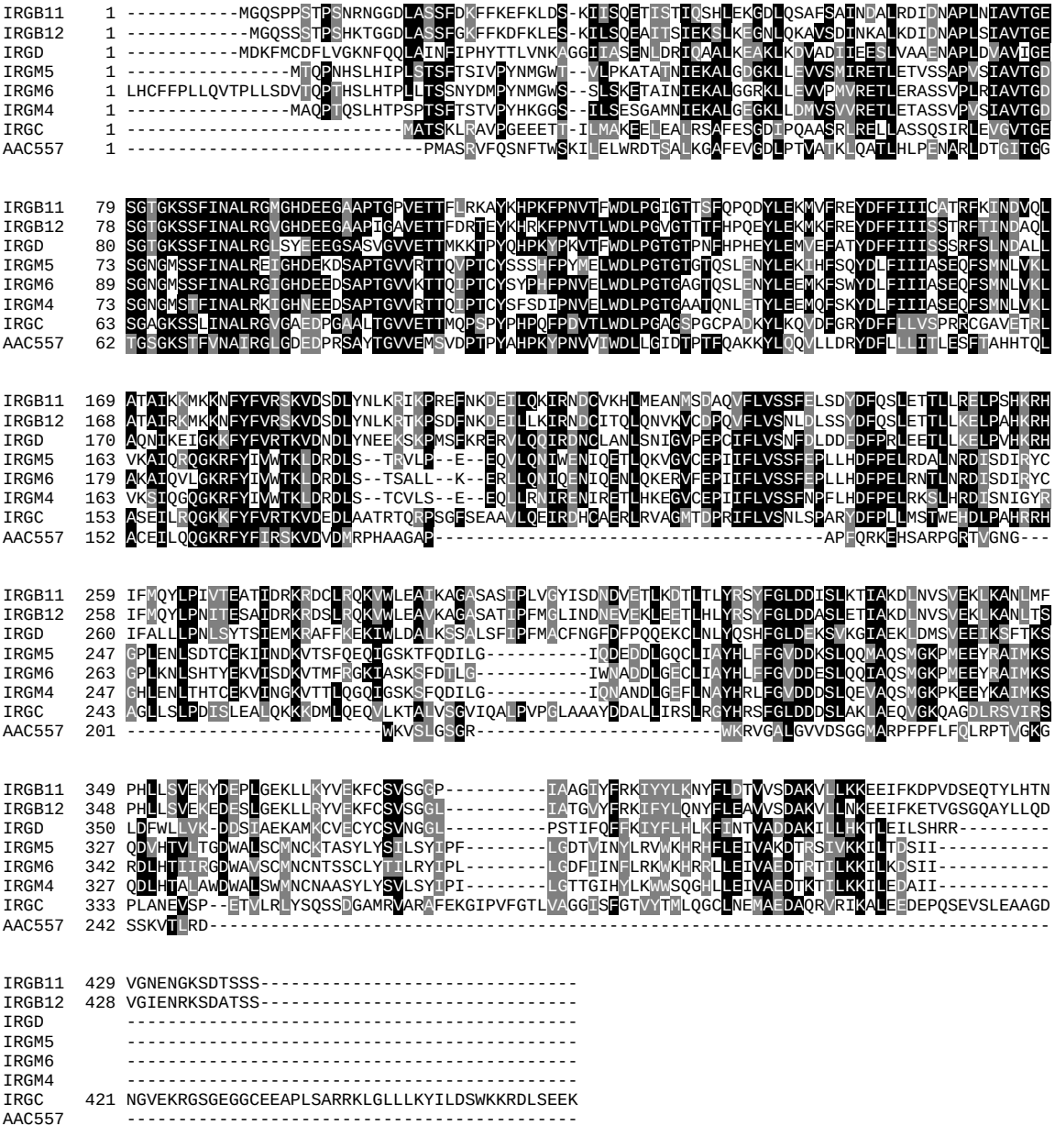





Irgb2 333 LKANIKSPHL--------FSDEPD-TSLTEKLLKYIGNP-----------YFSKVFHLQNYFIDTVASDAKIILSKEELFTEQVSSFNS
Irgb7 348 LKANIKSPHL--------LSDEPD-TSLTEKLLKYIGNP-----------YFSKVFHLQNYFIDTVASDVKIILSKEELFTEQVSSFNS
Irgb5 342 LKANIKSPHL--------LSDEPD-TSLTEKLLKYIGNP-----------YFSKVFHLQNYFIDTVASDVKIILSKEELFTEQVSSFNS
Irgb9 342 LKANIKSPHL--------LSDEPD-TSLTEKLLKYIGNP-----------YFSKVFHLQNYFIDTVASDVKIILSKEELFTEQVSSFNS
Irgb14(rat) 342 LKANIKSPHL--------LSYEPD-ISLRDKLLKYISHP-----------YFSKVFHLQNYFIDAVASDVKLILSKEELLTNKVRSFNS
Irgb3 329 IKAHLRFLQL--------FTKNND-MSFKEKLLKYIEYISCVTGGPLASGLYFRKTYYWQSLFIDTVASDAKSLLNKEEFLSEKPGSCLS
Irgb4 349 IKAHLRSLQL--------FTKNND-MSFKEKLLKYIEYISCVTGGPLASGLYFRKTYYWQSLFIDTVASDAKSLLNKEEFLSEKPGSCLS
Irgb8 329 IKAHLRSLQL--------LTKNND-MSFKEKLLKYIEYISCVTGGPLASGLYFRKTYYWQSLFIDTVASDAKSLLNKEEFLSEKPGSCLS
Irgb1 349 IKAHLRSLQL--------LTKNND-MSFKEKLLKYIEYISCVTGGPLASGLYFSKTYYWQSLFIDTVASDAKSLLNKEEFLSEKPGSCLS
Irgb13(rat) 349 IKAHIKSLHL--------LTENKD-MSFGEKLLKYIEYISSFTGGPLASGLYFRKTYYWKSLFIDTVASDAKALLNKEAFLSEKPGLRVS
Irgb6 327 FKVHLRFPHL--------FAEHND-ESLEDKLFKYIKHISSVTGGPVAAVTYYRMAYYLQNLFLDTAANDAIALLNSKALFEKKVGPYIS
Irgb10(rat) 335 LKSFINSPHL--------LSCEMN-ESVSDKMVKIMEKIFAVTGGLIATGLYFRKSYYMQNYFLDTVSEDAKILLKKKVFLQGSEDSE--
Irgb10 ----------
Irga3 342 LIKNLKSPNL--------LKCKEETLEELLLNCVEK--FASANGGLLAAGLYFRKTYYLQFHFLDTVAEDAKVLLKAAQTHFAHSF----
Irga8 342 LKIKLKSPYL--------LELEEETLGGLILNCVEK--FASANGGLLATGLYFRKTYYLQFHFLDTVAEDAKVLLKEAY----------
Irga16(rat) 328 LKKENLNLLI--------CWKLRKRKHEYCFGTVWRNLLQLINDGFLATGLYFGKTYFLQTYFLDTVTEEPKVLLKEAYSKNIAQTQLAH
Irga4 341 VKKTMKTPHL--------LKKYRE-ETFRNDFKKLYSTFG----RLLAVGLYFPAIYYLQLHILDTVTEDAKVLLRWKYSKPRSNSTYP-
Irga7 342 VKEIMKSPHL--------LKTNGK-ETLGEKLLKYLEKFETATGGLLAVGLYFRKTYYLQLHFLDTVIEDAKVLLRWKYSKPRSNSTYP-
Irga15(rat) 343 LKEILKSPHL--------LKTDRE-ETLQDKLLKYLEIFASANGGLLATGLYFRKTYYLQLHFLDTVAEDAKVLLQWKYSKH--------
Irga12(rat) 340 LEAMMKSPIV--------FKPTDE-ETIHERLSRYYHDYCSANGHLFTDDRDLREISYLKYYFLDIVTEDAKTLLKEICVRNKLVSN---
Irga11(rat) 341 LEAMMKSPIV--------FKPTDE-ETIHERLSRYYHDYCSANGHLFTDDRDLREISYLKYYFLDIVTEDAKTLLKEICVRNKLVSN---
Irga13(rat) 340 LEAKMKSPIV--------FKPTDE-ETIHERLSRYYRDFCLANGYLVTQNLYLREIFYLKFYFLDIVTEDSKTLLKEICLRNKLVS----
Irga2 335 VRAMIKSPAV--------FTPTDE-ETIQERLSRYNQEFCLANGYLLPKN-HCREILYLKLYFLDMVTEDAKTLLKEICLRN--------
Irga6 341 VEAMIKSPAV--------FKPTDE-ETIQERLSRYIQEFCLANGYLLPKNSFLKEIFYLKYYFLDMVTEDAKTLLKEICLRN--------
Irga1 293 ----------YLH----------
Irga5(rat) 338 LRKIIKSPYL--------LETKKR-KALEGMLLKYMEKSASANGGLLATGLYFRKSFYLQLLFLDTVAEDAKVLLRETHSRN--------
Irga5(Edited) ----------
Irgd(rat) 346 IKGQLKCLDF--------WSFVKD-DSIIARARSAGEAFCSVKGGLGSSVVQALKVYYMRTQFLNVVVEDAKHLLRKMETVNIA------
Irgd 346 IKGELKCLDF--------WSLVKD-NSIIAQATSAAEAFCAVKGGPESSAFQALKVYYRRTQFLNIVVDDAKHLLRKIETVNVA------
Irgc(rat) 325 LRSVIRSPLANEVSPETVLRLYSQSSDGAMRVARAFERGIPVFGTLVAGGISFGTVYTMLQGCLNEMAEDAQRVRIKALEEDETQG-EVS
Irgc 324 LRSVIRSPLANEVSPETVLRLYSQSSDGAMRVARAFERGIPVFGTLVAGGISFGTVYTMLQGCLNEMAEDAQRVRIKALEEDEPQGGEVS
Irgm3(rat) 319 TRALQFQDL--------IKMDRRLRLMMCFVVNILLRVLGSPWWFGLWDVVTRYFRHQRQKRIIEIVAKNTKTSLRFALEDYTLPPEILC
Irgm3 343 SRALQFQDL--------IKMDRRLELMMCFAVNKFLRLLESSWWYGLWNVVTRYFRHQRHKLVIEIVAENTKTSLRKALKDSVLPPEIH-
Irgm2(rat) 320 YKASIESQE--------IQGYQQDGWPLVWLHRPVIQFFSTGLDRVPCCFYSPHHRYTQQKGVLDETAGKTKNFLWKILKDSISHLQKT-
Irgm2 330 FNTSVESQE--------TQRYQQDDWVLARLYRTGTRVGSIGFDYVKCCFTSHHSRCKQQKDILDETAAKAKEVLLKILRLSIPHP----
Irgm1(rat) 318 YKANIKSQD--------FHTLRRADWKLRLMTCTTVNALFCLFKFLPCLCHCFKR--MRHKRMLLLVAKDTKNILKKILMDAVSPPQI--
Irgm1 332 YKDNMKSQN--------FYTLRREDWKLRLMTCAIVNAFFRLLRFLPCVCCCLRR--LRHKRMLFLVAQDTKNILEKILRDSIFPPQI--
Irgb2 402 KASPYREESVGKVFPVSPGSTFLFHFFEMFQSDSDKLCHVHVLLLLTSWGLSGETVT
Irgb7 417 KASPYREESVGEVFPVGPGSTFLFHFFEMFQSDSDKLCHVHVLLLLTSWGLSGETVT
Irgb5 411 KASLYREESVGKVFPVGPGSTFLFHFIEMFQSDSDELCHVHVLLLLTSGGLSSETVT
Irgb9 411 KASPYWEESVGKVFPVGPGSTFLFHFFEMFQSDSDKLCHVHVLLLLTSWGLSGETVT
Irgb14(rat) 411 NVSRQLF----------
Irgb3 410 DLPEYWETGMEL----------
Irgb4 430 DLPEYWETGMEL----------
Irgb8 410 DLPEYWETGMEL----------
Irgb1 430 DLPEYWETGMEL----------
Irgb13(rat) 430 DHTEYWEAGMEL----------
Irgb6 408 EPPEYWEA----------
Irgb10(rat) ----------
Irgb10 ----------
Irga3 ----------
Irga8 ----------
Irga16(rat) 410 SCRQLRDQRVNSRESLDKFML----------
Irga4 ----------
Irga7 ----------
Irga15(rat) ----------
Irga12(rat) ----------
Irga11(rat) ----------
Irga13(rat) ----------
Irga2 ----------
Irga6 ----------
Irga1 ----------
Irga5(rat) ----------
Irga5(Edited) ----------
Irgd(rat) ----------
Irgd ----------
Irgc(rat) 414 LEAAGDNAVEKRSSGEGTSEEAPLSTRRKLGLLLKYILDSWKRRDLSEDK-------
Irgc 414 LEAAGDNLVEKRSTGEGTSEEAPLSTRRKLGLLLKYILDSWKRRDLSEDK-------
Irgm3(rat) 401 EGSGVPSSGIQAASGSFCIEP----------
Irgm3 ----------
Irgm2(rat) ----------
Irgm2 ----------
Irgm1(rat) ----------
Irgm1 ----------

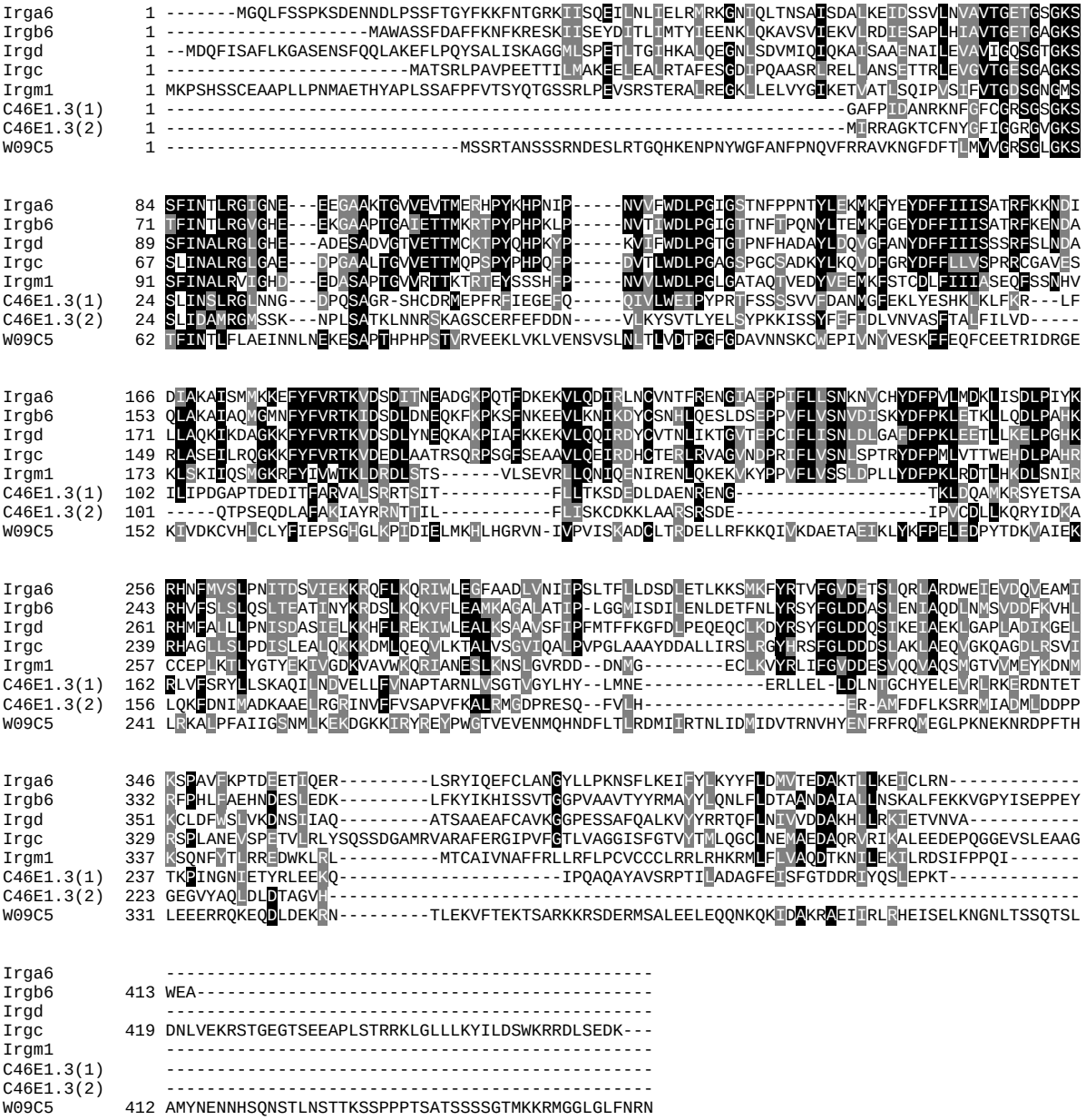
V.8.Appendix Figure 4. Multiple alignment of dog p47 GTPases.

Alignment was performed using BCM search launcher with default options and highlighted using Boxshade server version 3.21. AA557 is the abbreviation of (AACN010048557). Highly degraded pseudogene (AACN010088820) is excluded from the protein alignment.



V.9.Appendix Figure 5. Multiple alignment of C. elegans and classical mouse p47 GTPases.

Alignment was performed using BCM search launcher with default options and highlighted using Boxshade server version 3.21. C46E1.3 was devided into two sequence and edited to align properly with other p47 GTPases.



VI.REFERENCES

Altmann, S. M., Mellon, M. T., Johnson, M. C., Paw, B. H., Trede, N. S., Zon, L. I., and Kim, C. H. (2004). Cloning and characterization of an Mx gene and its corresponding promoter from the zebrafish, Danio rerio. Dev Comp Immunol *28*, 295-306.

Anderson, K. V., and Nusslein-Volhard, C. (1984). Information for the dorsal--ventral pattern of the Drosophila embryo is stored as maternal mRNA. Nature *311*, 223-227.

Anderson, S. L., Carton, J. M., Lou, J., Xing, L., and Rubin, B. Y. (1999). Interferon-induced guanylate binding protein-1 (GBP-1) mediates an antiviral effect against vesicular stomatitis virus and encephalomyocarditis virus. Virology *256*, 8-14.

Ausubel, F. M. (2005). Are innate immune signaling pathways in plants and animals conserved? Nat Immunol *6*, 973-979.

Bekpen, C., Hunn, J. P., Rohde, C., Parvanova, I., Guethlein, L., Dunn, D. M., Glowalla, E., Leptin, M., and Howard, J. C. (2005). The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. Genome Biol *6*, R92.

Benech, P., Vigneron, M., Peretz, D., Revel, M., and Chebath, J. (1987). Interferon-responsive regulatory elements in the promoter of the human 2',5'-oligo(A) synthetase gene. Mol Cell Biol *7*, 4498-4504.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., *et al.* (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416-422.

Beutler, B. (2004). Inferences, questions and possibilities in Toll-like receptor signalling. Nature *430*, 257-263.

Beutler, B. (2005). The Toll-like receptors: analysis by forward genetic methods. Immunogenetics, 1-8.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., *et al.* (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. Nature *391*, 485-488.

Bishop, J. G., Dean, A. M., and Mitchell-Olds, T. (2000). Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. Proc Natl Acad Sci U S A *97*, 5322-5327.

Boehm, U., Guethlein, L., Klamp, T., Ozbek, K., Schaub, A., Fütterer, A., Pfeffer, K., and Howard, J. C. (1998). Two families of GTPases dominate the complex cellular response to interferon-γ. J Immunol *161*, 6715-6723.

Boehm, U., Klamp, T., Groot, M., and Howard, J. C. (1997). Cellular responses to interferon-γ. Annu Rev Immunol *15*, 749-795.

Borst, P. (2002). Antigenic variation and allelic exclusion. Cell *109*, 5-8.

Burdon, J. J., and Thrall, P. H. (2003). The fitness costs to plants of resistance to pathogens. Genome Biol *4*, 227.

Carlow, D. A., Teh, S.-J., and Teh, H.-S. (1998). Specific antiviral activity demonstrated by TGTP, a member of a new family of interferon-induced GTPases. J Immunol *161*, 2348-2355.

Chang, K. C., Hansen, E., Foroni, L., Lida, J., and Goldspink, G. (1991). Molecular and functional analysis of the virus- and interferon-inducible human MxA promoter. Arch Virol *117*, 1-15.

Charles A. Janeway Jr., P. T., Mark Walport, Mark J. Shlomchik (2005). Immunobiology, the immune system in health and disease, In (New York: Garland Science Publising).

Cheng, Y. S., Becker-Manley, M. F., Chow, T. P., and Horan, D. C. (1985). Affinity purification of an interferon-induced human guanylate-binding protein and its characterization. J Biol Chem *260*, 15834-15839.

Cheng, Y. S., Patterson, C. E., and Staeheli, P. (1991). Interferon-induced guanylate-binding proteins lack an N(T)KXD consensus motif and bind GMP in addition to GDP and GTP. Mol Cell Biol *11*, 4717-4725.

Christoffels, A., Koh, E. G., Chia, J. M., Brenner, S., Aparicio, S., and Venkatesh, B. (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol Biol Evol *21*, 1146-1151.

Cicchetti, G., Maurer, P., Wagener, P., and Kocks, C. (1999). Actin and phosphoinositide binding by the ActA protein of the bacterial pathogen *Listeria monocytogenes*. J Biol Chem *274*, 33616-33626.

Collazo, C. M., Yap, G. S., Sempowski, G. D., Lusby, K. C., Tessarollo, L., Woude, G. F. V., Sher, A., and Taylor, G. A. (2001). Inactivation of LRG-47 and IRG-47 Reveals a Family of Interferon {gamma}-inducible Genes with Essential, Pathogen-specific Roles in Resistance to Infection. J Exp Med *194*, 181-188.

Collet, B., and Secombes, C. J. (2001). The rainbow trout (Oncorhynchus mykiss) Mx1 promoter. Structural and functional characterization. Eur J Biochem *268*, 1577-1584.

Comeron, J. M. (1999). K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics *15*, 763-764.

Costas, J., and Naveira, H. (2000). Evolutionary history of the human endogenous retrovirus family ERV9. Mol Biol Evol *17*, 320-330.

Crocker, P. R., Jefferies, W. A., Clark, S. J., Chung, L. P., and Gordon, S. (1987). Species heterogeneity in macrophage expression of the CD4 antigen. J Exp Med *166*, 613-618.

Dar, A., Munir, S., Vishwanathan, S., Manuja, A., Griebel, P., Tikoo, S., Townsend, H., Potter, A., Kapur, V., and Babiuk, L. A. (2005). Transcriptional analysis of avian embryonic tissues following infection with avian infectious bronchitis virus. Virus Res *110*, 41-55.

David, M. (2002). Signal transduction by type I interferons. Biotechniques *Suppl*, 58-65.

Di Cristofano, A., Strazullo, M., Longo, L., and La Mantia, G. (1995). Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. Nucleic Acids Res *23*, 2823-2830.

Dobzhansky, T. (1973). Nothing in biolgy makes sense except in the light of evolution. The American Biology Teacher *35*, 125-129.

Dobzhansky, T. (1982). Genetics and the Origin of Species (New York: Colombia University Press).

Dombrowski, J. E., and Raikhel, N. V. (1995). Isolation of a cDNA encoding a novel GTP-binding protein of Arabidopsis thaliana. Plant Mol Biol *28*, 1121-1126.

Dovhey, S. E., Ghosh, N. S., and Wright, K. L. (2000). Loss of interferon-gamma inducibility of TAP1 and LMP2 in a renal cell carcinoma cell line. Cancer Res *60*, 5789-5796.
Fang, F. C. (2004). Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. Nat Rev Microbiol *2*, 820-832.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution *39*, 783-791.

Field, C. M., al-Awar, O., Rosenblatt, J., Wong, M. L., Alberts, B., and Mitchison, T. J. (1996). A purified Drosophila septin complex forms filaments and exhibits GTPase activity. J Cell Biol *133*, 605-616.

Flajnik, M. F. (2004). Immunology: another manifestation of GOD. Nature *430*, 157-158.
Flajnik, M. F., and Du Pasquier, L. (2004). Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol *25*, 640-644.

Flor, H. H. (1971). Current Status of the gene for gene concept. Annual review of Phytopthology *9*, 275.

Floyd-Smith, G., Wang, Q., and Sen, G. C. (1999). Transcriptional induction of the p69 isoform of 2',5'-oligoadenylate synthetase by interferon-beta and interferon-gamma involves three regulatory elements and interferon-stimulated gene factor 3. Exp Cell Res *246*, 138-147.

Frank, S. A. (1994). Recognition and polymorphism in host-parasite genetics. Philos Trans R Soc Lond B Biol Sci *346*, 283-293.

Friedman, R. L., and Stark, G. R. (1985). alpha-Interferon-induced transcription of HLA and metallothionein genes containing homologous upstream sequences. Nature *314*, 637-639.

Fritig, B., Heitz, T., and Legrand, M. (1998). Antimicrobial proteins in induced plant defense. Curr Opin Immunol *10*, 16-22.

Galan, J. E., and Bliska, J. B. (1996). Cross-talk between bacterial pathogens and their host cells. Annu Rev Cell Dev Biol *12*, 221-255.

Ghosh, A., Uthaiah, R., Howard, J., Herrmann, C., and Wolf, E. (2004). Crystal structure of IIGP1: a paradigm for interferon-inducible p47 resistance GTPases. Mol Cell *15*, 727-739.

Gilly, M., Damore, M. A., and Wall, R. (1996). A promoter ISRE and dual 5' YY1 motifs control IFN-gamma induction of the IRG-47 G-protein gene. Genetics *179*, 237-244.

Gilly, M., and Wall, R. (1992). The IRG-47 gene is IFN-gamma induced in B cells and encodes a protein with GTP-binding motifs. J Immunol *148*, 3275-3281.

Gongora, C., Degols, G., Espert, L., Hua, T. D., and Mechti, N. (2000). A unique ISRE, in the TATA-less human Isg20 promoter, confers IRF-1-mediated responsiveness to both interferon type I and type II. Nucleic Acids Res *28*, 2333-2341.

Gordien, E., Rosmorduc, O., Peltekian, C., Garreau, F., Brechot, C., and Kremsdorf, D. (2001). Inhibition of hepatitis B virus replication by the interferon-inducible MxA protein. J Virol *75*, 2684-2691.

Guenzi, E., Topolt, K., Cornali, E., Lubeseder-Martellato, C., Jorg, A., Matzen, K., Zietz, C., Kremmer, E., Nappi, F., Schwemmle, M.*, et al.* (2001). The helical domain of GBP-1 mediates the inhibition of endothelial cell proliferation by inflammatory cytokines. EMBO J *20*, 5568-5577.

Gutierrez, M. G., Master, S. S., Singh, S. B., Taylor, G. A., Colombo, M. I., and Deretic, V. (2004). Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. Cell *119*, 753-766.

Haldane, J. B. S. (1949). Disease and Evolution. La Ricerca Scientifica *19 Supplemento Anno 19°*, 68-75.

Haller, O., Acklin, M., and Staeheli, P. (1987). Influenza virus resistance of wild mice: wild-type and mutant Mx alleles occur at comparable frequencies. J Interferon Res *7*, 647-656.

Haller, O., and Kochs, G. (2002). Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity. Traffic *3*, 710-717.

Hoegg, S., and Meyer, A. (2005). Hox clusters as models for vertebrate genome evolution. Trends Genet *21*, 421-424.

Hoffmann, J. A., Kafatos, F. C., Janeway, C. A., and Ezekowitz, R. A. (1999). Phylogenetic perspectives in innate immunity. Science *284*, 1313-1318.

Hong, Z., Bednarek, S. Y., Blumwald, E., Hwang, I., Jurgens, G., Menzel, D., Osteryoung, K. W., Raikhel, N. V., Shinozaki, K., Tsutsumi, N., and Verma, D. P. (2003). A unified nomenclature for Arabidopsis dynamin-related large GTPases based on homology and possible functions. Plant Mol Biol *53*, 261-265.

Hoshino, K., Takeuchi, O., Kawai, T., Sanjo, H., Ogawa, T., Takeda, Y., Takeda, K., and Akira, S. (1999). Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. J Immunol *162*, 3749-3752.

Hug, H., Costas, M., Staeheli, P., Aebi, M., and Weissmann, C. (1988). Organization of the murine Mx gene and characterization of its interferon- and virus-inducible promoter. Mol Cell Biol *8*, 3065-3079.

Janeway, C. A. (1989). Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harbor Symp Quant Biol *54*, 1-13.

Janeway, C. A., Jr., and Medzhitov, R. (2002). Innate immune recognition. Annu Rev Immunol *20*, 197-216.

Janzen, C., Kochs, G., and Haller, O. (2000). A monomeric GTPase-negative MxA mutant with antiviral activity. J Virol *74*, 8202-8206.

Jin, H., Yamashita, T., Ochiai, K., Haller, O., and Watanabe, T. (1998a). Characterization and expression of the Mx1 gene in wild mouse species. Biochemical Genetics *36*, 311-322.

Jin, H. K., Yamashita, T., Ochiai, K., Haller, O., and Watanabe, T. (1998b). Characterization and expression of the *Mx1* gene in wild mouse species. Biochemical Genetics *36*, 311-322.

Johnson, M. E., Viggiano, L., Bailey, J. A., Abdul-Rauf, M., Goodwin, G., Rocchi, M., and Eichler, E. E. (2001). Positive selection of a gene family during the emergence of humans and African apes. Nature *413*, 514-519.

Kaiser, F. (2005) Molecular Caracterization and Functional Analysis of Interferon-inducible 47 kDa GTPase IIGP, Ph.D, Berlin University, Berlin.

Kaiser, F., Kaufmann, S. H., and Zerrahn, J. (2004). IIGP, a member of the IFN inducible and microbial defense mediating 47 kDa GTPase family, interacts with the microtubule binding protein hook3. J Cell Sci *117*, 1747-1756.

Kayihan, G. C., Huber, D. A., Morse, A. M., White, T. L., and Davis, J. M. (2005). Genetic dissection of fusiform rust and pitch canker disease traits in loblolly pine. Theor Appl Genet *110*, 948-958.

Kimbrell, D. A., and Beutler, B. (2001). The evolution and genetics of innate immunity. Nat Rev Genet *2*, 256-267.

Klamp, T., Boehm, U., Schenk, D., Pfeffer, K., and Howard, J. C. (2003). A giant GTPase, VLIG-1, is inducible by interferons. J Immunol *171*, 1255-1265.

Kochs, G., Haener, M., Aebi, U., and Haller, O. (2002a). Self-assembly of human MxA GTPase into highly-ordered dynamin-like oligomers. J Biol Chem *in press*.

Kochs, G., and Haller, O. (1999). GTP-bound human MxA Protein interacts with the nucleocapsids of Thogoto virus (Orthomyxoviridae). J Biol Chem *274*, 4370-4376.

Kochs, G., Janzen, C., Hohenberg, H., and Haller, O. (2002b). Antivirally active MxA protein sequesters La Crosse virus nucleocapsid protein into perinuclear complexes. Proc Natl Acad Sci U S A *99*, 3153-3158.

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Selection in the evolution of gene duplications. Genome Biol *3*, RESEARCH0008.

Kotenko, S. V., Gallagher, G., Baurin, V. V., Lewis-Antes, A., Shen, M., Shah, N. K., Langer, J. A., Sheikh, F., Dickensheets, H., and Donnelly, R. P. (2003). IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex. Nat Immunol *4*, 69-77.

Kuhen, K. L., and Samuel, C. E. (1997). Isolation of the interferon-inducible RNA-dependent protein kinase Pkr promoter and identification of a novel DNA element within the 5'-flanking region of human and mouse Pkr genes. Virology *227*, 119-130.

Kumar, S., Mitnik, C., Valente, G., and Floyd-Smith, G. (2000). Expansion and molecular evolution of the interferon-induced 2'-5' oligoadenylate synthetase gene family. Mol Biol Evol *17*, 738-750.

Kumar, S., Tamura, K., and Nei, M. (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci *10*, 189-191.

Lafuse, W. P., Brown, D., Castle, L., and Zwilling, B. S. (1995). Cloning and characterization of a novel cDNA that is IFN-gamma-induced in mouse peritoneal macrophages and encodes a putative GTP-binding protein. J Leukoc Biol *57*, 477-483.

Lammas, D. A., Stober, C., Harvey, C. J., Kendrick, N., Panchalingam, S., and Kumararatne, D. S. (1997). ATP-induced killing of mycobacteria by human macrophages is mediated by purinergic P2Z(P2X7) receptors. Immunity *7*, 433-444.

Lanier, L. L. (1998). NK cell receptors. Annu Rev Immunol *16*, 359-393.

Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol *317*, 41-72.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet *20*, 116-122.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., and Hoffmann, J. A. (1996). The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell *86*, 973-983.

Lenhard, B., Sandelin, A., Mendoza, L., Engstrom, P., Jareborg, N., and Wasserman, W. W. (2003). Identification of conserved regulatory elements by comparative genome analysis. J Biol *2*, 13.

Lewontin, R. C. (1978). Adaptation. Sci Am *239*, 212-218, 220, 222 passim.

Lewontin, R. C. (2002). Directions in evolutionary biology. Annu Rev Genet *36*, 1-18.

Lindenmann, J., Lane, C. A., and Hobson, D. (1963). The Resistance of A2g Mice to Myxoviruses. J Immunol *90*, 942-951.

Ling, J., Pi, W., Bollag, R., Zeng, S., Keskintepe, M., Saliman, H., Krantz, S., Whitney, B., and Tuan, D. (2002). The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. J Virol *76*, 2410-2423.

Lower, R., Lower, J., and Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci U S A *93*, 5177-5184.

Lund, J., Sato, A., Akira, S., Medzhitov, R., and Iwasaki, A. (2003). Toll-like receptor 9-mediated recognition of Herpes simplex virus-2 by plasmacytoid dendritic cells. J Exp Med *198*, 513-520.

MacMicking, J., Taylor, G. A., and McKinney, J. (2003). Immune control of tuberculosis by IFN-gamma-inducible LRG-47. Science *302*, 654-659.

MacMicking, J. D. (2004). IFN-inducible GTPases and immunity to intracellular pathogens. Trends Immunol *25*, 601-609.

MacMicking, J. D., Nathan, C., Hom, G., Chartrain, N., Fletcher, D. S., Trumbauer, M., Steven, K., Xie, Q., Sokol, K., Hutchinson, N., *et al.* (1995). Altered responses to bacterial infection and endotoxic shock in mice lacking inducible nitric oxide synthase. Cell *81*, 641-650.

Martens S, P. I., Zerrahn J, Griffiths G, Schell G, Reichmann G, Howard JC. (2005). Disruption of *Toxoplasma gondii* Parasitophorous Vacuoles by the Mouse p47-Resistance GTPases. PLoS Pathog *1*, e24.

Martens, S., Sabel, K., Lange, R., Uthaiah, R., Wolf, E., and Howard, J. C. (2004). Mechanisms Regulating the Positioning of Mouse p47 Resistance GTPases LRG-47 and IIGP1 on Cellular Membranes: Retargeting to Plasma Membrane Induced by Phagocytosis 1. J Immunol *173*, 2594-606.

Martinsohn, J. T., Sousa, A. B., Guethlein, L. A., and Howard, J. C. (1999). The gene conversion hypothesis of MHC evolution: a review. Immunogenetics *50*, 168-200.

Mashimo, T., Glaser, P., Lucas, M., Simon-Chazottes, D., Ceccaldi, P. E., Montagutelli, X., Despres, P., and Guenet, J. L. (2003). Structural and functional genomics and evolutionary relationships in the cluster of genes encoding murine 2',5'-oligoadenylate synthetases. Genomics *82*, 537-552.

Mayr, E. (1978). Evolution. Sci Am *239*, 46-55.

Mayr, E. (2001). What evolution is (New York: Basic Books).

Medzhitov, R., and Janeway, C., Jr. (2000). Innate immunity. N Engl J Med *343*, 338-344.

Medzhitov, R., and Janeway, C. A. J. (1999). Innate immune induction of the adaptive immune response. Cold Spring Harb Symp Quant Biol *64*, 429-435.

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A., Jr. (1997). A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature *388*, 394-397.

Melen, K., Ronni, T., Broni, B., Krug, R. M., von Bonsdorff, C. H., and Julkunen, I. (1992). Interferon-induced Mx proteins form oligomers and contain a putative leucine zipper. J Biol Chem *267*, 25898-25907.

Mestas, J., and Hughes, C. (2004). Of mice and not men: differences between mouse and human immunology. J Immunol *172*, 2731-2738.

Miller, I. J., and Bieker, J. J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. Mol Cell Biol *13*, 2776-2786.

Modiano, N., Lu, Y. E., and Cresswell, P. (2005). Golgi targeting of human guanylate-binding protein-1 requires nucleotide binding, isoprenylation, and an IFN-gamma-inducible cofactor. Proc Natl Acad Sci U S A *102*, 8680-8685.

Monteiro, R. C., and Van De Winkel, J. G. (2003). IgA Fc receptors. Annu Rev Immunol *21*, 177-204.

Nathan, C., and Shiloh, M. U. (2000). Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. Proc Natl Acad Sci U S A *97*, 8841-8848.

Nathan, C. F., Murray, H. W., Wiebe, M. E., and Rubin, B. Y. (1983). Identification of interferon-gamma as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. J Exp Med *158*, 670-689.

Nelson, D. E., Virok, D. P., Wood, H., Roshick, C., Johnson, R. M., Whitmire, W. M., Crane, D. D., Steele-Mortimer, O., Kari, L., McClarty, G., and Caldwell, H. D. (2005). Chlamydial IFN-gamma immune evasion is linked to host infection tropism. Proc Natl Acad Sci U S A *102*, 10658-10663.

O'Brien, S. J. a. E., J.F. (1988). Interactive influence of infectious disease and genetic diversity in natural populations. Trends in Ecology & Evolution *3*, 254-259.

Ogura, Y., Bonen, D. K., Inohara, N., Nicolae, D. L., Chen, F. F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R. H., *et al.* (2001). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature *411*, 603-606.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K. R., Grimson, A., Hachiya, T., Hentze, M. W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell *12*, 1187-1200.

Ouellette, A. J., and Selsted, M. E. (1996). Paneth cell defensins: endogenous peptide components of intestinal host defense. Faseb J *10*, 1280-1289.

Pancer, Z., Amemiya, C. T., Ehrhardt, G. R., Ceitlin, J., Gartland, G. L., and Cooper, M. D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature *430*, 174-180.

Parvanova, I. A. (2005) Analysis of the role of the p47 GTPase IIGP1 in Resistance against Intracellular Pathogens, PhD., University of Cologne, Koeln.

Pearson, A. M. (1996). Scavenger receptors in innate immunity. Curr Opin Immunol *8*, 20-28.

Pfefferkorn, E. R. (1984). Interferon gamma blocks the growth of Toxoplasma gondii in human fibroblasts by inducing the host cells to degrade tryptophan. Proc Natl Acad Sci U S A *81*, 908-912.

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Huffel, C. V., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., *et al.* (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science, 2085-2088.

Praefcke, G., Geyer, M., Schwemmle, M., Kalbitzer, H., and Herrman, C. (1999). Nucleotide-binding characteristics of human guanylate-binding protein 1 (hGBP1) and identification of the third GTP-binding motif. J Mol Biol *292*, 321-332.

Praefcke, G. J. K., and McMahon, H. T. (2004). The dynamin superfamily: universal membrane tubulation and fission molecules? Nature Reviews Molecular Cell Biology *5*, 133-147.

Prakash, B., Renault, L., Praefcke, G. J. K., Wittinghofer, A., and Herrmann, C. (2000). Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. Nature *403*, 567-571.

Rausher, J. K. a. M. D. (2001a). Two modes of host-enemy coevolution. Population ecology *43*, 3-14.

Rausher, M. D. (2001b). Co-evolution and plant resistance to natural enemies. Nature *411*, 857-864.

Rehli, M. (2002). Of mice and men: species variations of Toll-like receptor expression. Trends Immunol *23*, 375-378.

Reich, N., Evans, B., Levy, D., Fahey, D., Knight, E., Jr., and Darnell, J. E., Jr. (1987). Interferon-induced transcription of a gene encoding a 15-kDa protein depends on an upstream enhancer element. Proc Natl Acad Sci U S A *84*, 6394-6398.

Reichelt, M., Stertz, S., Krijnse-Locker, J., Haller, O., and Kochs, G. (2004). Missorting of LaCrosse virus nucleocapsid protein by the interferon-induced MxA GTPase involves smooth ER membranes. Traffic *5*, 772-784.

Rigby, M. C., Hechinger, R. F., and Stevens, L. (2002). Why should parasite resistance be costly? Trends Parasitol *18*, 116-120.

Risso, A. (2000). Leukocyte antimicrobial peptides: multifunctional effector molecules of innate immunity. J Leukoc Biol *68*, 785-792.

Roach, J. C., Glusman, G., Rowen, L., Kaur, A., Purcell, M. K., Smith, K. D., Hood, L. E., and Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci U S A *102*, 9577-9582.

Robinson, C. M., Shirey, K. A., and Carlin, J. M. (2003). Synergistic transcriptional activation of indoleamine dioxygenase by IFN-gamma and tumor necrosis factor-alpha. J Interferon Cytokine Res *23*, 413-421.

Routledge, S. J., and Proudfoot, N. J. (2002). Definition of transcriptional promoters in the human beta globin locus control region. J Mol Biol *323*, 601-611.

Ryals, J., Dierks, P., Ragg, H., and Weissmann, C. (1985). A 46-nucleotide promoter segment from an IFN-alpha gene renders an unrelated promoter inducible by virus. Cell *41*, 497-507.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol *4*, 406-425.

Schaible, U. E., and Kaufmann, S. H. (2004). Iron and microbial infection. Nat Rev Microbiol *2*, 946-953.

Schopf, J. W. (1993). Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. Science *260*, 640-646.

Schug, J., and Overton, G. C. (1997). Modeling transcription factor binding sites with Gibbs Sampling and Minimum Description Length encoding. Proc Int Conf Intell Syst Mol Biol *5*, 268-271.

Schumacher, B., Bernasconi, D., Schultz, U., and Staeheli, P. (1994). The chicken Mx promoter contains an ISRE motif and confers interferon inducibility to a reporter gene in chick and monkey cells. Virology *203*, 144-148.

Schumacher, B., and Staeheli, P. (1998). Domains mediating intramolecular folding and oligomerization of MxA GTPase. J Biol Chem *273*, 28365-28370.

Schwemmle, M., Weining, K. C., Richter, M. F., Schumacher, B., and Staeheli, P. (1995). Vesicular stomatitis virus transcription inhibited by purified MxA protein. Virology *206*, 545-554.

Shivdasani, R. A., and Orkin, S. H. (1996). The transcriptional control of hematopoiesis. Blood *87*, 4025-4039.

Singh, G., and Lykke-Andersen, J. (2003). New insights into the formation of active nonsense-mediated decay complexes. Trends Biochem Sci *28*, 464-466.

Song, B. D., and Schmid, S. L. (2003). A molecular motor or a regulator? Dynamin's in a class of its own. Biochemistry *42*, 1369-1376.

Sorace, J. M., Johnson, R. J., Howard, D. L., and Drysdale, B. E. (1995). Identification of an endotoxin and IFN-inducible cDNA: possible identification of a novel protein family. J Leukoc Biol *58*, 477-484.

Staeheli, P., Grob, R., Meier, E., Sutcliffe, J., and Haller, O. (1988). Influenza virus-susceptible mice carry *Mx* genes with a large deletion or a nonsense mutation. Mol Cell Biol *8*, 4518-4523.

Staeheli, P., Prochazka, M., Steigmeier, P. A., and Haller, O. (1984). Genetic control of interferon action: mouse strain distribution and inheritance of an induced protein with guanylate-binding property. Virology *137*, 135-142.

Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H., and Schreiber, R. D. (1998). How cells respond to interferons. Annu Rev Biochem *67*, 227-264.

Staskawicz, B. J., Mudgett, M. B., Dangl, J. L., and Galan, J. E. (2001). Common and contrasting themes of plant and animal diseases. Science *292*, 2285-2289.

Summers, K., McKeon, S., Sellars, J., Keusenkothen, M., Morris, J., Gloeckner, D., Pressley, C., Price, B., and Snow, H. (2003). Parasitic exploitation as an engine of diversity. Biol Rev Camb Philos Soc *78*, 639-675.

Svensson, A. C., Raudsepp, T., Larsson, C., Di Cristofano, A., Chowdhary, B., La Mantia, G., Rask, L., and Andersson, G. (2001). Chromosomal distribution, localization and expression of the human endogenous retrovirus ERV9. Cytogenet Cell Genet *92*, 89-96.

Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T., Takeda, K., and Akira, S. (1999). Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity *11*, 443-451.

Tanaka, H., and Samuel, C. E. (1994). Mechanism of interferon action: structure of the mouse PKR gene encoding the interferon-inducible RNA-dependent protein kinase. Proc Natl Acad Sci U S A *91*, 7995-7999.

Taylor, G., Collazo, C., Yap, G., Nguyen, K., Gregorio, T., Taylor, L., Eagleson, B., Secrest, L., Southon, E., Reid, S., *et al.* (2000). Pathogen-specific loss of host resistance in mice lacking the IFN-gamma-inducible gene IGTP. PNAS *97*, 751-755.

Taylor, G. A. (2004). p47 GTPases: regulators of immunity to intracellular pathogens. Nature Reviews Immunology *4*, 100-109.

Taylor, G. A., Jeffers, M., Largaespada, D. A., Jenkins, N. A., Copeland, N. G., and Woude, G. F. (1996). Identification of a novel GTPase, the inducibly expressed GTPase, that accumulates in response to interferon gamma. J Biol Chem *271*, 20399-20405.

Tenen, D. G., Hromas, R., Licht, J. D., and Zhang, D. E. (1997). Transcription factors, normal myeloid development, and leukemia. Blood *90*, 489-519.

Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M., and Bergelson, J. (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature *423*, 74-77.

Trowsdale, J., and Parham, P. (2004). Mini-review: defense strategies and immunity-related genes. Eur J Immunol *34*, 7-17.

Uthaiah, R., Praefcke, G. J. M., Howard, J. C., and Herrmann, C. (2003). IIGP-1, a interferon-g inducible 47 kDa GTPase of the Mouse, is a slow GTPase showing co-operative enzymatic activity and GTP-dependent multimerisation. J Biol Chem *278*, 29336-29343.

Uthaiah, R. C. (2002) Biochemical, Structural and Cellular Studies on IIGP1, a Member of p47 Family of GTPases, PhD., University of Cologne, Koeln.

van der Bliek, A. M. (1999). Functional diversity in the dynamin family. Trends Cell Biol *9*, 96-102.

Varki, A. (2001). Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. Am J Phys Anthropol *Suppl 33*, 54-69.

Vestal, D. J. (2005). The guanylate-binding proteins (GBPs): proinflammatory cytokine-induced members of the dynamin superfamily with unique GTPase activity. J Interferon Cytokine Res *25*, 435-443.

Vivier, E., and Biron, C. A. (2002). Immunology. A pathogen receptor on natural killer cells. Science *296*, 1248-1249.

Wagner, A. (1998). The fate of duplicated genes: loss or new function? Bioessays *20*, 785-788.

Wagner, A. (2002). Selection and gene duplication: a view from the genome. Genome Biol *3*, reviews1012.

Warnock, D. E., Hinshaw, J. E., and Schmid, S. L. (1996). Dynamin self-assembly stimulates its GTPase activity. J Biol Chem *271*, 22310-22314.

Weinberg, J. B. (1998). Nitric oxide production and nitric oxide synthase type 2 expression by human mononuclear phagocytes: a review. Mol Med *4*, 557-591.

Wilkinson, M. F. (2005). A new function for nonsense-mediated mRNA-decay factors. Trends Genet *21*, 143-148.

Wilson, E. O. (2000). A global biodiversity map. Science *289*, 2279.

Wright, K. L., White, L. C., Kelly, A., Beck, S., Trowsdale, J., and Ting, J. P. (1995). Coordinate regulation of the human TAP1 and LMP2 genes from a shared bidirectional promoter. J Exp Med *181*, 1459-1471.

Yamanaka, R., Kim, G. D., Radomska, H. S., Lekstrom-Himes, J., Smith, L. T., Antonson, P., Tenen, D. G., and Xanthopoulos, K. G. (1997). CCAAT/enhancer binding protein epsilon is preferentially up-regulated during granulocytic differentiation and its functional versatility is determined by alternative use of promoters and differential splicing. Proc Natl Acad Sci U S A *94*, 6462-6467.

Yarovinsky, F., Zhang, D., Andersen, J. F., Bannenberg, G. L., Serhan, C. N., Hayden, M. S., Hieny, S., Sutterwala, F. S., Flavell, R. A., Ghosh, S., and Sher, A. (2005). TLR11 activation of dendritic cells by a protozoan profilin-like protein. Science *308*, 1626-1629.

Zerrahn, J., Schaible, U. E., Brinkmann, V., Guhlich, U., and Kaufmann, S. H. (2002). The IFN-inducible Golgi- and endoplasmic reticulum- associated 47-kDa GTPase IIGP is transiently expressed during listeriosis. J Immunol *168*, 3428-3436.

VII.SUMMARY

The interferon-inducible p47 GTPases are probably the most powerful resistance system in the mouse against intracellular pathogens. It is shown that the genome of the C57BL/6 mouse contains 23 p47 GTPase genes on chromosomes 7, 11 and 18 of which only 6 have previously been described. Among these are 2 probable pseudogenes. Of the 6 p47s thus far published, four have been knocked out and all of them have pathogen-sensitive phenotypes. By implication, others among this large family are also probably functional and required for normal pathogen resistance. Published differences in pathogen resistance profile, extensive divergent sequence evolution and radically differentiated intracellular behaviour suggest that the individual proteins have been selected for distinct functions, no doubt against distinct intracellular pathogens or pathogen classes.

Surprisingly, there are no reports of p47 GTPase function in human. The human genome has only one complete p47 GTPase gene, (*IRGC*) on chromosome 19, which is 90% identical at the protein level to mouse *Irgc*. *IRGC* is expressed in testis and syntenic between the two species. A p47 gene fragment (*IRGM*) is present on human chromosome 5 in a region syntenic to mouse chromosome 18 and mouse chromsome 11. This fragment is transcribed in 5 different spliced forms but no protein is detected. The expression profile of *IRGM* is regulated by a ERV9 retroviral elements containing promoter. Both of the human genes, *IRGC* and *IRGM*, are not induced by interferons. Therefore, human has no interferon-inducible p47 GTPase resistance system.

This different distribution of p47 GTPases in the two mammals has led to a broader investigation of the systematics of these interesting proteins. It is shown that the human has lost the immunologically functional members during mammalian evolution. The p47 GTPases are documented down to the dog, rat and bony fishes and shown that dramatic gain and also loss of the family member is going on in these ancient taxonomic groups as well. Variable number of p47 GTPases in different species is probably a mechanism to generate diversity of p47 GTPases acting on pathogens which are usually known to be fast evolving.

VIII.ZUSAMMENFASSUNG

Die Interferon induzierbaren p47-GTPasen sind vermutlich das wirkungsvollste Resistenzsystem gegen intrazelluläre Pathogene in der Maus. In dieser arbeit es wurde gezeigt, dass das Genom der C57BL/6 Maus dreiundzwanzig p47-GTPase-Gene, auf den Chromosomen 7, 11 und 18, enthält, von welchen sechs früher bereits beschrieben wurden. Zwei diser Gene sind wahrscheinlich Pesudogene. Vier der früher sechs publizierten p47-GTPasen wurden durch gezielte Mutagenese deaktiviert, die Phänotypen zeigten alle eine erhöhte Anfälligkeit gegenüber Pathogenen. In diesen Zusammenhang wird vermutet, dass auch die anderen Mitglieder dieser großen Familie funktional sind und für eine normale Resistenz gegenüber Pathogenen benötigt werden. Die veröffentlichten Unterschiede in der vermittelten Resistenz gegenüber Pathogenen, eine umfassende divergierende Sequenzevolution und radikale Unterschiede im intrazellulären Verhalten, lassen vermuten, dass die einzelnen Proteine für verschiedene Funktionen ausgewählt wurden, ohne Zweifel gegen verschiedene intrazelluläre Pathogene beziehungsweise Pathogenklassen.

Überaschenderweise, wurde keine p47-GTPase Funktion bei Menschen berichtet. Das menschliche Genom hat nur ein vollständiges p47-GTPase-Gen (IRGC) auf dem Chromosom 19, dieses ist auf dem Proteinlevel zu 90% mit Irgc aus der Maus identisch. IRGC wird in den Testes exprimiert und ist synthenisch in den beiden Arten. Ein p47-GTPase-Genfragment (IRGM) befindet sich auf dem menschlichen Chromosom 5, in einer Region welche synthenisch zu Maus Chromosom 18 und 11 ist. Dieses Fragment wird in fünf verschiedenen Spleißformen transkribiert, ein Protein wurde nicht detektiert. Das Expressionsprofil des IRGM wird durch einen Promotor reguliert, der ein ERV9 retrovirales Element enthält. Beide menschlichen Gene, IRGC und IRGM, sind durch Interferone nicht induzierbar. Deshalb besitzt der Mensch kein Interferon induzierbares p47-GTPase Resistenzsystem.

Die unterschiedliche Verteilung der p47-GTPasen in den zwei Säugern veranlasste eine umfassende Untersuchung der Systematik dieser interessanten Proteine. Es wurde gezeigt, dass der Mensch die in die Immunologie involvierten Mitglieder dieser Familie während der Säugerevolution verloren hat. Die p47-GTPasen wurden bis Hund, Ratte und den Knochenfischen dokumentiert. Es wurde ebenfalls ein dramatischer zugewinn und Verlust von Familienmitgliedern in dieser taxonomisch alten Gruppe gezeigt. Die variable Anzahl der p47-GTPasen in unterschiedlichen Arten ist vermutlich ein Mechanismus um Diversität in der Familie der p47-GTPasen, deren Mitglider auf gewöhnlich schnell evolvieren Pathogene einwirken, zu generieren.

IX.ACKNOWLEDGEMENT

I would like to thank Jonathan Howard not only supervising me from the beginning of this study but also showing and teaching me what real science is “we did something”. I also thank for his patience, help and discussions during my PhD and all through the tedious preparation of this text.

I would like to thank also Mathias Cramer who showed me how to generate antisera and for his help thorough all bureaucracy. It was pleasure to talk and work with you.

I would like to thank Maria Leptin, for sharing her data about fish p47 GTPases.

Many thanks to Revathy Uthaiah who was always there when I fell down and for very many pleasant moments, realisations, discussions, suggestions and for her help during the preparation of this thesis. Senior, it is and was fantastic to be friend of you.

To Ashwani, for being next to me all the time. It was full of enjoyment all this time that we spent together in cologne. Also for helping, encouraging, rescuing and philosophical discussions, I will not forget especially the one about the dream.

To Hicham, who encouraged me during the rough times. for his help and especially for the scientific discussions, suggestions, criticisms that we spent hours and also for delicious food and tea. It was really nice to discuss science with you.

To Dr. Ralf Max Leonhardt not only his wonderful scientific discussions, and suggestions but also for his help especially preparation of this text and providing wonderful environment during my Ph.D. I also express my sincere appreciation to his patience of listening my Turkish poems all day without any complains.

To Julia Hunn, for kindly providing her figure and correcting my mistakes all the time. also for her help in the lab. whenever I needed.

To Natasa Papic, for her help, and providing solutions, reagents, materials. Especially, I would like to thank for her understanding in hard times.

To Iana, for her friendship and for making life a lot easier and funny.

To Metodi, who was my flat mate in the beginning of this journey, for his help especially during the writing my thesis and helping, suggestions for critical experiments.

To Sascha for his scientific discussions and time for me, it was really nice to talk and spent time with you.

To Christoph, For providing nice environment in the lab. and for your helpful discussions and nice chess matches (40-3).

To Jia, for his help, especially during the preparation of this text.

To Michael, who help me through many situations.

To Yang and Niko, for their patience and helping me.

To Robert, my destiny friend (GMS story), thank you for your help

To Steffi, Gaby, Bettina, Rita, for everything, you made for me.

To Inci Togan, who was my previous supervisor, and Sezai Turkel, for teaching me and supporting me all the time.

To Luis, Leo, and Yen Yen for everything especially for the movie nights. I spent wonderful times with you.

To my friends, Abi, Gogce, Zeynep, Gaye, Irfan Hocam, Serdar, Goksu, Fatih, Ali (Personel), Saban, Murat, Thank you for keeping me alive and in balance.

To all my friends that I didn't mention here. Thank you.

I am grateful to my family; my mother Asiye, my uncles Metin, Cahit, Nedim, my aunt Miyase, Makbule my brother Erol and my best friends Mete, Senol, Murat for their support and encouragements.

X.ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt habe, die benutzen Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlauf oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solcheVeröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Jonathan C. Howard betreut worden.

Köln, Dec 2005.

Bekpen C, Hunn JP, Rohde C, Parvanova I, Guethlein L, Dunn DM, Glowalla E, Leptin M and Howard JC (2005). The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell-autonomous resistance mechanism in the human lineage. *Genome Biol.* 6, R92

XI.LEBENSLAUF

Name: Cemalettin BEKPEN

Geburtsdatum: 07-07-1974

Geburtsort: Bolu, Turkei

Staatsangehorigkeit: Turkisch

Familienstand: Ledig

Wohnort: Paul-Schalluck Strasse 25 WE 7,
50939 Köln

Instituts-Adresse: Institut für Genetik
Zülpicher Str. 47/ 50674 Köln

Ausbildung:

1981-1989 Grundschule, Bolu, Turkei

1989-1992 Mathematisch-Naturwissenschaftliches Gymnasium, Bolu, Turkei

1994-1995 Studium der Biologie an der Universitaet zu Hacettepe “Hacettepe Universitesi” Ankara, Turkei

1995-1998 Studium der Biologie an der Universitaet zu AIBU “Abant Izzet Baysal Universitesi” Bolu, Turkei

1999-2001 Diplomarbeit in Genetik an der bilogischen Fakultaet der Unversitaet zu ODTU “Orta Dogu Teknik Universitesi” Ankara, Turkei

ab WS 2001/2002 Promotionstudiengang Biologie (Fachrichtung Genetik) Institut fuer Genetik, Universitaet zu Koeln, Abteilung von Prof. Dr. Jonathan C. Howard