# EVOLUTIONARY DYNAMICS OF THE ALPHA-ACTININ GENE FAMILY

Thesis

Presented to the Graduate Council of
Southwest Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

Master of Science

by

Jamie D. Dixson

San Marcos, Texas

December, 2001

Evolutionary Dynamics of the Alpha-Actinin Gene Family


1478 Stallion Springs Drive
Fischer, Texas 78623
Jamie.dixson@excite.com



Printed in the United States of America

ACKNOWLEDGEMENTS

I would like to thank all of those who supported me in my undertaking of this study which includes, but is not limited to Dr. Dana M. García, who allowed me as an inexperienced graduate student to join her lab and work on this project. She was unwavering in her intellectual contribution and insight, constructive criticism and skepticism. I truly feel that her contribution was pivotal in the culmination of this study. Dr. Michael R. J. Forstner too was invariable in his support of my studies. He provided the "sanity support" which only someone with a similar background as myself could provide. His ingenuity and understanding of genetics is truly vast, and it was that understanding which led me to strive for a deeper knowledge of genetics. Dr. Joseph R. Koke was also a strong influence on my studies. Through watching Dr. Koke in class and in his dealings with research questions I learned that not all things are as obvious as they seem and that it is not bad to question others for their own benefit. I would also like to thank my family which has stood behind me through 21 years of school. My wife, Joy, deserves as much credit for the work which gave rise to this thesis as I do. She was there when frustration turned to anger and was always willing to listen; she sacrificed many sleepless nights wondering when I would be home, and she took care of things around the house when I was away. Ultimately I would like to thank the Lord for the guidance and the drive which He has given me. There have been many doors which would not have opened for me without His hand.

This work was supported in part by faculty research enhancement grants awarded to Dana M. García, Joseph R. Koke and M. R. J. Forstner by Southwest Texas State University, NSF grant #IBN0077666 and a NIH/Bridges grant #GM58375-01A1.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

**Abstract**

The alpha-actinin gene family is a member of the spectrin superfamily of proteins. Other members of this superfamily include the α- and β-spectrins, $\beta_{heavy}$ spectrin and dystrophin. Each member of this superfamily has a unique actin cross-linking distance consequent to the specific number of spectrin repeats within the central repeat region of the superfamily. Alpha-actinin is thought to be the least derived of the spectrin superfamily due to its possession of the smallest number of spectrin repeats. Though many α-actinins are known, the relationships among them are as yet undetermined. I performed phylogenetic analyses on α-actinin sequences to establish the kinship among α-actinin isoforms and to taxonomically treat the sequences. In addition, I sequenced and analyzed a previously unknown rat α-actinin 3. Analyses support four main isoforms of α-actinin. I also investigated homogenizing evolution within the spectrin repeats of the gene family. This was accomplished using gene conversion analyses and a new analysis which implements a continuum of the two known modes of homogenizing evolution: birth-and-death evolution and concerted evolution. This conceptual leap to a continuous model has not previously been made. Furthering these generalized studies of the α-actinin gene family, I used RT-PCR, cloning and automated DNA sequencing to obtain 77bp of sequence for what may be yet another member of the spectrin superfamily, the G3.5 antigen.

*Chapter 1*

# THE EVOLUTION OF ALPHA-ACTININ


## Summary

**Genomic characterization requires much more than sequencing efforts. Post-sequencing analyses often focus on groups of similar genes. Gene families comprise large portions of the genomes of higher vertebrates. The rate of genomic data production exceeds the ability of the scientific community to analyze it; therefore a current reappraisal of gene family evolution is underway. This reappraisal will clarify relatedness among and between gene families as well as help to decipher the modes of gene family evolution. One such family, the α-actinin gene family is part of the spectrin superfamily. There are four known loci which encode α-actinins 1, 2, 3 and 4. Of the eight domains in α-actinin, the actin-binding domain is the most highly conserved. The spectrin repeats of the α-actinin gene family are believed to have undergone concerted evolution, a mode of evolution which results in homogenization of homologous sequences. Here we present evidence gained through phylogenetic analyses that of the four α-actinins, α-actinin 2 is the least derived followed by α-actinin 3 and then α-actinins 1 and 4. Such**

**resolution of the kinship among α-actinins may allow for future genetic treatment directed at members of this gene family which have been associated with cancer and nemaline myopathy in humans. In addition we have resolved the kinship among the four spectrin repeats found in α-actinin(s). Our results were surprising in two ways: 1) The apparent relatedness of the α-actinin genes appears to deviate from parsimony regarding the functional gene product and 2) the order of divergence of the spectrin repeats best supported by my data would have had a low probability of occurrence.**

## Introduction

Advances in the genetic analyses of both functional and evolutionary aspects of gene families have been catalyzed by the completion of the human genome and the genomes of other organisms. Gene families comprise a large percentage of genomic content and their intra- and inter-specific divergence accounts for a comparable amount of diversity. In higher vertebrates, gene families and/or gene super-families comprise much of the functional portion of the genome (Hughes, 1999, pp. 143 and 212). It is necessary for both applied and theoretical work that we understand the mode(s) and tempo by which gene families evolve. In order to maintain high levels of fidelity and efficacy, applied biomedical engineering should target recently duplicated genes for genetic alteration when seeking to cure disease associated with the altered product of a gene pair. Such cures would be possible due to the fact that at the point of duplication and for a period of time thereafter, the duplicate gene is superfluous

to the original copy. This situation provides an avenue for genetic engineering while limiting risks associated with inactivation of single copy genes.

An emerging theory states that gene families may maintain high levels of protein homogeneity and low levels of nucleotide homogeneity through a mode of evolution in which genes undergo frequent duplication and subsequent purifying selection. This form of gene family evolution has been termed birth-and-death evolution (Gojobori and Nei, 1984; Nei et al., 2000, 1997; Ota and Nei, 1994). In families where this mode of evolution exists, one might expect to find pseudogenes suitable for the type of treatment proposed above. However, within many gene families there is a degree of inter-locus similarity which cannot be explained by homology alone as many of the loci are known to be the result of ancient divergences. In order to explain this similarity many authors have proposed concerted evolution (Nenoi et al., 1998; Sharp and Li, 1987; Tan et al., 1993; Thomas et al., 1997; Vrana, et al., 1996), with gene conversion as the form of concerted evolution most often proposed as an influence in gene family evolution. Gene conversion is a non-reciprocal recombination event that allows interlocus genetic exchange without rendering the donor gene deficient (Dover, 1982; Graur and Li, 1999, pg. 308; Ohta, 1983). Gene conversion is facilitated by the nucleotide sequence of the genes involved. Therefore one would expect homogeneity at both the nucleotide and amino acid level if gene conversion had occurred. However, Nei et al. (2000) noted that, in many cases where concerted evolution had previously been concluded, there was considerable divergence at the nucleotide level. Through the exploration of the proportion of synonymous substitutions per synonymous site they concluded that the ubiquitin gene family

had undergone birth-and-death evolution rather than concerted evolution (Nei et al., 2000).

Birth-and-death evolution and concerted evolution may not be mutually exclusive. One would expect recently duplicated loci to have some predisposition to inter-locus recombination and thus gene conversion, especially if duplication occurs in tandem on the same chromosome. Therefore there are two instances where it may be hard to distinguish concerted evolution from birth-and-death evolution. The first instance involves families where gene conversion occurs in the apparent absence of birth-and-death evolution. If gene conversion were to occur at a low rate, synonymous substitutions could take place among related loci thus creating profiles which would appear to mimic birth-and-death evolution (Nei et al., 2000). In addition, the result of rapid birth-and-death evolution is homogeneity at the nucleotide level which would appear to indicate concerted evolution. If birth-and-death evolution and gene conversion were to act on a gene family at the same time and if the rate of each were great enough, they could not be distinguished from each other within a given species. As Nei et al. (2000) observed, inter-specific analysis allows the two to be distinguished based on factors such as chromosomal position, nucleotide composition and the number of gene copies. For this reason the amount of nucleotide sequence available for a given gene family may force one to infer one of these two forms of evolution when in fact the other may be the dominating force or both may be acting together. Therefore, repeated analysis of gene families will be required to determine the actual mode of evolution.

In response to a call for re-evaluation of gene family evolution (Nei et al., 2000), we have begun exploration into the evolution of α-actinins. These genes provide not only the potential for testing the alternate hypotheses of how homogeneity is maintained in gene families, but also may lead to the development of genetic treatments directed toward human diseases such as cancer and nemaline myopathy with which α-actinins have been shown to be associated (Honda et al., 1998; Jockusch et al., 1980; Nikolopoulos, 2000).

Alpha-actinins are a diverse group of cytoskeletal proteins which belong to a larger group of proteins referred to as the spectrin super-family. The super-family is characterized by the ability to bind actin (Baron et al., 1987; Beggs et al., 1992; Dubreuil, 1991) and the presence of a specific number of spectrin repeats. The super-family contains α-actinin(s), α- and β-spectrin(s), $\beta_{heavy}$ spectrin(s) and dystrophin(s). Alpha-actinins contain four spectrin repeats and are therefore by virtue of parsimony considered more primitive than β-spectrin which has 17 repeats, α-spectrin which has 20 repeats, dystrophin which has 24 repeats and $\beta_{heavy}$ spectrin which has 30 repeats (Muse et al., 1997; Thomas et al., 1997). The observed divergence among spectrin repeats is thought to have arisen during two distinct periods of evolution. In the first period, an α-actinin-like ancestor underwent a series of duplications which ultimately gave rise to the families presented above. The array structure of the spectrin repeats allowed duplication events to occur within the repeats. Enduring such events allowed each family to achieve a unique molecular length, selectively restrained by the resulting actin-cross-linking distance (Muse et al., 1997; Thomas et al., 1997).

Alpha-actinin assembles as a rod shaped homodimer with a sub-unit molecular mass of 94-103 kDa (reviewed by Blanchard et al., 1989). However, it has been demonstrated that at least two of the four main isoforms of α-actinin (α-actinins 2 and 3) form heterodimers (Chan et al., 1998). The primary nucleotide and/or amino acid structure of this gene family consists of two calponin-homology (actin-binding) domains at the N-terminus followed by four spectrin repeats and two EF-hand domains at the C-terminus (reviewed by Dubreuil et al., 1991).

The family is functionally divided by the ability to bind calcium. Isoforms which bind calcium undergo a conformational deformation, upon binding, which causes the release of actin from the actin-binding domains. Therefore, sensitivity to calcium yields two distinct groups which generally define muscle (calcium insensitive) and non-muscle (calcium sensitive) isoforms (Burridge et al., 1981; Noegel et al., 1987). With its ability to bind and cross-link actin, α-actinin is located in areas where such functionality is required, such as muscle and non-muscle cells including liver, kidney, small intestine and brain (Beggs et al., 1992). Furthermore, the organismal diversity is currently known to encompass *Dictyostelium discoideum, Trichomonas vaginalis, Caenorhabditis elegans, Drosophila melanogaster, Gallus gallus, Rattus norvegicus, Mus musculus,* and *Homo sapiens* (Addis et al., 1998; Barstead et al., 1991; Beggs et al., 1992; Bricheux et al., 1998; El-Husseini et al., 2000; Endo and Masaki, 1984; Fyrberg et al., 1990; Geisert et al., 1991).

Resolution of the diversity of α-actinin requires unified examination and resolution of the resulting family tree. Likewise the system appears to provide

an avenue for re-evaluation of the evolutionary mode within this minimally-examined gene family. It is the goal of this study to resolve the gene tree for modern α-actinins, to determine whether gene conversion has occurred in the spectrin repeats and to make inferences as to whether α-actinin undergoes birth-and-death evolution.

## Materials and Methods

### *Reverse Transcription*

The sequences for all primers used in reverse transcription, PCR and RACE are listed in Table 1.1. Poly(A)mRNA from rat skeletal muscle (strain=Spague Dawley) was obtained from Sigma (St. Louis). A 2.0 μg portion was reverse transcribed using the primer RVSAA which was designed based on the *Mus musculus* sequence for α-actinin 3 (AF093775). The pre-reaction consisted of 2.5μl 400mM Tris-HCl, 2.5μl 400mM KCl, 1μl 300mM $MgCl_2$, 5μl 100mM dithiothreitol (DTT), 2.5μl 10mM dNTP, 2.0μl actinomycin D (2mg/ml), 0.38μl RNAguard (Amersham, Piscataway, NJ) and 12.83μl cellulose-nitrate-filtered (0.22μm) $dH_20$. The pre-reaction was then added to the reaction mixture. The reaction mixture was prepared as follows: a solution of 1μl RVSAA (10pmol/μl), 3.57μl (2.0μg) poly(A)mRNA and 16.43μl RNase free $dH_20$ was incubated at 65°C for 15 minutes. After combining the reaction mix and the pre-reaction mix, 0.8μl (16u) of AMV Reverse Transcriptase was added. The reaction was then allowed to incubate for one hour at 50°C. Following incubation, the reaction was diluted with 450μl of 10mM Tris-HCl/10mM EDTA (pH 7.5). It was then stored at -30°C.

*cDNA Amplification*

In order to amplify the cDNA corresponding to rat α-actinin 3, the primers FWDAA and RVSAA were used in a PCR reaction which incorporated the product from the above reverse transcription reaction as template. The PCR reaction consisted of 10μl 10X standard PCR buffer, 5μl FWDAA (10 pmol/μl), 5μl RVSAA (10pmol/μl), 5μl diluted cDNA, 2μl dNTP, 62.5 μl $dH_2O$, 10μl 15mM $MgCl_2$ and 0.5μl (2.5 units) *Taq* polymerase. The reaction was then subjected to thermal cycling according to the following cycling parameters: 94°C for 2 minutes; 40 cycles of 94°C for 1 minute, 55°C for 2 minutes and 72°C for 3 minutes. This regimen was followed by a final extension at 72°C for 7 minutes coupled with a 4°C hold. In order to evaluate amplification the samples were subjected to electrophoresis in a 0.8% agarose gel, stained with ethidium bromide and visualized on a UV light table. Once confirmation of amplification was obtained, the entire PCR reaction was loaded into an agarose gel, subjected to electrophoresis and purified by gel extraction using a Nucleospin gel extraction kit (Clontech, Palo Alto, CA). Following gel extraction, 2μl of the extracted product was electrophoresed, as before, to determine the concentration of the extracted product. The plasmid pGEM (ABI, Foster City, CA) was used as the concentration standard for visual concentration evaluation.

*Rapid Amplification of cDNA Ends*

The cDNA template used in the RACE reactions was produced as before with a few alterations. Oligo $d(T)_{20}$ was the reverse transcription primer and a

42°C, 2 hour incubation was used. cDNA ends were amplified using classic RACE as described by Cowell and Austin (1997, pp. 61-75). Gene specific primers FAA6 and FAA7 were used in 3' RACE and RAA6 and RAA7 were used in 5' RACE. Subsequent sequencing was performed with FAA7 and RAA7 to obtain the sequence across the 3' and 5' ends of the cDNA.

**Table 1.1** Primers used for RT, PCR and DNA sequencing.

| Primer Name | Primer Sequence 5' to 3' |
|---|---|
| FWDAA | GGCTCAGGCGCAGCTTGGTCTGC |
| FAA2 | CAAGAGCCTGAGCAAGATGAT |
| FAA3 | GCCCACAGCCTGATCTTTATT |
| FAA4 | AAGCTAGTGTCCATCGGTGC |
| FAA5 | AGGCTACCACTTGTTCAGCC |
| FAA6 | CCTATGTTTCCTGCTTCTACC |
| FAA7 | CCCTGTTGGCAGCAGCAGTC |
| FAA8 | GCGAAGCTTGCATAGTCAATG |
| RVSAA | CCATTCTCTCTAGTGCATCCCTCCTCTTC |
| RAA2 | CAACCCCTACATCACCCTCAGTTCGC |
| RAA3 | CAACCTCCATTGCCCGCACCATC |
| RAA4 | CTACATTACCCCAGAAGAGCTGCGGAG |
| RAA5 | GTCCTCTTCAATGTTCTCGATC |
| RAA6 | CTGGTTTCGGACATAGCCAATGC |
| RAA7 | CTGCTGCCGGGCCAGCTCCTCC |
| RAA8 | CAGCTAAGACCCGGTTGTCCTAGACGTT |
| RAA9 | GCAGCCTCGGTGCTACCGAACCGTCGCCGATAA |

*Automated DNA Sequencing*

Cycle sequencing reactions were performed using 40-100ng of gel-extracted product as template with an Applied Biosystems second generation Big-Dye dye terminator cycle sequencing kit (ABI, Foster City, CA). A typical sequencing reaction consisted of 1μl template, 0.5μl primer, 3μl Big Dye and 4.5μl $dH_2O$. The reactions were carried out for 35-40 cycles of 96°C for 5 seconds, 50°C for 10 seconds and 60°C for 4 minutes. Synthetic oligonucleotides were designed after each sequencing series in order to extend the sequenced fragments across the entire length of the cDNA. The contiguous sequence was

assembled using the computer program Sequencher (v. 4.1)(Gene Codes Corp., Arbor, MI).

*Phylogenetic Analysis*

Gene family analysis presents problems unique among phylogenetic investigations. The problems associated with examination of gene families which experience birth-and-death evolution and/or concerted evolution include underestimation of branch length and topological mis-interpretation within and among groups (Jackson et al., 1996). In order to minimize the contribution of these anomalies to the final conclusion of this analysis we chose to use Maximum Parsimony (MP), Neighbor Joining (NJ) and Quartet Puzzling (QP) analyses. Due to the varied objective function among these algorithms an error incurred from a single method may be identified and rejected based on the data derived from the other two analyses.

The sequence for rat α-actinin 3, as well as, sequences from mammals and/or birds for α-actinin (Table 1.2) were aligned using Clustal X version 1.8 (www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html). In addition, a single non-vertebrate sequence was included in the alignment. This sequence was the α-actinin sequence from *Trichomonas vaginalis*. The alignment was modified to contain ~748 bases from the actin-binding domain of the cDNA sequences as this domain is the most conserved of the α-actinin domains (Beggs et al., 1992; Hammonds, 1987) and thus best suited for phylogenetic analyses. The resulting alignment was then used as input for the Phylip package (Felsenstein, 1993) and PUZZLE (Strimmer and Haeseler, 1996).

Phylogenetic analyses included parsimony, neighbor-joining, and quartet puzzling. In all analyses the sequence corresponding to the *Trichomonas vaginalis* α-actinin was designated as the outgroup. NJ analysis was performed using Kimura-2-parameter corrected distances. A transition/ transversion ratio of 1.0 was used where applicable. Nodal support was assessed by bootstrap analyses of 1000 iterations. QP analysis was performed with 1000 puzzling steps under the HKY model of evolution (Hasegawa et al., 1985).

Time of divergence estimates were made using the rodent and human sequences only. Kimura-2-parameter corrections were made to the distance matrices. The time of divergence between primates and rodents was assumed to be 122 million years ago (mya)(Kitano et al., 1999), and the dating of subsequent divergences was made using this estimate in the algorithm provided by Graur and Li (1999, pp. 271-273)

**Table 1.2** Accession numbers, revised names and citations of sequences used in phylogenetic reconstructions. *"Trich. vag."* =*Trichomonas vaginalis* and "AA"=α-actinin.

| Accession # | Name | Revised Name | Citation |
|---|---|---|---|
| M86406 | Human AA2 | N/A | Beggs et al., 1992 |
| AF248643 | Mouse AA2 | N/A | Submitted 2000, N. Yang and K. North |
| M86407 | Human AA3 | N/A | Beggs et al., 1992 |
| AF450248 | Rat AA3 | N/A | Current study |
| AF093775 | Mouse AA3 | N/A | Submitted 1998, C. Birkenmeier et al. |
| J03486 | Chicken actinin | Chicken AA"1.1" | Baron et al., 1991 |
| M74143 | Chicken AA | Chicken AA"1" | Baron et al., 1991 |
| M95178 | Human non-muscle AA | Human AA"1" | Youssoufian et al., 1990 |
| AF115386 | Rat AA1 | N/A | Submitted 1998, T. Schulz and P. Seeburg |
| D26597 | Chicken AA | Chicken AA"4" | Imamura et al., 1994 |
| AJ289242 | Mouse AA4 | N/A | Dear et al., 2000 |
| RNU19893 | Rat AA | Rat AA"4.1" | Geisert et al., 1996 |
| X13874 | Chicken Skele. Mus. AA | Chicken AA"2" | Arimura et al., 1988 |
| AF190909 | Rat AA4 | N/A | El-Husseini et al., 2000 |
| U48734 | Human non-muscle AA | Human AA"4.1" | Nikolopoulous et al., 2000 |
| D89980 | Human AA4 | N/A | Honda et al., 1998 |
| AF072678 | *Trich. vag.* AA1 | *Trich. Vag.* "AA" | Bricheux et al., 1998 |

*Gene Conversion Detection*

The program GeneConv (Sawyer, 1999) was used to detect probable gene conversion events among spectrin repeats using an alignment of the individual spectrin repeats for each of the four α-actinin isoforms from human and rodent. The goal was to find qualitative evidence for gene conversion; therefore, mismatches and a maximum overlap value of 150 were allowed. The results were examined based on the percentage of total events detected as occurring between a given pair of repeats.

A distance network was constructed from the alignment used in the GeneConv analysis. Time of divergence estimates were made as in the phylogenetic analysis of the actin-binding domain (above). Results were analyzed based on periods of evolution rather than points of divergence.

## Results

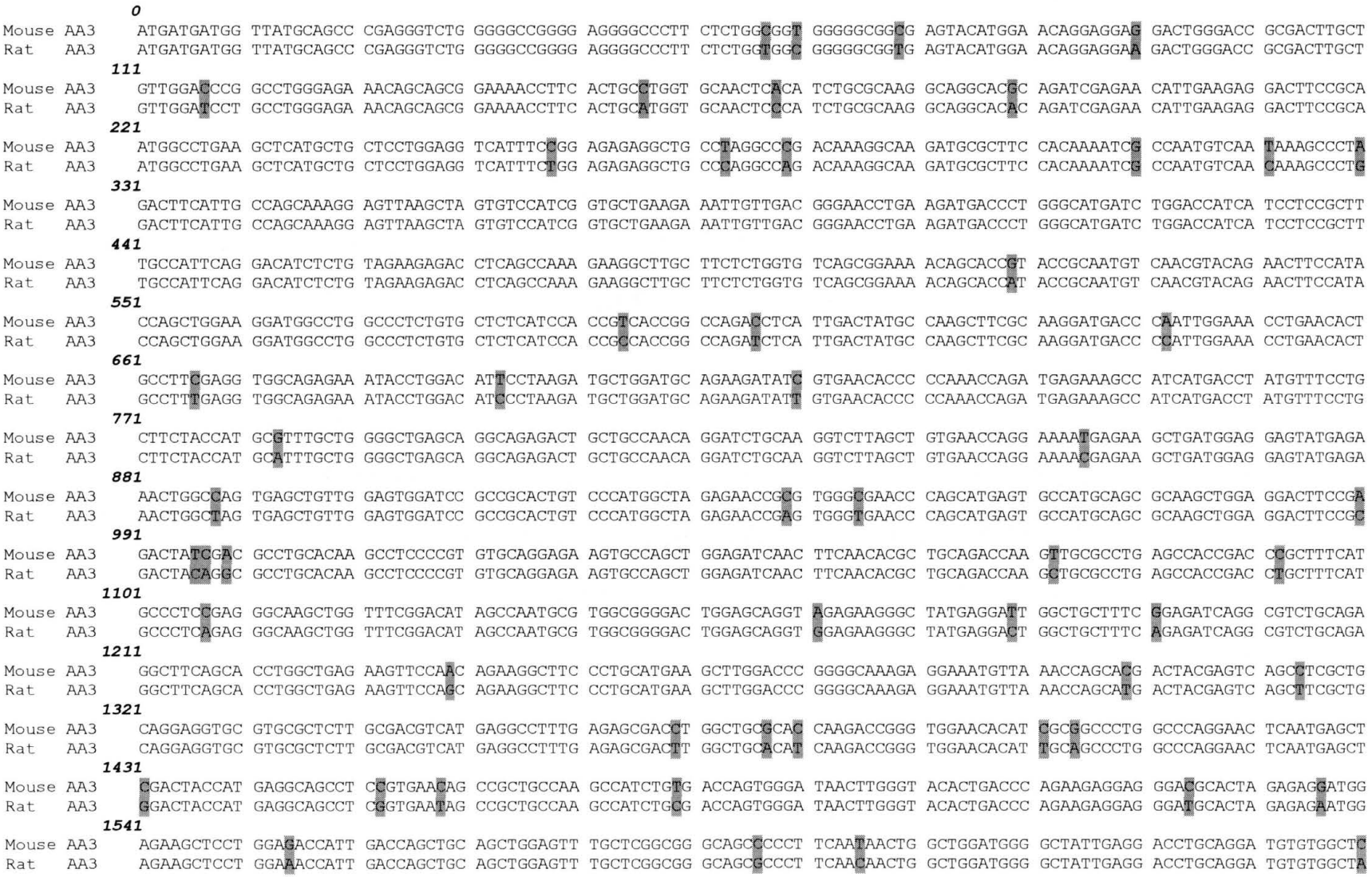
*Sequencing of a cDNA encoding rat α-actinin 3*

A cDNA clone for a rat α-actinin was isolated and sequenced utilizing RT-PCR and automated DNA sequencing. The clone proved to contain 2868 nucleotides of rat origin, 2700 of which are found within a single open reading frame which appears to encode rat α-actinin 3. In addition to the single open reading frame, the clone contained 39 bases within the 5' un-translated region and 129 bases within the 3' un-translated region (AF450248). The nucleotide identity, within the open reading frame, between rat and mouse α-actinin 3 is 95.8%. Using the inferred amino acid sequence the clone was found to be identical to mouse α-actinin 3 at all but 1 of 900 residues, translating to 99.9%

identity. The single bi-morphism at residue 711 is the result of a single non-synonymous substitution at nucleotide 2133. In addition to this single non-synonymous substitution, there are 113 additional synonymous substitutions. Alignments between rat and mouse α-actinin 3 at both the nucleotide and amino acid levels may be found in Figures 1.1 and 1.2, respectively.

*α-Actinin Phylogenetic Analyses*

All phylogenetic reconstructions were performed on an alignment of ~748 bases from within the actin-binding domain of fifteen different α-actinin nucleotide sequences taken from GenBank and the cDNA sequence for rat α-actinin 3 obtained in the current study (Table 2.1). Topological results from MP, NJ and QP analyses may be found in Figure 1.3. All methods of re-construction revealed four major groups of α-actinins. Groups were defined based on the chromosomal location of the paralogues for human α-actinins 1, 2, 3 and 4. A BLAST search of the human genome as well as a literature search revealed that each paralogue occurs at a different locus within the human genome (Beggs et al., 1992; Honda et al., 1998; Nikolopoulos, 2000) and can therefore be used as a guide to proper arrangement and subsequent designation of α-actinins. The current classification of α-actinins places type 1 α-actinins as being orthologous to the human isoform which is found to be encoded on chromosome 14, likewise type 2 α-actinins are found on chromosome 1, while types 3 and 4 α-actinins are found on chromosomes 11 and 19, respectively. This classification scheme is compatible with the groupings which we obtained in all phylogenetic analyses. The accession numbers as well as the original names for the proteins used in this


0
Mouse AA3 ATGATGATGG TTATGCAGCC CGAGGGTCTG GGGGCCGGGG AGGGGCCCTT CTCTGGCGGT GGGGGCGGCG AGTACATGGA ACAGGAGGAG GACTGGGACC GCGACTTGCT
Rat AA3 ATGATGATGG TTATGCAGCC CGAGGGTCTG GGGGCCGGGG AGGGGCCCTT CTCTGGTGGC GGGGGCGGTG AGTACATGGA ACAGGAGGAA GACTGGGACC GCGACTTGCT
111
Mouse AA3 GTTGGACCCG GCCTGGGAGA AACAGCAGCG GAAAACCTTC ACTGCCTGGT GCAACTCACA TCTGCGCAAG GCAGGCACGC AGATCGAGAA CATTGAAGAG GACTTCCGCA
Rat AA3 GTTGGATCCT GCCTGGGAGA AACAGCAGCG GAAAACCTTC ACTGCATGGT GCAACTCCCA TCTGCGCAAG GCAGGCACAC AGATCGAGAA CATTGAAGAG GACTTCCGCA
221
Mouse AA3 ATGGCCTGAA GCTCATGCTG CTCCTGGAGG TCATTTCCGG AGAGAGGCTG CCTAGGCCCG ACAAAGGCAA GATGCGCTTC CACAAAATCG CCAATGTCAA TAAAGCCCTA
Rat AA3 ATGGCCTGAA GCTCATGCTG CTCCTGGAGG TCATTTCTGG AGAGAGGCTG CCCAGGCCAG ACAAAGGCAA GATGCGCTTC CACAAAATCG CCAATGTCAA CAAAGCCCTG
331
Mouse AA3 GACTTCATTG CCAGCAAAGG AGTTAAGCTA GTGTCCATCG GTGCTGAAGA AATTGTTGAC GGGAACCTGA AGATGACCCT GGGCATGATC TGGACCATCA TCCTCCGCTT
Rat AA3 GACTTCATTG CCAGCAAAGG AGTTAAGCTA GTGTCCATCG GTGCTGAAGA AATTGTTGAC GGGAACCTGA AGATGACCCT GGGCATGATC TGGACCATCA TCCTCCGCTT
441
Mouse AA3 TGCCATTCAG GACATCTCTG TAGAAGAGAC CTCAGCCAAA GAAGGCTTGC TTCTCTGGTG TCAGCGGAAA ACAGCACCGT ACCGCAATGT CAACGTACAG AACTTCCATA
Rat AA3 TGCCATTCAG GACATCTCTG TAGAAGAGAC CTCAGCCAAA GAAGGCTTGC TTCTCTGGTG TCAGCGGAAA ACAGCACCAT ACCGCAATGT CAACGTACAG AACTTCCATA
551
Mouse AA3 CCAGCTGGAA GGATGGCCTG GCCCTCTGTG CTCTCATCCA CCGTCACCGG CCAGACCTCA TTGACTATGC CAAGCTTCGC AAGGATGACC CAATTGGAAA CCTGAACACT
Rat AA3 CCAGCTGGAA GGATGGCCTG GCCCTCTGTG CTCTCATCCA CCGCCACCGG CCAGATCTCA TTGACTATGC CAAGCTTCGC AAGGATGACC CCATTGGAAA CCTGAACACT
661
Mouse AA3 GCCTTCGAGG TGGCAGAGAA ATACCTGGAC ATTCCTAAGA TGCTGGATGC AGAAGATATC GTGAACACCC CCAAACCAGA TGAGAAAGCC ATCATGACCT ATGTTTCCTG
Rat AA3 GCCTTTGAGG TGGCAGAGAA ATACCTGGAC ATCCCTAAGA TGCTGGATGC AGAAGATATT GTGAACACCC CCAAACCAGA TGAGAAAGCC ATCATGACCT ATGTTTCCTG
771
Mouse AA3 CTTCTACCAT GCGTTTGCTG GGGCTGAGCA GGCAGAGACT GCTGCCAACA GGATCTGCAA GGTCTTAGCT GTGAACCAGG AAAATGAGAA GCTGATGGAG GAGTATGAGA
Rat AA3 CTTCTACCAT GCATTTGCTG GGGCTGAGCA GGCAGAGACT GCTGCCAACA GGATCTGCAA GGTCTTAGCT GTGAACCAGG AAAACGAGAA GCTGATGGAG GAGTATGAGA
881
Mouse AA3 AACTGGCCAG TGAGCTGTTG GAGTGGATCC GCCGCACTGT CCCATGGCTA GAGAACCGCG TGGGCGAACC CAGCATGAGT GCCATGCAGC GCAAGCTGGA GGACTTCCGA
Rat AA3 AACTGGCTAG TGAGCTGTTG GAGTGGATCC GCCGCACTGT CCCATGGCTA GAGAACCGAG TGGGTGAACC CAGCATGAGT GCCATGCAGC GCAAGCTGGA GGACTTCCGC
991
Mouse AA3 GACTATCGAC GCCTGCACAA GCCTCCCCGT GTGCAGGAGA AGTGCCAGCT GGAGATCAAC TTCAACACGC TGCAGACCAA GTTGCGCCTG AGCCACCGAC CCGCTTTCAT
Rat AA3 GACTACAGGC GCCTGCACAA GCCTCCCCGT GTGCAGGAGA AGTGCCAGCT GGAGATCAAC TTCAACACGC TGCAGACCAA GCTGCGCCTG AGCCACCGAC CTGCTTTCAT
1101
Mouse AA3 GCCCTCCGAG GGCAAGCTGG TTTCGGACAT AGCCAATGCG TGGCGGGGAC TGGAGCAGGT AGAGAAGGGC TATGAGGATT GGCTGCTTTC GGAGATCAGG CGTCTGCAGA
Rat AA3 GCCCTCAGAG GGCAAGCTGG TTTCGGACAT AGCCAATGCG TGGCGGGGAC TGGAGCAGGT GGAGAAGGGC TATGAGGACT GGCTGCTTTC AGAGATCAGG CGTCTGCAGA
1211
Mouse AA3 GGCTTCAGCA CCTGGCTGAG AAGTTCCAAC AGAAGGCTTC CCTGCATGAA GCTTGGACCC GGGGCAAAGA GGAAATGTTA AACCAGCACG ACTACGAGTC AGCCTCGCTG
Rat AA3 GGCTTCAGCA CCTGGCTGAG AAGTTCCAGC AGAAGGCTTC CCTGCATGAA GCTTGGACCC GGGGCAAAGA GGAAATGTTA AACCAGCATG ACTACGAGTC AGCTTCGCTG
1321
Mouse AA3 CAGGAGGTGC GTGCGCTCTT GCGACGTCAT GAGGCCTTTG AGAGCGACCT GGCTGCGCAC CAAGACCGGG TGGAACACAT CGCGGCCCTG GCCCAGGAAC TCAATGAGCT
Rat AA3 CAGGAGGTGC GTGCGCTCTT GCGACGTCAT GAGGCCTTTG AGAGCGACTT GGCTGCACAT CAAGACCGGG TGGAACACAT TGCAGCCCTG GCCCAGGAAC TCAATGAGCT
1431
Mouse AA3 CGACTACCAT GAGGCAGCCT CCGTGAACAG CCGCTGCCAA GCCATCTGTG ACCAGTGGGA TAACTTGGGT ACACTGACCC AGAAGAGGAG GGACGCACTA GAGAGGATGG
Rat AA3 GGACTACCAT GAGGCAGCCT CGGTGAATAG CCGCTGCCAA GCCATCTGCG ACCAGTGGGA TAACTTGGGT ACACTGACCC AGAAGAGGAG GGATGCACTA GAGAGAATGG
1541
Mouse AA3 AGAAGCTCCT GGAGACCATT GACCAGCTGC AGCTGGAGTT TGCTCGGCGG GCAGCCCCCT TCAATAACTG GCTGGATGGG GCTATTGAGG ACCTGCAGGA TGTGTGGCTC
Rat AA3 AGAAGCTCCT GGAAACCATT GACCAGCTGC AGCTGGAGTT TGCTCGGCGG GCAGCGCCCT TCAACAACTG GCTGGATGGG GCTATTGAGG ACCTGCAGGA TGTGTGGCTA

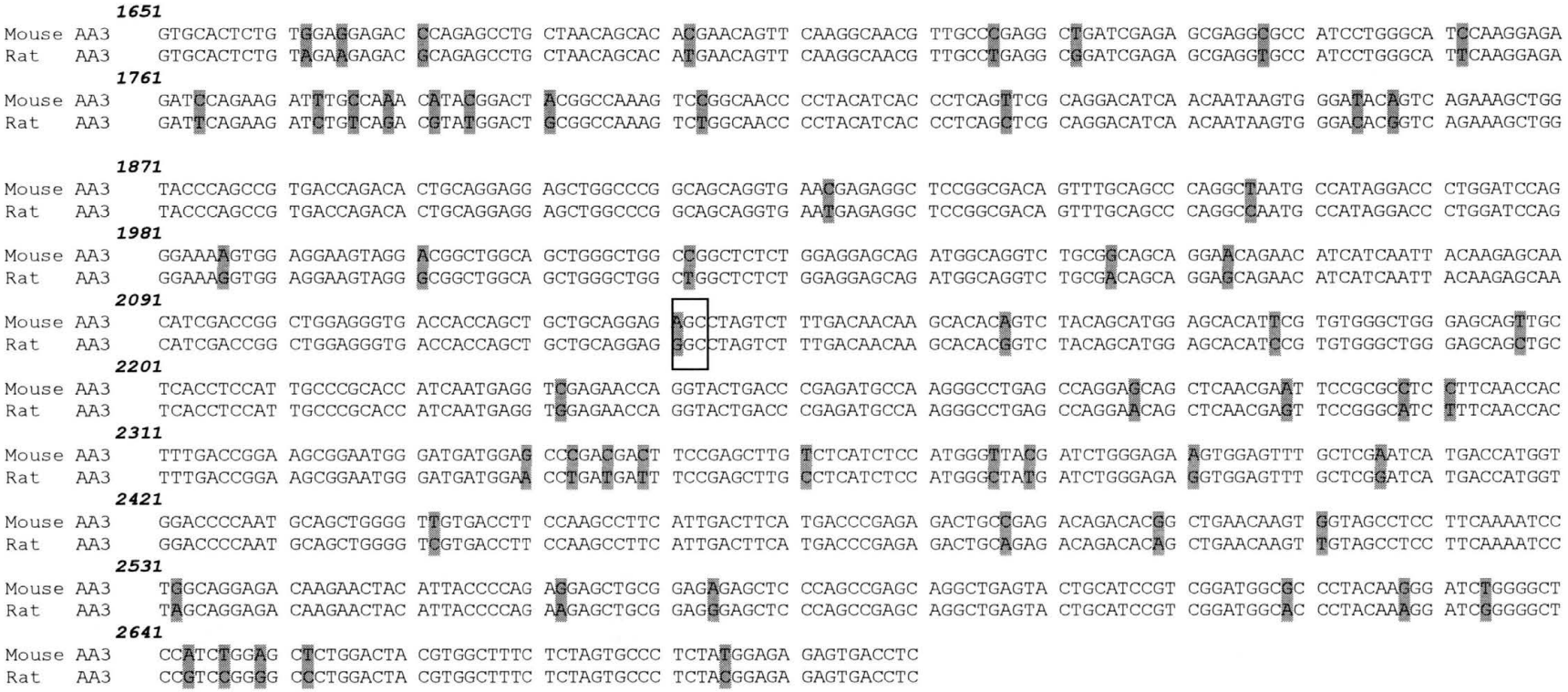



**Figure 1.1.** A pair-wise alignment of the sequence for mouse α-actinin 3 (AF093775) and the cDNA sequence for rat α-actinin 3 (AF450248) which was obtained in this study. The sequences are 95.8% identical at the nucleotide level and 99.9% identical at the amino acid level. Bases which are divergent between the two sequences are shaded. A single non-synonymous amino-acid substitution occurred as a result of the divergence between this pair of sequences. The triplet encoding the substitution is boxed.

```
            1
Mouse AA3    MMMVMQPEGL GAGEGPFSGG GGGEYMEQEE DWDRDLLLDP AWEKQQRKTF TAWCNSHLRK AGTQIENIEE DFRNGLKLML LLEVISGERL PRPDKGKMRF
Rat   AA3    MMMVMQPEGL GAGEGPFSGG GGGEYMEQEE DWDRDLLLDP AWEKQQRKTF TAWCNSHLRK AGTQIENIEE DFRNGLKLML LLEVISGERL PRPDKGKMRF

          101
Mouse AA3    HKIANVNKAL DFIASKGVKL VSIGAEEIVD GNLKMTLGMI WTIILRFAIQ DISVEETSAK EGLLLWCQRK TAPYRNVNVQ NFHTSWKDGL ALCALIHRHR
Rat   AA3    HKIANVNKAL DFIASKGVKL VSIGAEEIVD GNLKMTLGMI WTIILRFAIQ DISVEETSAK EGLLLWCQRK TAPYRNVNVQ NFHTSWKDGL ALCALIHRHR

          201
Mouse AA3    PDLIDYAKLR KDDPIGNLNT AFEVAEKYLD IPKMLDAEDI VNTPKPDEKA IMTYVSCFYH AFAGAEQAET AANRICKVLA VNQENEKLME EYEKLASELL
Rat   AA3    PDLIDYAKLR KDDPIGNLNT AFEVAEKYLD IPKMLDAEDI VNTPKPDEKA IMTYVSCFYH AFAGAEQAET AANRICKVLA VNQENEKLME EYEKLASELL

          301
Mouse AA3    EWIRRTVPWL ENRVGEPSMS AMQRKLEDFR DYRRLHKPPR VQEKCQLEIN FNTLQTKLRL SHRPAFMPSE GKLVSDIANA WRGLEQVEKG YEDWLLSEIR
Rat   AA3    EWIRRTVPWL ENRVGEPSMS AMQRKLEDFR DYRRLHKPPR VQEKCQLEIN FNTLQTKLRL SHRPAFMPSE GKLVSDIANA WRGLEQVEKG YEDWLLSEIR

          401
Mouse AA3    RLQRLQHLAE KFQQKASLHE AWTRGKEEML NQHDYESASL QEVRALLRRH EAFESDLAAH QDRVEHIAAL AQELNELDYH EAASVNSRCQ AICDQWDNLG
Rat   AA3    RLQRLQHLAE KFQQKASLHE AWTRGKEEML NQHDYESASL QEVRALLRRH EAFESDLAAH QDRVEHIAAL AQELNELDYH EAASVNSRCQ AICDQWDNLG

          501
Mouse AA3    TLTQKRRDAL ERMEKLLETI DQLQLEFARR AAPFNNWLDG AIEDLQDVWL VHSVEETQSL LTAHEQFKAT LPEADRERGA ILGIQGEIQK ICQTYGLRPK
Rat   AA3    TLTQKRRDAL ERMEKLLETI DQLQLEFARR AAPFNNWLDG AIEDLQDVWL VHSVEETQSL LTAHEQFKAT LPEADRERGA ILGIQGEIQK ICQTYGLRPK

          601
Mouse AA3    SGNPYITLSS QDINNKWDTV RKLVPSRDQT LQEELARQQV NERLRRQFAA QANAIGPWIQ GKVEEVGRLA AGLAGSLEEQ MAGLRQQEQN IINYKSNIDR
Rat   AA3    SGNPYITLSS QDINNKWDTV RKLVPSRDQT LQEELARQQV NERLRRQFAA QANAIGPWIQ GKVEEVGRLA AGLAGSLEEQ MAGLRQQEQN IINYKSNIDR

          701
Mouse AA3    LEGDHQLLQE SLVFDNKHTV YSMEHIRVGW EQLLTSIART INEVENQVLT RDAKGLSQEQ LNEFRASFNH FDRKRNGMME PDDFRACLIS MGYDLGEVEF
Rat   AA3    LEGDHQLLQE GLVFDNKHTV YSMEHIRVGW EQLLTSIART INEVENQVLT RDAKGLSQEQ LNEFRASFNH FDRKRNGMME PDDFRACLIS MGYDLGEVEF
                        *
          801
Mouse AA3    ARIMTMVDPN AAGVVTFQAF IDFMTRETAE TDTAEQVVAS FKILAGDKNY ITPEELRREL PAEQAEYCIR RMAPYKGSGA PSGALDYVAF SSALYGESDL
Rat   AA3    ARIMTMVDPN AAGVVTFQAF IDFMTRETAE TDTAEQVVAS FKILAGDKNY ITPEELRREL PAEQAEYCIR RMAPYKGSGA PSGALDYVAF SSALYGESDL
```

**Figure 2.2.** An alignment of the inferred amino-acid sequence for mouse α-actinin 3 (AF093775) and the inferred amino-acid sequence for rat α-actinin 3 (AF450248) obtained in this study. The sequences show 99.9% identity and are only divergent at a single position depicted by "*."

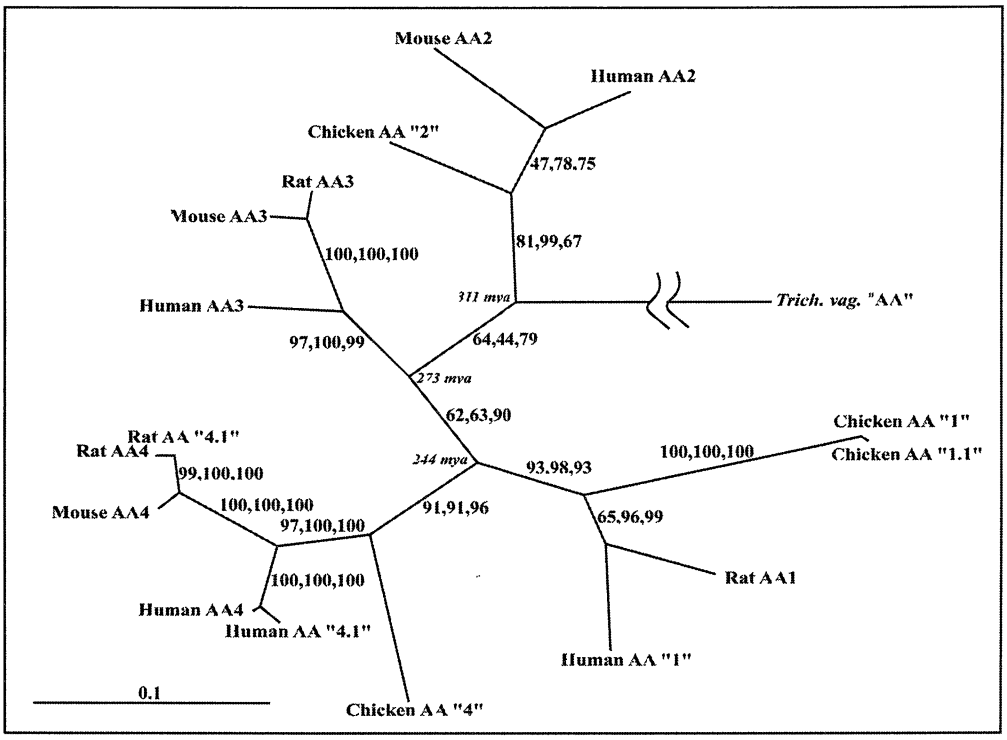



**Figure 1.3.** A composite of the maximum parsimony (MP), neighbor joining (NJ) and quartet puzzling (QP) phylogenetic re-constructions for the α-actinin gene family. The topology was supported by each method and the dimensions are derived from maximum likelihood distances used in the QP re-construction. Time of divergence estimates derived from Kimura-2-parameter corrected distances are represented at the respective nodes. Bootstrap values (%) of 1000 iterations in each method are presented on the branches corresponding to the respective internal nodes. The first value on a given branch represents the MP bootstrap value followed by the NJ bootstrap value and then the percentage of puzzling steps out of 1000 which support the observed branching pattern.

study are listed in Table 1.2. In addition, this table contains the names we propose as replacements while remaining consistent with the currently accepted naming system. In cases where the original name given to a sequence is consistent with the current system of nomenclature, that name was used in our phylogenetic re-constructions (Table 1.2). However, in cases where it was necessary to assign a new name for a given sequence we included the name within quotation marks. This designation is intended to clarify any

discrepancies between our subsequent discussion and the pre-existing classification of these proteins.

The three phylogenetic reconstructions performed in this study each support the same general topology. In all three cases reconstruction yields four major groups. In addition, each of these groups contains one human isoform from one location in the human genome. It appears that type 2 α-actinins are the least derived of the three isoforms, followed by type 3 and then types 1 and 4. Using Kimura-2-parameter distance data along with information as to the time of divergence between rodents and primates, it was determined that type 2 α-actinins are ~311 million years old followed in sequential order by type 3 which is ~273 million years old and types 1 and 4 which are both ~244 million years old. Bootstrap values for the MP and NJ re-constructions as well as the reliability values from 1000 puzzling steps in the QP reconstruction show a high level of support for each internal node in the composite tree (Figure 1.3). The nodes separating types 4, 3 and 1 α-actinins and the node which separates mouse and human α-actinin 2 show no bootstrap support in two analyses (44% (NJ) and 47% (MP) respectively) while all other nodes within the tree are supported (62-100%). Notably each of the unsupported nodes is well supported in each of the other two analyses.

Within all analyses, the mono-phyletic group for type 1 α-actinins contains chicken AA "1," chicken AA "1.1," rat AA1 and human AA "1." The type 2 α-actinins include mouse AA2, chicken AA "2" and human AA2. The type 3 α-actinins contain mouse, rat and human AA3, and type 4 α-actinins

encompass chicken AA "4," rat AA4, rat AA "4.1," mouse AA4, human AA4 and human AA "4.1." A number after the decimal point in these names indicates that the sequence is likely a different form of the original gene sequenced for that species. In each of these bifurcations confidence limits (not shown) for the ML distances (depicted in tree topology in Figure 1.3) are very low indicating the possibility of an allelic bifurcation. This is supported by the

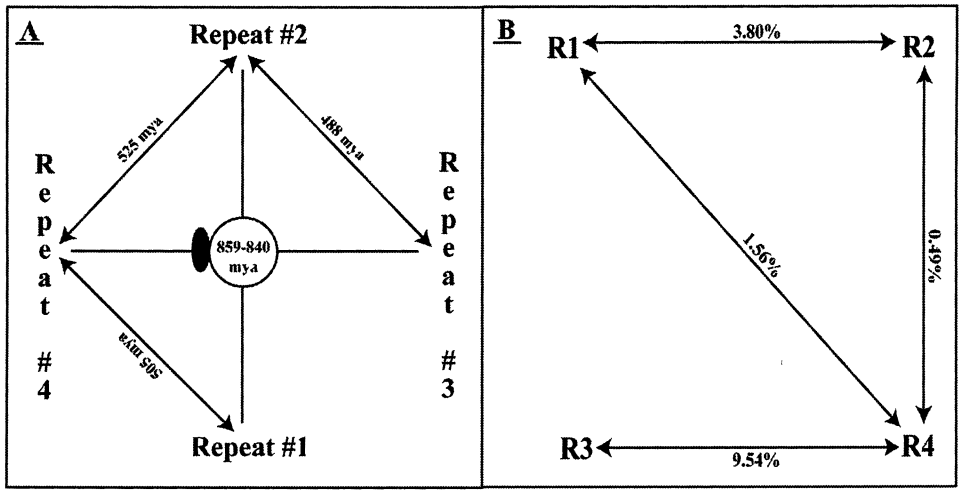



**Figure 1.4. A)** A graph inferred using Kimura-2-parameter correction as the basis for time of divergence estimates for α-actinin repeats. Representative α-actinin sequences from rodents and primates were used. This representation indicates that there have been two major periods of α-actinin repeat similarity. The first period which occurred 859-840mya may have been the original divergence of the repeats through duplication. The second period which ceased ~500mya was most likely a period of homogenizing evolution (either concerted or birth-and-death evolution). The shaded oval indicates that the graph does not present evidence but rather inference that repeat #4 appeared during this time period. **B)** A graph indicating relative amounts of gene conversion among α-actinin spectrin repeats based on quantitative treatment of data derived from analysis of α-actinin spectrin repeats using the program GeneConv (Sawyer, 1999).

fact that there is evidence of allelic forms of α-actinin 4 in the literature (Honda et al., 1998, Nikolopoulos et al., 2000). Finally, a BLAST search of the human genome places both human type 4 sequences at the same locus.

*Detection of Concerted Evolution*

In order to evaluate whether or not gene conversion was detectable among the spectrin repeats of primates and rodents, each repeat from each isoform of α-actinin from both groups were aligned. From the resulting alignment, gene conversion was detected using GeneConv (Sawyer, 1999). The program identified 1027 probable gene conversion events. The events were identified as being significant at the 95% confidence level using the default permutated p-value calculation of the program. Of the events identified, 9.54% were between repeats 1 and 2, 3.80% were between repeats 3 and 4, 1.56% were between repeats 1 and 4, and 0.49% were between repeats 2 and 4. The remaining 84.62% of the events were intra-repeat events. These data are summarized in Figure 1.4B. A similar analysis of the entire sequence for α-actinin sequences identified no gene conversion events in the actin-binding domain, a large proportion in the spectrin repeats and few events within the calcium-binding domain (data not shown).

Using Kimura-2-parameter corrected distances for the spectrin repeats from rodent and human type 1-4 isoforms of α-actinin along with a time of divergence estimate of 122mya from rodents to primates (Kitano et al., 1999) divergence times were estimated between each pair of spectrin repeats. Some divergence times, indicated by arrows in Figure 1.4A, appear to indicate the last time gene conversion occurred in the repeat pair. In this explosion type representation of the divergence among individual spectrin repeats of alpha-actinin, one can see that there are two periods of time during which divergence among spectrin repeats occurred. The first period marked by repeat duplication

took place 859-840 million years ago. The second period of evolution, marked by gene conversion or possibly birth-and-death evolution, occurred from the time of duplication until ~500mya. Assuming these two periods of evolution, a divergence time analysis of the repeats reveals that repeats 1 and 2 evolved in concert using repeat 4 as a shuttle. Likewise, repeats 4 and 3 evolved in concert using repeat 2 as a shuttle. However, the repeat pairs 1 and 4, 4 and 2 as well as 2 and 3 evolved in direct concert.

## Discussion

The MP, NJ and QP phylogenetic re-constructions of the α-actinin gene family are consistent and well supported. Each of the phylogenetic re-constructions support four evolutionary types of α-actinin. Our results substantiate group membership for all previously designated α-actinins (9 out of 17) and successfully assign membership to the previously unassigned α-actinins (8 out of 17)(Table 1.2). In some cases a node was not supported by one of the reconstruction methods used. However, lack of nodal support for those nodes was disregarded based on the high levels of support provided by the other two methods. Therefore I conclude that the tree presented in Figure 1.3 represents the true topology for this gene family.

Concerted evolution and birth-and-death evolution are two modes of evolution which, for the sake of this discussion will, be referred to as homogenizing evolution. Both appear to play a role in the evolution of gene families which exhibit a degree of similarity not expected under a strictly divergent model of evolution (Ota and Nei, 1994). It has been postulated that

most, if not all, gene families undergo some degree of homogenizing evolution (Gojobori and Nei, 1984). In order to determine the particular modes of evolution which act in gene families, an in-depth study of individual gene families is necessary. In an attempt to determine the role of gene conversion as a form of homogenizing evolution acting on the spectrin repeats of the α-actinin gene family we used the program GeneConv (Sawyer, 1999). The program identified 1027 areas where gene conversion may be expected to have occurred. The graphical representation of the data derived from this analysis may be found in Figure 1.4. From these data one might infer that gene conversion between repeats 1 and 3 was a rare event in the evolutionary history of this gene family. Likewise events between repeats 2 and 4 were also rare. In addition, events between repeats 1 and 4 were not as frequent as events between repeats 3 and 4 and 1 and 2. I looked at the possible diverging orders assuming single-repeat duplication events and found a model which fits these data (Figure 1.5B). According to the model the third spectrin repeat is the most primitive. It was duplicated giving rise to repeat 4, which was then duplicated to give rise to repeat 1. Repeat 2 is the most modern of the repeats according to this model as it was derived from repeat 1. This orientation may be summarized as ${}_{4}1,{}_{1}2,3,{}_{3}4.$ or graphically as in Figure 1.5B.

Due to relative proximity, the unequal crossing over event which gave rise to repeat 1 from repeat 4 would have had a relatively low probability of occurrence. Perhaps some functionality was imposed by repeat 4 thus leading to the ${}_{4}1,3,{}_{3}4$ rather than the ${}_{3}1,3,{}_{3}4$ origination which represents another feasible model. This point may be supported by the fact that Beggs et al. (1992) found

that the most variable region among all available α-actinins at the time of their study was from amino acids 570-700. This region composes a large portion of the fourth spectrin repeat. The fact that this region is divergent seems to weaken my argument until one considers the fact that the region is highly conserved

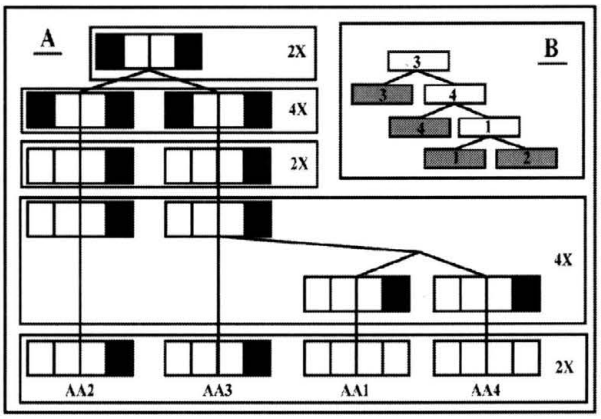



**Figure 1.5 A)** A diagram representing the dosage response of α-actinin spectrin repeats. Under this model an essential function was imparted by repeats 1 and 4 in the ancestral gene. The selective pressure was to maintain a maximum of two repeats with the unknown function. Therefore, excess function depicted by "4X" was eliminated from the system each time that it arose. **B)** A diagram representing the divergence order of the spectrin repeats found in α-actinin. The diagram is read much like a phylogenetic reconstruction. The box with the number "3" within it, at the top of the diagram indicates that repeat #3 is the most ancestral of the repeats.

between human α-actinins 2 and 3 and conserved between human α-actinin 2 and chicken α-actinin "2" (Beggs et al., 1992).

Further support for an essential function of repeat 4 may be concluded from our phylogenetic results (Figure 1.3) which indicate that α-actinins 2 and 3 are the least derived of the known α-actinins. Therefore it can be inferred that the functionality imparted by this region may have been lost not only in repeat 1, but also in repeat 4 from α-actinins 1 and 4 (Figure 1.5A). It can also be proposed

that if this were the case, repeat 1 may have lost its function as the protein structure, under selective pressure, became more efficient at performing the hypothetical task. If the function were catalytic in nature, then it is possible that a dosage relationship may have existed. According to the dosage dependent model (Figure 3A) it seems likely that the loss of function in repeat 1 occurred after the divergence of α-actinins 2 and 3 yet before the divergence of α-actinins 1 and 4 from α-actinin 3. This may also help explain the divergence between repeats 1 and 4. Perhaps repeat 1 remained functional for a period of time followed by duplication of the gene.

Having a duplicate gene may have created an excess of function with regard to the catalytic activity of the region in question. This would have relieved the selective pressure on maintenance of the catalytic activity at two sites within a single molecule as the maintenance of one active site in two molecules could impart the same benefit as maintenance of two active sites in one molecule. This is true only if one of two proteins are expressed in both the same locale and with similar regulation, and it is known that present day α-actinin 2 is expressed in this fashion with regard to α-actinin 3 (Beggs et al., 1992). In fact it has been estimated that ~16% of the world population has a congenital deficiency of α-actinin 3 as a result of a single deleterious mutation. However, this deficiency imparts no apparent disadvantage (North et al., 1999). Therefore it can be inferred that α-actinin 2 may in fact share functionality with α-actinin 3 which provides compensation for the loss of α-actinin 3. The event which gave rise to the $_4$1, $_1$2,3, $_3$4 orientation would have also had a relatively low probability. However, with each successive duplication event the probabilities of

both in-register and out-of-register crossing-over would have increased. Therefore the duplication event which gave rise to repeat 2 would not have required as much selective pressure, to be maintained, as the event which gave rise to repeat 1.

In an alternate scenario repeats 3 and 4 would have arisen early in the history of the gene family after which they were duplicated as a cassette, giving rise to repeats 1 and 2, respectively. This model has the advantage over the previous scenario of explaining the high rate of gene conversion between repeats 3 and 4 (Figure 1.4B). One would expect the most primitive of the repeats to have undergone the highest rate of gene conversion if time is the determining factor. In this case repeats 3 and 4 would have had the most time for gene conversion and would therefore exhibit the most detectable gene conversion events. However cassette style duplication events may not fit the data, based on the apparent difference between the number of events between repeats 3 and 4 versus the number of events between repeats 1 and 2. If repeats 1 and 2 had been the result of a cassette duplication of repeats 3 and 4, then repeats 1 and 2 would have been at the same point in their concerted convergence as repeats 3 and 4 at the time of duplication. This would result in an equal number of detectable conversion events between 1 and 2 as there are between repeats 3 and 4. In addition, events between supposed homologous repeats would be prominent. However, limited gene conversion was detected between repeats 2 and 4 and none between repeats 1 and 3. Therefore we find minimal evidence for cassette style duplication in these data.

In addition to direct gene conversion analysis, we have identified evidence for homogenizing evolution in a time-of-divergence analysis of the repeats. Time-of-divergence analysis yielded two clusters of divergence times. One cluster occurred 859-840mya. The second occurred ~500mya. Thomas et al. (1997) concluded that concerted evolution in spectrins ceased ~500mya. Therefore, the second cluster of divergence times in α-actinins may not indicate divergence through duplication, but rather divergence through the cessation of homogenizing evolution. At first glance one might conclude that Figure 1.4A indicates a diverging order. However distance data may not reveal the true relatedness of repeated sequences if homogenizing evolution has occurred. Therefore the value of Figure 1.4A is in the evolutionary periods which it reveals. The estimate of 859-840mya since the divergence by duplication may be low considering the fact that subsequent gene conversion would have obscured some of the divergence between repeats. However, repeats 1 and 3, which exhibited no evidence for gene conversion in Figure 1.4B, yield the time of divergence estimate of ~859 mya. In the case of repeats 3 and 4, no evidence was obtained to place them into the cluster indicating divergence through duplication; instead the divergence between these two sequences indicates a time of divergence in the 500mya range. This may indicate that as a result of primitive gene conversion, repeats 3 and 4 have experienced more extensive gene conversion than occurred between repeats 1 and 3 and repeats 1 and 2. The second period of evolution indicated by the peripheral arrows and dates in Figure 1.4A is indicative of a period when homogenizing evolution ceased to operate within the spectrin repeats. This period of time is in direct agreement with the estimate made by

Thomas et al. (1997) concerning the spectrin super-family. However in contrast to their study, this study examined only α-actinin sequences.

Figure 1.6 represents an inferred timeline of the evolutionary history for α-actinins. In this timeline one can observe that the divergence of the spectrin repeats occurred ~859-840mya followed by a period of homogenizing evolution which ceased ~500mya. From that time until ~311mya a single α-actinin may have persisted or the family may have maintained only one α-actinin of a host of α-actinins predicted under a birth-and-death model of evolution. Insight into this period will be gained through the exploration of birth-and-death evolution in this gene family (Chapter 2). The single gene was duplicated ~311mya giving rise to α-actinin 2. Approximately 273mya α-actinin 3 appeared followed by α-actinins 1 and 4 which came into existence ~244mya. From these data we predict that organisms which diverged from mammals and birds more than ~311mya will have only one α-actinin in common with the α-actinins presented in this study. However, this does not mean that organisms which diverged before this date will exhibit only one α-actinin. It is always possible that modes of evolution such as converging evolution and/or birth-and-death evolution may have been experienced in this gene family. Therefore as previously indicated, I do not intend for this study to represent the definitive history of α-actinins. Instead it is intended as one avenue for subsequent studies to evaluate predictions against data as yet unavailable.

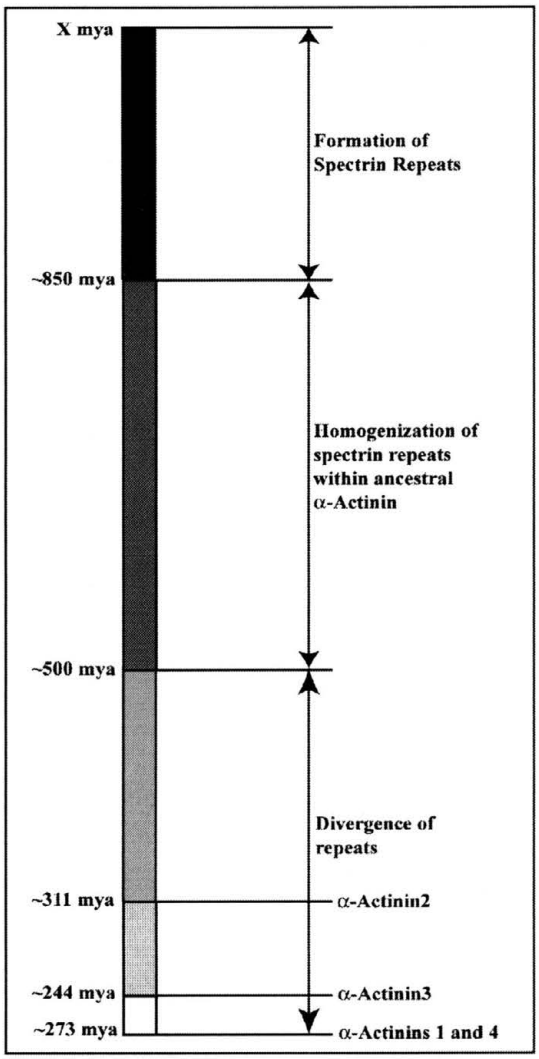



**Figure 1.6** A timeline representation of the time of divergence data for the α-actinin gene family. The "mya" designations indicate points of shift in the evolution of the gene family. Periods of evolution are indicated by labels between dividers while duplication events are indicated by labels on the dividers. "X" mya indicates an indefinite point before formation of the first spectrin repeat.

*Chapter 2*

# A NEW MODEL FOR HOMOGENIZING EVOLUTION


## Summary

**Homogenizing evolution is a process by which repetitive units of DNA maintain more similarity than expected under a strictly divergent mode of evolution. The two known modes of homogenizing evolution with regard to gene families are concerted evolution and birth-and-death evolution. Birth-and-death evolution works with purifying selection to maintain protein level homogeneity while allowing divergence at the nucleotide level. Concerted evolution occurs primarily through continuous non-reciprocal inter- and intra-locus genetic exchange or gene conversion. This repeated exchange maintains homogeneity of a gene at the DNA level and as a consequence at the protein level. These two modes of evolution have been treated as distinct process. However, I make the conceptual leap to a single model under which both processes act concurrently and propose a mathematical equation to estimate the relative contribution of each in the evolution of gene families. This model will allow for the general characterization of gene families with regard to homogenizing evolution.**

## Introduction

Birth-and-death evolution and concerted evolution are two modes of evolution which act to homogenize repeating units of multi-gene families (Klein et al., 1993; Nei et al., 1997; Nei et al., 2000; Sharp et al., 1987; Tan et al, 1993; Wittzell et al., 1999; Zimmer, 1980). Collectively these two modes of evolution may

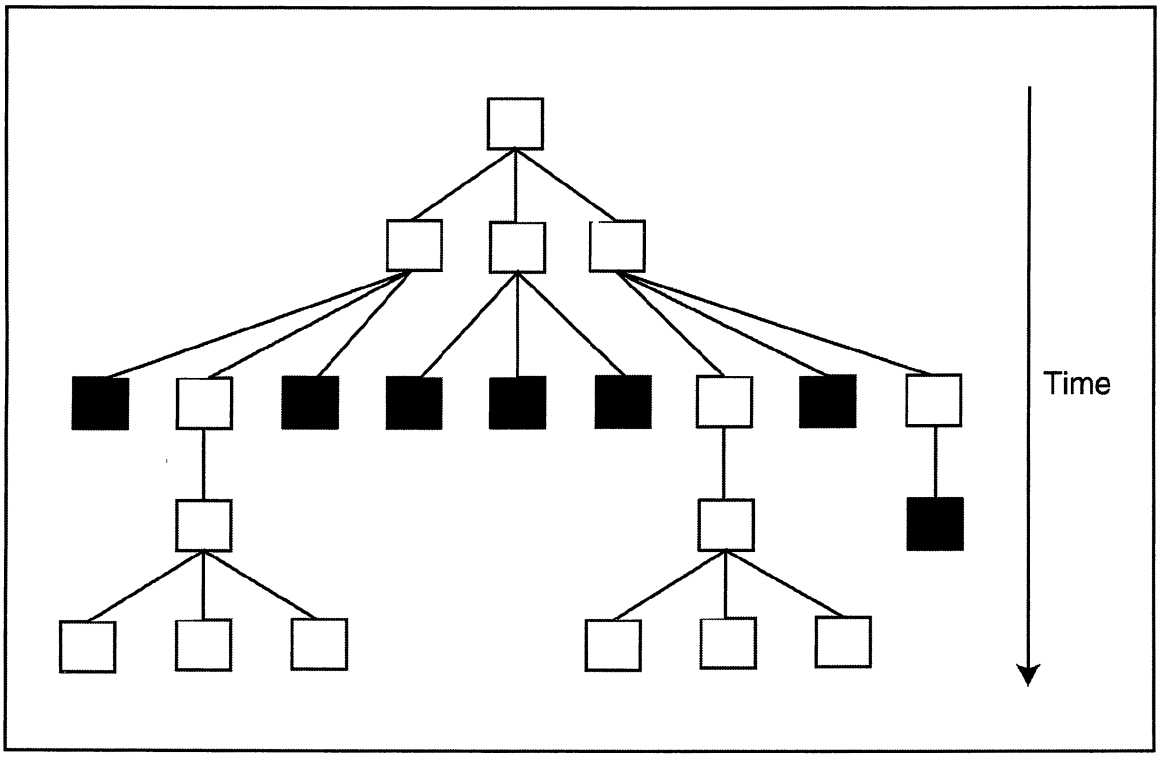



**Figure 2.1.** The mechanism for birth-and-death evolution as proposed by Nei et al. (2000 and 1997). Shaded boxes represent pseudogenes and open boxes represent functional genes. Genes are duplicated and other genes are converted to pseudogenes. Therefore rapid birth-and-death evolution may create homogeneity in a given gene family.

be referred to as homogenizing evolution. Birth-and-death evolution confers homogeneity to gene family members through gene duplication coupled with conversion of some genes to pseudogenes (Nei et al., 2000)(Figure 2.1). Concerted evolution on the other hand primarily conveys homogeneity through

repeated non-reciprocal inter-locus and intra-locus recombination or gene conversion (For the purpose of this chapter, gene conversion and concerted evolution will be used synonymously.) (Graur and Li, 1999; Ohta, 1991; Högstrand and Böhme, 1997). Though a single gene conversion event is non-reciprocal, the flow of information as an overall process is bi-directional since conversion events among members of a gene family are likely to occur more than once when large amounts of time are considered. If combined, birth-and-death evolution and concerted evolution can create homogeneity of gene families within a species and diversity among species as illustrated in Figure 2.2. The fundamental difference in the detection of these modes of evolution is in the level of sequence homogeneity each produces. Concerted evolution homogenizes gene family members at the DNA level and as a consequence maintains amino acid level homogeneity. The primary requirement for gene conversion (concerted evolution) is nucleotide homogeneity (Graur and Li, 1999; Walsh, 1987). Therefore concerted evolution positively reinforces its own action. If the rate of birth-and-death evolution is great enough, it may also convey DNA level homogeneity in a single gene (unit) lineage (Figure 2.1). However it has been shown that the primary force acting to maintain homogeneity in gene families which undergo birth-and-death evolution is purifying selection (Nei et al., 2000), a mode of evolution which results in removal of all alleles from the population except the most beneficial allele with respect to fitness (Graur and Li, 1999, pg. 41). Therefore gene families which have undergone strong birth-and-death evolution also will have undergone strong purifying selection. This derivation comes from the realization that if purifying selection did not

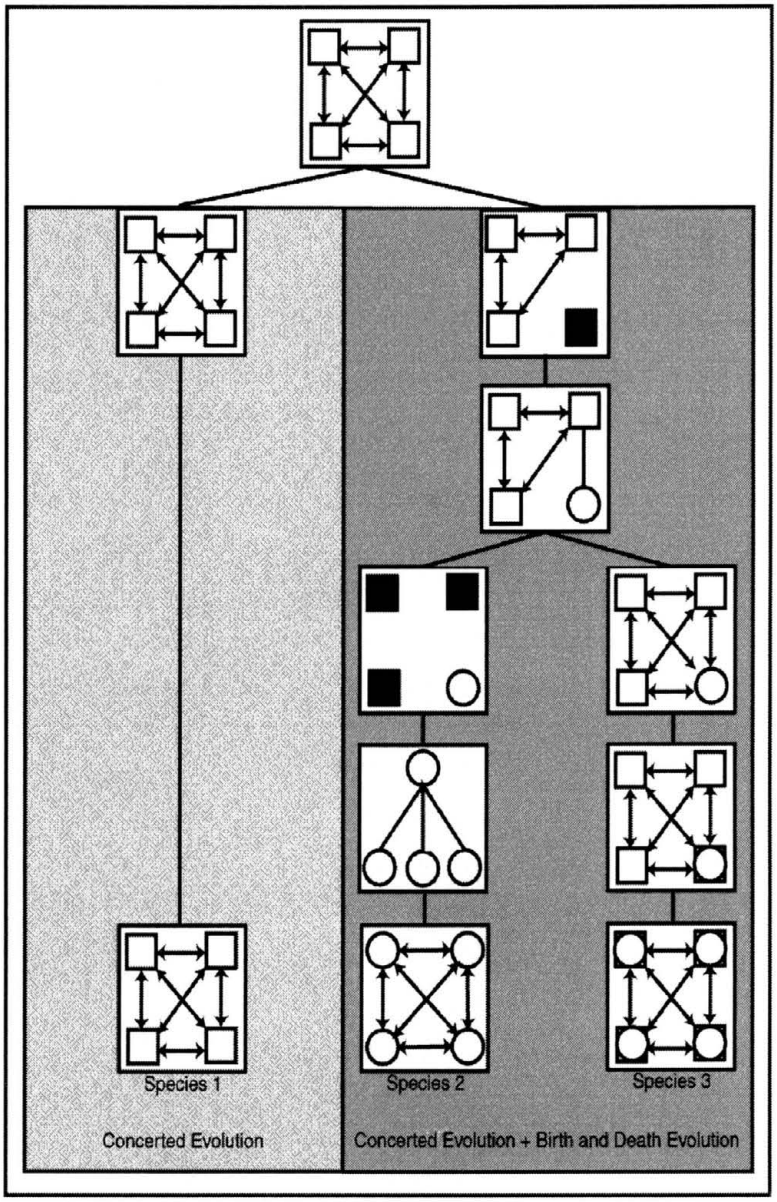



**Figure 2.2.** An illustration demonstrating how concerted evolution and birth-and-death evolution may work together to diversify a given gene family while in other cases a single lineage may remain unchanged with respect to the original product. In the lineage leading to species 1 only concerted evolution has occurred; therefore, the gene family has remained identical to the ancestral gene family. In the lineage leading to species 2 the family underwent birth-and-death evolution which removed three of the original genes from the family. Following the period of birth-and-death evolution a single surviving gene diverged and was then duplicated giving rise to a gene family which is divergent from the original gene family. In the lineage leading to species 3 the gene family underwent a very brief period of birth-and-death evolution which allowed one gene to diverge from the rest followed by dispersion of the divergence to the other gene family members.

positively correlate with strong birth-and-death evolution, gene family members experiencing only this mode of evolution would be free to diverge following duplication events. Given a sufficient amount of time under the birth-and-death model, such divergence would lead to dissolution of the gene family rather than its maintenance.

Here I present the development of a continuous model of birth-and-death evolution and concerted evolution. This model is based on the assumption that these modes of evolution are actually components of the same process (homogenizing evolution). The goal of this study was to create a model for homogenizing evolution and establish its reliability and utility with regard to gene family analysis. In addition I present a brief re-evaluation of the active mode(s) of homogenizing evolution in α-actinins, the HLA-A, B, and C antigens of the MHC-complex and the ubiquitins. Therefore, with the exception of the spectrin repeats of α-actinin, the number of sequences used in each evaluation was fewer than 10. However, due to the treatment of the sequences with regard to analysis, comparisons ranged from 3-495.

## Materials and Methods

### *Conceptualization of a Continuum*

Considering the consequences of concerted evolution and birth-and-death evolution, a continuum of the two processes as presented in Figure 2.3A was compiled using the ratio of the nucleotide saturation value to the amino acid saturation value. Each saturation value was defined as the number of divergent characters divided by the number of characters considered. If strong concerted evolution acts on a gene family, then a 1:1 relationship may exist between the

number of nucleotide substitutions and the number of amino acid substitutions (Nei et al., 2000). Therefore the 1:1 relationship was used as the concerted evolution extreme as presented in the proposed continuum (Figure 2.3A).

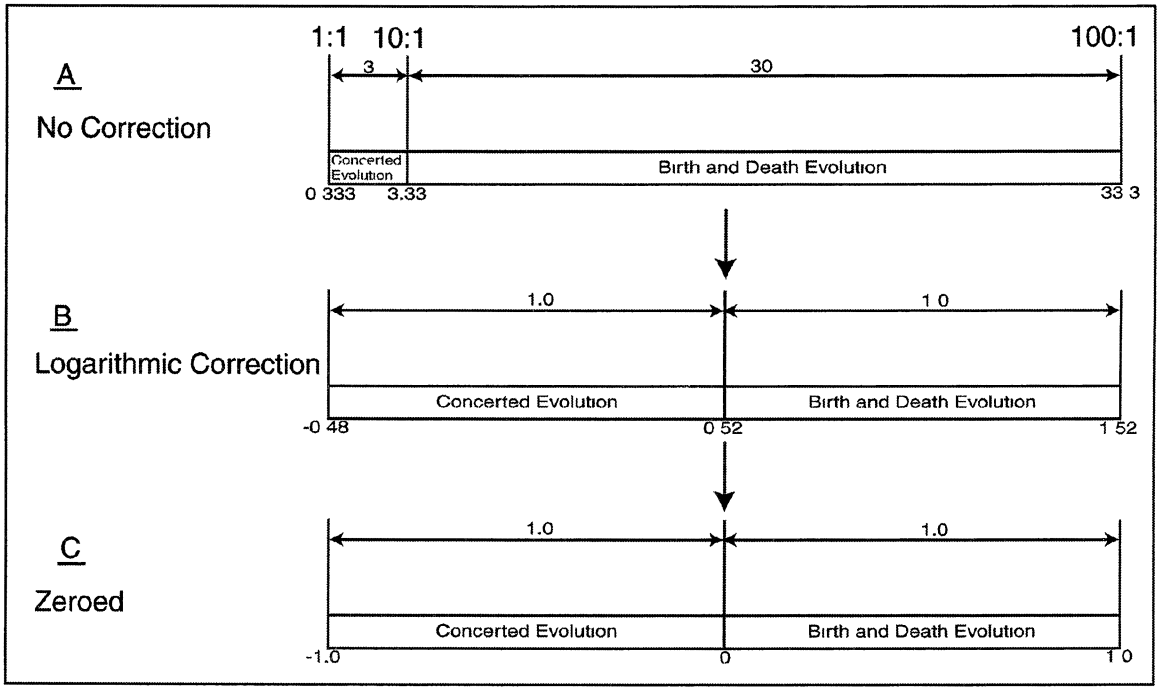



**Figure 2.3.** The proposed continuum for homogenizing evolution **A)** with no correction, **B)** with correction for logarithmic weight and **C)** with correction for scale.

The saturation level for birth-and-death evolution was defined as complete nucleotide divergence with no amino acid divergence. However considering the genetic code, this extreme case is not plausible. Therefore a 100:1 relationship between nucleotide substitution and amino acid substitution was used. This extreme was determined based on the relationship in certain comparisons of the ubiquitin gene family (see results). Use of a 1:1 ratio and a 100:1 ratio as the two extremes allowed for the implementation of a boundary between the two extremes. This boundary was defined as the 10:1 relationship due to its centrality with regard to the 100:1 and 1:1 ratios and consideration of

the 3:1 relationship of nucleotides/ codon in biological systems. These ratios when used to define a continuum based on the relationship of nucleotide level saturation and amino acid level saturation without logarithmic correction gives excess emphasis to birth-and-death evolution (Figure 2.3A). Therefore the logarithm of that relationship normalizes the continuum with respect to the weight given to each process (Figure 2.3B). The continuum was then scaled by subtracting the mid-point value from each value (Figure 2.3C). The resulting midpoint value is then 0 and the extremes values are transformed to –1.0 and 1.0. Consequently, the final continuum is as it appears in Figure 2.3C, with concerted evolution being implied by a negative value for the statistic and birth-and-death evolution being indicated by a positive value.

Gene families were placed into the continuum by employing the equation described above (Figure 2.4). To avoid undefined values for the index, when the amino acid divergence equaled zero, a value of 1 was used as an approximation. This correction factor assumes that a family which exhibits zero amino acid substitutions and many nucleotide substitutions is experiencing high levels of purifying selection. A high rate of purifying selection indicates that any amino acid substitution which is maintained must be neutral with respect to protein function. Therefore a value of zero in the equation carries the same effective result as substituting a value of 1 for this extreme. Use of a value of less than 1 assigns excess emphasis toward birth-and-death evolution. Considering the information implied by this continuum, the statistic used to describe the placement of gene families in the continuum was named the homogenization index (HI)(Figure 2.4).

$$HI=\left(Log\left(\frac{(N/Tn)}{(A/Ta)}\right)\right)-0.52$$

**Figure 2.4.** The formula used to find the homogenization index (HI) value for a given pair-wise comparison. "N" is the number of nucleotide differences, "Tn" is the total number of nucleotides considered, "A" is the number of amino acid differences and "Ta" is the total number of amino acids considered. If a value of zero for "A" and/or "N" is encountered it should be treated as 1 (see text).

*Modeling of the Continuum*

Intra- and inter-specific gene family analyses were performed on α-actinins, ubiquitins and MHC-complex genes. Homogenization indices as defined in Figure 2.4 were calculated for each group. All sequences used in this study may be found in Table 2.1. Groupings of sequences were defined to facilitate inter- and intra-specific evaluation of the sensitivity of the continuum. Ubiquitin, MHC-complex and α-actinin data were obtained from the GenBank database (www.ncbi.nlm.nih.gov). Alignment was performed using Clustal X version 1.8 (www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html), and distance matrices were constructed using total character differences in the phylogenetic analysis package, PAUP (v. 4.0B5)(Swofford, 1999). The central repeats of the α-actinin genes were analyzed as if each repeat were a separate gene. Ubiquitin data was divided into 3 groups; the first group contained separated repeats from the B, C and D domains of a single polyubiquitin locus in *Zea mays* (corn). The second group contained corresponding data with chicken and corn while the third group contained contiguous polyubiquitin loci from human, rat and mouse. The MHC-complex HLA-A, HLA-B and HLA-C antigens were examined

as distinct genes using loci from human and chimpanzee. Inter-specific chimpanzee to human, intra-specific chimpanzee and intra-specific human analyses were performed.

**Table 2.1.** The names, accession numbers and citations for all sequences evaluated using the HI value described in Figure 2.4.

| Accession # | Name | Citation |
|---|---|---|
| M86406 | Human α-actinin 2 | Beggs et al., 1992 |
| M86407 | Human α-actinin 3 | Beggs et al., 1992 |
| AF450248 | Rat α-actinin 3 | Current study |
| AF190909 | Rat α-actinin 4 | El Husseini et al., 2000 |
| D89980 | Human α-actinin 4 | Honda et al., 1998 |
| AF115386 | Rat α-actinin 1 | Submitted 1998, T. Schulz and P. Seeburg |
| M95178 | Human α-actinin "1" | Youssoufian et al., 1990 |
| AF248643 | Mouse α-actinin 2 | Submitted 2000, N. Yang and K. North |
| M26880 | Human ubiquitin | Wiborg et al., 1985 |
| D16554 | Rat polyubiquitin | Hyashi et al., 1994 |
| X51703 | Mouse ubiquitin | Finch et al., 1990 |
| U29162 | Corn ubiquitin | Liu et al., 1995 |
| U29158 | Corn polyubiquitin | Liu et al., 1995 |
| U29159 | Corn polyubiquitin | Liu et al., 1995 |
| M11100 | Chicken ubiquitin | Bond and Schlesinger, 1985 |
| AF115460 | Chimpanzee HLA-B | Matsui et al., 1999 |
| D83031 | Human HLA-C | Submitted 1996, H. Wang |
| AF115464 | Chimpanzee HLA-C | Matsui et al., 1999 |
| L77702 | Human HLA-A | Lienert et al, 1996 |
| AF115463 | HLA-A | Matsui et al., 1999 |

After construction of the distance matrices, the data was compiled into a Microsoft Excel (1997) worksheet (Redmond, WA) and the HI (Figure 2.4) was calculated for each group. The average HI value and the standard deviation for each group was recorded. Extensive statistical treatment of these analyses was omitted from the analysis because the homogenization index is intended as a metric.

## Results

A summary of the data may be found in Table 2.2. The raw data used for each group may be found in the Appendix. The human and rodent comparisons for the spectrin repeats of α-actinin yielded a HI value of –0.60 (Figure 2.5). The corn polyubiquitin data when considered alone yielded a HI value of 0.45, and

**Table 2.2.** HI and standard deviation values for groups studied.

| Unit | N | Group | HI | Stand. Dev. |
|---|---|---|---|---|
| Ubiquitin | 104 | Corn | 0.45 | 0.10 |
| Ubiquitin | 9 | Corn/Chicken | 0.69 | 0.15 |
| $Ubiquitin_1$ | 6 | Rat/Human/Mouse | 1.0 | 0.07 |
| Spectrin Repeats | 495 | Human/Rat/Mouse | -0.60 | 0.18 |
| HLA-A, B, C | 14 | Human/Chimpanzee | -0.71 | 0.04 |
| HLA-A, B, C | 3 | Human | -0.71 | 0.05 |
| HLA-A, B, C | 3 | Chimpanzee | -0.69 | 0.01 |

the same comparison when considered between corn and chicken yielded a HI value of 0.69 (Figure 2.5). The contiguous polyubiquitin sequences for the human, rat and mouse comparisons yielded a homogenizing index value of 1.0, placing them at the end of the continuum indicative of pure birth-and-death

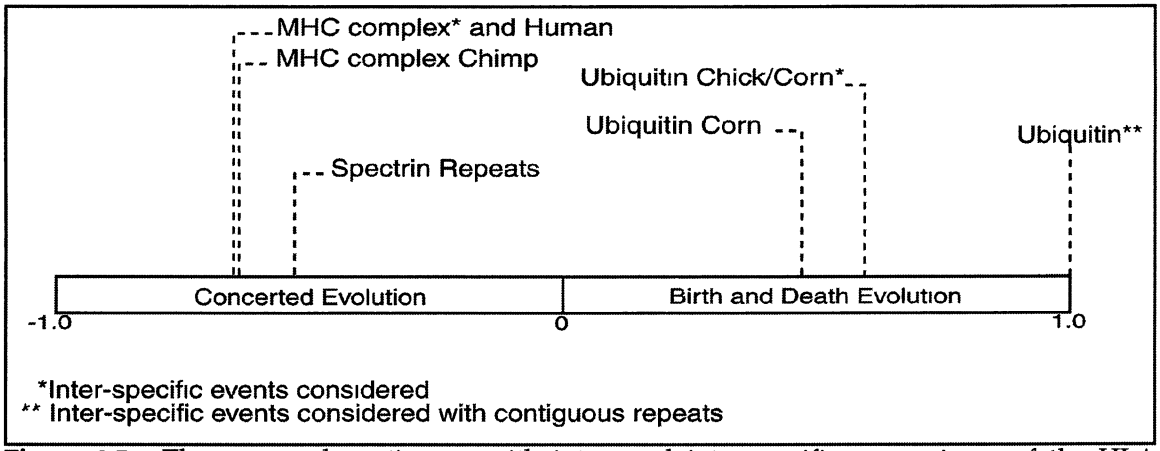



**Figure 2.5.** The proposed continuum with inter- and intra-specific comparisons of the HLA antigens from the MHC-complex, the α-actinin spectrin repeats of human, rat and mouse and the ubiquitins. Placement of the groups into the continuum was determined by the mean HI value for the group. Notice that speciation and higher levels of repeat structure in the ubiquitins increase the detection of birth-and-death evolution.

evolution (Figure 2.5). The MHC-complex data for the human/chimpanzee comparison yielded a homogenizing index value of –0.71 placing them closer to the concerted evolution extreme of the continuum than to the midpoint (Figure 2.5). The intra-specific comparisons for human and chimpanzee yielded HI values of –0.71 and –0.69 respectively (Figure 2.5).

**Discussion**

Use of the proposed HI allows birth-and-death evolution and concerted evolution to be considered as opposing extremes of a single process represented as a continuum (Figure 2.5). When the polyubiquitin loci are considered as a contiguous sequence and are compared between human, rat and mouse, intense birth-and-death evolution (HI=1.0) is predicted. This validates using the 100:1 ratio of nucleotide substitution to amino acid substitution as the birth-and-death extreme. When the sequences are split for comparison of internal repeat units within a single species, a lower HI value of 0.45 is obtained (Figure 2.5). In addition comparison of the internal repeats of the family between two distantly related species (corn and chicken) yields a HI value of 0.69. The relative strength of these values is of greater importance than the overall value. An extreme value for the inter-specific comparison of this gene family, considering contiguous repeats at a single locus, indicates that birth-and-death evolution has had a major impact on the evolution of the gene family. However a smaller value for the inter-specific analysis of the individual repeats indicates that the repeats have undergone birth-and-death evolution but with concerted evolution playing a more significant role when compared to the inter-specific analysis of the gene as a whole unit. In addition the intra-specific analysis of the internal repeats

yielded an even smaller value for the HI. From these data it can be inferred that inter-specific analysis yields more evidence of birth-and-death evolution than intra-specific analysis, indicating that speciation events may also be considered birth events. Inversely, the detection of concerted evolution is greater in intra-specific analyses since there is a lower level of birth-and-death evolution due to the absence of speciation. The finding of concerted evolution in such comparisons agrees with the findings of some authors (Sharp et al., 1987; Tan et al., 1993); however, Nei et al. (2000) presented overwhelming evidence of birth-and-death evolution in the ubiquitin gene family. By demonstrating that a scale can be derived to describe the modes of gene family evolution these analyses have provided a precedent for my work.

To summarize, the ubiquitin data indicate that speciation is, in fact, a birth process with regard to birth-and-death evolution (Figure 2.5). In addition, the suppositional 100:1 extreme is validated for birth-and-death evolution. Furthermore, regardless of the presence or absence speciation the homogenization indices indicate that the ubiquitin gene family has undergone higher levels of birth-and-death evolution than concerted evolution. Finding, birth-and-death evolution as the dominant form of homogenizing evolution in the ubiquitins is in direct agreement with the findings of others (Nei et al., 2000).

The MHC-complex genes considered between human and chimpanzee yield data consistent with concerted evolution, and the HI for this comparison is identical to the intra-specific human comparison and only minimally different from the value for the intra-specific chimpanzee comparison (Table 2.2). The placement of the gene family on the concerted evolution side of the continuum

provides support for the utility of the homogenization index in determining the contribution of concerted evolution as others have concluded a role for concerted evolution in the homogenization of this gene family (Klein, 1993; Wittzell, 1999). Furthermore, the similarity of the intra- and inter-specific analyses indicate that due to concerted evolution the ~5.5 million years of divergence between human and chimpanzee (Kumar and Hedges, 1998) has been insufficient to allow the formation of distinct evolutionary patterns within this gene family (Figure 2.5). In accord with the concept of a continuum, my data also support a role of birth-and-death evolution in the gene family as concluded by Nei et al. (1997). This agrees with the allusion of Wittzell et al. (1999) that concerted evolution and birth-and-death evolution need not be treated as mutually exclusive entities.

The spectrin repeats of the α-actinin gene family are believed to have undergone concerted evolution (Dubreil et al., 1991; Muse et al., 1997; Thomas et al., 1997). The family has not previously been scrutinized with regard to birth-and-death evolution and may therefore serve to corroborate the conclusion about the concerted evolution side of the continuum based on the MHC-complex loci. Our analyses indicate that the family has undergone a level of homogenizing evolution consistent with concerted evolution (Figure 2.5). In addition the inter-specific analysis yielded a value very close to the intra-specific HI value (Table 2.2). This relationship is expected under a concerted evolution model as the mutations allowed with minimal contribution from birth-and-death evolution are distributed rapidly to repeat sequences. Therefore I conclude that sequence homogeneity with regard to the spectrin repeats of the α-actinin gene family is

maintained by concerted evolution and that birth-and-death evolution makes a minimal contribution.

Although support for the index developed here, with respect to birth-and-death evolution is provided in an investigation of the ubiquitin gene family (Nei et al., 2000), the concerted evolution extreme is not as well supported in the literature. rRNA genes which are believed to have undergone intense concerted evolution (Nei et al., 1997) cannot be analyzed using the homogenization index as they are not expressed as proteins. However few will contest the idea that concerted evolution can rapidly distribute mutation to all members of a gene family and therefore gene families experiencing rapid concerted evolution in its extreme form will show zero amino acid substitutions and zero nucleotide substitutions. Realizing that such values yield little usable data, the extreme of a 1:1 relationship has been previously acknowledged (Nei et al., 2000) and was therefore applied in the proposed continuum.

In addition to the experimental validation of this model, a continuous model of homogenizing evolution can be described based on conceptually derived theory. Birth-and-death evolution and concerted evolution are conceptually different as previously presented; however, they may not be exclusive processes with respect to each other. If one considers birth-and-death evolution to be any rate of gene duplication (birth) coupled with any rate of gene death (pseudogene synthesis or gene deletion), then the only way for a gene family to arise is through the "birth" portion of birth-and-death evolution. In addition, by virtue of their persistence, gene families must experience some force which restricts its members from divergence-induced exclusion from the family.

Birth-and-death evolution may provide this force through rapid gene turnover and/or purifying selection, and concerted evolution may provide a similar force through gene conversion. Neither of these two processes arrest the action of the other. Due to its requirement of DNA sequence identity, it is probable that gene conversion occurs after a duplication event in gene families which experience birth-and-death evolution. A high rate of birth-and-death evolution may minimize this by converting genes to pseudogenes before concerted evolution has a chance to act. Taking all of this into account it can be inferred that all gene families experience some degree of integration between concerted evolution and birth-and-death evolution.

Birth is one part of birth-and-death evolution, and it is required for a gene family to arise. It can be argued that a family which experiences birth does not necessarily undergo birth-and-death evolution, however I disagree. The death process may occur very infrequently, yet there is always a chance of its occurrence. In addition a family which undergoes strong birth-and-death evolution is likely to undergo some small rate of concerted evolution due to similarity between two sequences following a duplication event. Thus concerted evolution and birth-and-death evolution occur concurrently in all gene families. Given these principles gene conversion actually requires birth-and-death evolution, and birth-and-death evolution may also require concerted evolution to assist purifying selection and to resurrect pseudogenes, or it may just be a consequence of birth-and-death evolution. In addition, gene conversion between a pseudogene produced by birth-and-death evolution and functional genes of a gene family predominately experiencing concerted evolution may allow the

family to experience slight divergence in order to diversify function. This expectation is consistent with the finding that class I genes and pseudogenes of the MHC-complex show similar proportions of synonymous and non-synonymous substitution, both among pseudogenes and between pseudogenes and functional genes (Hughes, 1994). Likewise, gene conversion between genes and pseudogenes is supported in the findings of Ohta (1982) concerning the generalized evolution of gene families. Gene conversion may instead allow for the evolution of a new gene family with divergent function from a gene family which generally experiences birth-and-death evolution. Therefore it is conceptually and theoretically probable that these two processes act continuously in the overall process of homogenizing evolution. In this case the frequency of each of these processes is inversely proportional to the frequency of the other. This conceptual leap to a single model does not appear to have been previously proposed.

In conclusion, homogenizing evolution is defined by the concurrent action of birth-and-death evolution and concerted evolution. Birth-and-death evolution may contribute more to inter-specific comparisons than to intra-specific comparisons due to speciation being a birth process. This requires awareness of the level of comparison being made from such analyses. Concerted evolution appears to be a much more confining process as intra-specific and inter-specific analyses of families at that end of the continuum show little variation. This accounts for the excess weight given to birth-and-death evolution in the uncorrected continuum (Figure 2.3A). In other words concerted evolution is very strong at its extreme when compared to birth-and-death evolution.

Concerted evolution in its extreme is also far less stable than birth-and-death evolution. Both birth-and-death evolution and concerted evolution are easily generalized to intra-specific analyses. However, it should be noted that if homologous gene families are available, both processes can be studied on all levels from the species level to phyla comparisons. In doing analyses which extend beyond the species boundary, it should be noted that birth-and-death-evolution is not restricted to the duplication of single genes, or arrays of genes, but can be extended to chromosomal duplication, whole genome duplication, lateral gene transfer and speciation. Accordingly, extinction and chromosomal deletion should be considered death processes. One should expect to find evidence of this in the data for individual gene families. Gene families which have experienced high levels of concerted evolution within a species will also exhibit evidence of the birth-and-death process which led to distinct taxonomic lineages.

Underlying the concept of this continuum is the lack of mutual exclusivity with regards to the concerted evolution and birth-and-death evolution. Therefore an inference about the contribution of each process may be made by considering that, at the mid-point, birth-and-death evolution makes the same contribution (50%) as concerted evolution (50%). Likewise both processes may approach 100% contribution at their respective extremes. However, since all gene families require a birth process and since concerted evolution is probable following gene duplication, neither process can account for 100% of the homogenizing mechanisms at work. On the other hand, it is likely that gene families will be found which have experienced birth and no death. This does not

exclude death from the family; it just indicates a very low rate of birth-and-death evolution.

In addition to the broad categorical basis of the continuum, several alternative applications of the continuum present themselves. Associated with genetic engineering of gene family members is the risk of side reaction with other members of the gene family. Such reactions could prove fatal if the incorrectly targeted gene is essential. Due to the ability of birth-and-death evolution to tolerate nucleotide divergence, families which undergo this mode of evolution may be less vulnerable to sequence-based molecular cross reaction and thus more amenable to genetic engineering. Alongside such medically based use of the continuum lies its value for designing oligonucleotide primers from known sequences to be used in the polymerase chain reaction with intentions of amplifying orthologous genes among highly divergent species. In this case gene families which have undergone intense concerted evolution will exhibit greater levels of DNA sequence identity among species. Therefore the ability to design primers based on a highly divergent comparison will have an increased probability of success when compared to gene families which have undergone high levels of birth-and-death evolution.

*Chapter 3*

# THE G3.5 ANTIGEN


## Summary

**Alpha-actinins are cytoskeletal proteins which function to cross-link actin filaments in both muscle and non-muscle cells. There are four known isoforms of α-actinin, two of which are found primarily in muscle cells while the others are found in non-muscle cells. The G3.5 antigen is a protein recognized by a monoclonal antibody identified in an attempt to immunologically characterize the protein portion of human brain plaque homogenates taken from individuals with multiple sclerosis. Later characterization of this protein has yielded some evidence indicating that the G3.5 antigen serves as a molecular cross-link between actin filaments and the intermediate filament network. In addition sequencing of peptide fragments taken from purified G3.5 antigen has shown high levels of similarity to α-actinin in three regions. We used three divergent amino acids from one of these peptide fragments along with DNA sequences for human α-actinin 2 to design primers for use in RT-PCR to amplify the cDNA encoding the G3.5 antigen. As a result we isolated a 77bp cDNA which shows ~29% identity to**

**rat α-actinin 3. In addition, using deer (*Odocoileus virginianus*) tissue we confirmed the finding of others that the G3.5 antigen is found in the exosarcomeric domain of skeletal muscle and that it does not co-localize with sarcomeric α-actinin.**

## Introduction

The gene families encoding α-actinin(s), α- and β-spectrin(s), $\beta_{heavy}$ spectrin(s) and dystrophin(s) make-up the spectrin superfamily. This superfamily is characterized by the ability to bind actin and the presence of a gene family-specific number of spectrin repeats (Muse et al., 1997; Thomas et al., 1997). Alpha-actinins are a diverse family of cytoskeletal proteins of which there are four known isoforms. The diversity observed among the α-actinin gene family is a result of alternative splicing as well as multiple genes (Parr et al., 1992). These proteins are expressed in a wide variety of cell types including muscle, liver, kidney, small intestine and brain (Beggs et al., 1992). The sub-unit structure of α-actinin is such that the dimerized protein has the ability to cross-link actin filaments. It is generally thought that α-actinin exists in the anti-parallel homodimer form (reviewed by Blanchard et al., 1989). However it has been shown that both *in vitro* and *in vivo* heterodimerization does occur between the two striated-muscle-specific isoforms, α-actinins 2 and 3 (Chan et al., 1998). The α-actinin protein monomer consists of two calponin-homology domains at the N-terminus followed by four spectrin repeats and two EF-hand domains at the C-terminus. The calponin-homology domains comprise the actin-binding domain, and the EF-hand domains convey calcium sensitivity (reviewed by Dubreuil et al., 1991). Muscle specific α-actinins are calcium insensitive due to

non-functional EF-hand domains, and the non-muscle isoforms are calcium sensitive due to functional EF-hand domains. It is thought that binding of the functional EF-hand domains to $Ca^{2+}$ causes a conformational change within the protein structure. This conformational deformation has been proposed as the controlling factor in actin binding (Burridge et al., 1981; Noegel et al., 1987). Therefore the ability to bind calcium as well as tissue-specific localization divides the family into two groups which are composed of muscle and non-muscle isoforms. In addition, muscle specific isoforms are derived from two distinct evolutionary lineages leading to α-actinins 2 and 3. Non-muscle isoforms are also derived from two distinct lineages which lead to α-actinins 1 and 4 (Figure 1.3).

A putative member of the α-actinin gene family, the G3.5 antigen, was named for the monoclonal antibody (G3.5 mAB) which was found to bind to the protein. This antibody was originally isolated in an attempt to immunologically characterize the protein portion of human brain plaque homogenates derived from individuals with multiple sclerosis (Malhotra et al., 1984). Previous studies involving the G3.5 antigen have focused on identifying the sub-cellular localization and the binding characteristics of this protein as well as its amino acid sequence. These studies have yielded evidence that the G3.5 antigen, like α-actinin has the ability to bind F-actin. Furthermore, a 19 residue polypeptide sequenced from G3.5 antigen purified from rat shows ~80% identity to human α-actinin 2 (Price et al., 1993). In addition, Bolanos et al. (1998) isolated two additional peptide fragments from bovine both of which show 100% identity to sequences derived from human α-actinin 2. These data support the supposition

that the G3.5 antigen is an uncharacterized α-actinin isoform. In contrast to these findings, the G3.5 antigen has been found to localize in the exo-sarcomeric domain of rat skeletal muscle and to bind α-actinin and desmin simultaneously *in vitro*, two properties not generally associated with α-actinins (Bolanos et al., 1998; Price et al., 1993). It has been proposed that these contrasting data are indicative of an α-actinin isoform which has the ability to cross-link actin to the intermediate filament network (Price et al., 1993).

In order to determine whether the G3.5 antigen is an α-actinin isoform or a novel protein, we initiated a molecular genetic study of the protein. The focus of the study was to isolate and sequence the cDNA which encodes the G3.5 antigen. We employed RT-PCR, cloning and automated DNA sequencing. The expectation was that the complete sequence of the cDNA for the G3.5 antigen along with phylogenetic analyses would allow for the definitive classification of the G3.5 antigen. Due to the previously described amino acid identity between α-actinin 2 and the G3.5 antigen (Bolanos et al., 1998; Price et al., 1993), we used mouse α-actinin 2 as a parsimony guide along with the peptide fragment sequenced by Price et al. (1993) in designing oligonucleotides to be used in RT-PCR (Bolanos et al., 1998). It was expected that the ~20% amino acid divergence observed by Price et al. (1993) would allow sufficient nucleotide divergence between known α-actinins and the G3.5 fragment for PCR-based purification and amplification of the cDNA encoding the antigen. In addition to this genetic methodology, we used immunostaining and confocal microscopy to confirm the localization of the G3.5 antigen in skeletal muscle (Bolanos et al., 1998 and Price et al., 1993). I extend their studies to deer skeletal muscle.

## Materials and Methods

### *Reverse Transcription*

The sequences for all primers used in reverse transcription and PCR are listed in Table 3.1. Poly(A)mRNA from rat skeletal muscle was obtained from Sigma (St. Louis, MO). A 2.0μg portion was reverse transcribed using the primer RAA4 which was designed based on the *Mus musculus* sequence for α-actinin 3 (AF093775) and is identical to *Homo sapiens* α-actinin 2 (M86406)(see Table 2.2) at all but 3 of 22 bases. The pre-reaction consisted of 2.5μl 400mM

**Table 3.1.** Primers used in reverse transcription, PCR and sequencing of a clone which may code for the G3.5 antigen.

| Primer Name | Sequence 5′ to 3′ |
|---|---|
| RAA4 | CTGCTGCCGGGCCAGCTCCTCC |
| RAA5 | GGCTCAGGCGCAGCTTGGTCTGC |
| G35UNQ-2 | CTTGCAATGCAAATCTTTGTGAAACACC |
| FWDPUC | CCATGATTACGAATTCGAGCTCGG |
| RVSPUC | TGCAGGTCGACTCTAGAGGAT |

Tris-HCl, 2.5μl 400mM KCl, 1μl 300mM $MgCl_2$, 5μl 100mM dithiothreitol (DTT), 2.5μl 10mM dNTP, 2.0μl actinomycin D (2mg/ml), 0.38μl RNAguard (Amersham, Piscataway, NJ) and 12.83μl cellulose-nitrate-filtered (0.22μm) $dH_20$. The pre-reaction was then added to the reaction mix. The reaction mix was prepared as follows: a solution of 1μl RAA4 (10pmol/μl), 3.57μl (2.0μg) poly(A)mRNA and 16.43μl RNase free $dH_20$ was incubated at 65°C for 15 minutes. After combining the reaction mix and the pre-reaction mix, 0.8μl (16u) of AMV reverse transcriptase was added. The reaction was then allowed to incubate for 1 hour at 50°C. Following incubation, the reaction was diluted with 450μl of 10mM Tris-HCl/10mM EDTA (pH 7.5). It was then stored at -30°C.

*cDNA Amplification*

In order to amplify the cDNA corresponding to rat G3.5 antigen, the primers G35UNQ-2 and RAA5 were used in a PCR reaction which incorporated the product from the above reverse transcription reaction as template. The PCR reaction consisted of 10μl 10X standard PCR buffer, 1μl RAA5 (10pmol/μl), 1μl G35UNQ-2 (10pmol/μl), 2.5μl diluted cDNA, 1μl dNTP (10mM), 62.5μl $dH_2O$, 10μl 17.5mM $MgCl_2$ and 0.5μl (2.5 units) *Taq* polymerase. The reaction was then subjected to thermal cycling according to the following cycling parameters: 94°C for 3 minutes; 40 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 4 minutes. This regimen was followed by a final extension at 72°C for 10 minutes coupled with a 4°C hold. In order to evaluate amplification the samples were subjected to electrophoresis in a 0.8% agarose gel, stained with ethidium bromide and visualized on a UV light table. Once confirmation of amplification was obtained, the PCR reaction was purified using a Qia-Prep PCR purification kit (Qiagen,Valencia, CA). Following purification, 2μl of the product was subjected to electrophoresis, as before, to determine the concentration of the extracted product. Visual comparison of the sample to the plasmid pGEM (ABI, Foster City, CA) was performed in order to assess the concentration of the product.

*Cloning*

In order to minimize the possibility of heterogeneous product, the clean PCR reaction was cloned into *E. coli* (strain JM105) using pUC 18 as the cloning vector. Ligation was carried out using a Sureclone Ligation kit (Amersham, Piscatoway, NJ). This kit incorporates plasmid and product using a blunt end ligation. For this blunt end ligation, 700ng of DNA from the above PCR reaction

was used. Subsequently, ligation was carried out using 140ng of the blunt end PCR product and 50ng of dephosphorylated pUC 18. Transformation was carried out by placing a well isolated colony of *E. coli*, which was cultured for 24 hours on an LB agar plate, into 250µl ice-cold $CaCl_2$ (0.05M) with 5µl of the ligated product from the Sureclone Ligation kit protocol. The tube containing the cells, recombinant plasmid and $CaCl_2$ was then agitated using a vortex mixer, placed on ice for 15 minutes, incubated at 42°C for 90 seconds, placed back on ice for 2 minutes followed by the addition of 250µl LB broth. The sample was then allowed to incubate at 37°C for 30 minutes as a recovery period for the cells. The sample was then centrifuged at 500Xg for 5 minutes followed by removal of the supernatant. The pelleted cells were resuspended in 100µl of LB broth and spread onto LB/Xgal/IPTG/ampicillin culture plates using a sterile glass rod. The cells were then incubated at 37°C for 24 hours.

After incubation, recombinant colonies were used to inoculate 2ml LB/ampicillin broth which was then incubated at 37°C for 24 hours in a shaking incubator. Following growth in liquid media, plasmid DNA was extracted using the alkali lysis method described in *Molecular Cloning* (Maniatis et al., 1989) with a few modifications. These modifications include substitution of a 3 minute drying of the final DNA pellet in a vacuum centrifuge rather than an air-dry, substitution of sodium acetate for potassium acetate in solution III (due to availability) and RNase exclusion from the final TE buffer used to re-dissolve the DNA pellet. Reconstituted DNA was then subjected to electrophoresis as before in order to evaluate the extracted plasmid DNA.. Once extraction was confirmed, the plasmid DNA was used as template in a PCR reaction. This was necessary as

several bands appeared in the electrophoretic gel from the plasmid samples. The PCR reaction was carried out using the same conditions as before with the pUC specific primers FWDPUC and RVSPUC (Table 3.1) and a 60°C annealing step. After completion of the PCR, the entire sample was subjected to electrophoresis as before. The resulting band which corresponded to the molecular weight of the product used in the ligation reaction was then gel extracted for use in subsequent automated DNA sequencing.

*Automated DNA Sequencing*

Cycle sequencing reactions were performed using 40-100ng of gel-extracted product as template with an Applied Biosystems second generation Big-Dye dye terminator cycle sequencing kit (ABI, Foster City, CA). Sequencing reactions consisted of 1μl template, 0.5μl primer, 3μl Big Dye and 4.5μl $dH_2O$. The reactions were carried out for 35 cycles of 96°C for 5 seconds, 50°C for 10 seconds and 60°C for 4 minutes. Clones were sequenced using the primers FWDPUC and RVSPUC.

*Sequence Analysis*

The contiguous sequence was assembled using the computer program Sequencher (v. 4.1) (Gene Codes Corp., Arbor, MI). Existing databases were searched for fragment similarity using the BLAST (Basic Local Alignment Search Tool) available at www.ncbi.nlm.nih.gov/BLAST/. LALIGN was used to search for multiple regions of similarity between two sequences (www.ch.embnet.org/software/LALIGN_form.html). Alignment of sequenced clones to known α-actinin sequences was performed on the GENESTREAM server using ALIGN (http://www2.igh.cnrs.fr/bin/align-guess.cgi).

*Immunostaining*

Deer skeletal muscle was fixed by freezing in a $CO_{2(solid)}$/hexane bath. Whole tissue was embedded in tissue-tek obtained from Sakura Fintek (Torrance, CA) and cryo-sectioned in 25μm thick sections. Sections were placed onto poly-L-lysine (0.1%) coated coverslips and allowed to dry overnight at 4°C. Sections were then permeabilized by washing 3X in phosphate buffered saline (PBS) containing 0.2% Tween-20 for 5 minutes each wash. Following permeabilization the sections were blocked in 20% powdered milk in PBS for 2 hours then labeled with the primary antibodies mouse anti-sarcomeric α-actinin (IgG)(1:100 in PBS) and mouse anti-G3.5 (IgM)(no dilution) for 2 hours followed by 3, 10 minute washes in PBS. Following labeling with primary antibodies, the sections were labeled in two steps with secondary antibodies specific for each primary antibody. The secondary antibodies were anti-mouse IgG (Cy-5)(1:1000) and anti-mouse IgM (FITC)(1:100). The secondary antibodies were applied in the same manner as the primary antibodies. Following the final wash, the sections were mounted in 90% glycerol containing 0.2% p-phenylenediamine, coverslips were sealed to slides using fingernail polish and slides were viewed on a Bio-Rad MRC 1024 confocal microscope.

## Results

*Sequencing of cDNA Clones*

A single clone was isolated and sequenced. This clone contained an insert within the multiple cloning site of pUC 18 which proved to be 132bp in length. The portion of the insert which was not subject to site directed mutation via PCR was 77bp in length. Translation of this product in six frames yielded one

inferred amino acid sequence which exhibited a region of identity to the G3.5 antigen sequence found by Price et al., (1993). In addition three of the inferred amino acids correspond to those found to be unique to the G3.5 antigen (Price et al., 1993) (Figure 3.1).

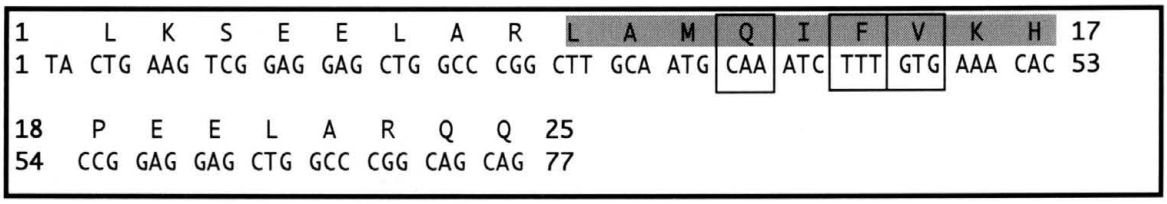



**Figure 3.1.** A single open reading frame of a clone from rat skeletal muscle which may encode the G3.5 antigen. The shaded amino acids are identical to a region found in the G3.5 antigen. In addition they found the amino acids corresponding to the boxed codons to be unique to the G3.5 antigen (Price et al., 1993).

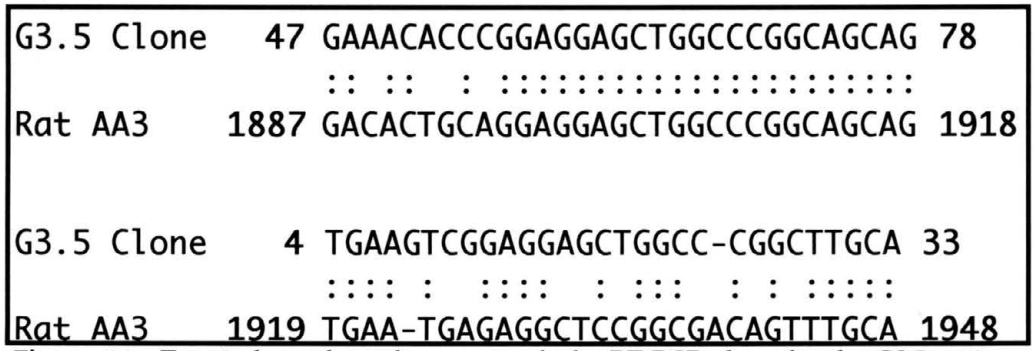



**Figure 3.2.** Two independent alignments of the RT-PCR clone for the G3.5 antigen which show 87.1% identity (top) and 66.7% identity (bottom). Identity is indicated by a dotted line between nucleotides.

A BLAST search of the translated fragment yielded no significantly similar matches. A BLAST search of the fragment at the nucleotide level yielded 100% similarity to *Mus musculus* α-actinin 3 (AF093775) from nucleotides 11-27 and 56-77 (Figure 3.2). Both regions align to the same region on the *Mus musculus* α-actinin 3 sequence. A search for multiple regions of identity between the cloned

fragment and rat α-actinin 3 (AF450248)(Figure 1.1) yielded two regions of similarity (Figure 3.2). Using the fragment with the highest identity as a guide, the sequence was aligned to rat α-actinin 3. The fragment was 29.8% identitical to rat α-actinin 3 in this region (Figure 3.3).

```
G3.5 Clone   1 TACTGAAG-TCGGAGGAGCT------------GGCCCGG 26
                 :   :: ::: ::::  :             : : :::
Rat AA3   1819 ACCCTCAGCTCGCAGGACATCAACAATAAGTGGGACACGG 1859

G3.5 Clone  27 CTTGCAA-----TGCAAATCTTTGT---GAAACACCCGGA 59
                : ::     : :  : :  ::    :: ::  : :::
Rat  AA3 1860 TCAGAAAGCTGGTACCCAGCCGTGACCAGACACTGCAGGA 1900

G3.5 Clone   60 GGAGCTGGCCCGGCAGCAG------------ 77
                ::::::::::::::::::
Rat  AA3   1901 GGAGCTGGCCCGGCAGCAGGTGAATGAGAGGC 1933
```

**Figure 3.3.** An alignment of the sequence for rat α-actinin 3 (AF450248) to the sequence for a putative G3.5 antigen clone. This alignment exhibits 29.8% identity between these fragments. Identity is indicated by a dotted line between nucleotides. The region of α-actinin consistent with the alignment is found predominately within the fourth spectrin repeat with ~20bp in the third spectrin repeat.

### *Immunostaining*

Deer skeletal muscle sections labeled for sarcomeric α-actinin and the G3.5 antigen were visualized by confocal microscopy. Visualization of the skeletal muscle sections labeled for both the G3.5 antigen and sarcomeric α-actinin revealed a distinct localization of each (Figure 3.4). The G3.5 antigen (green) appears to completely lack co-localization with sarcomeric α-actinin (blue) in Figure 3.4.

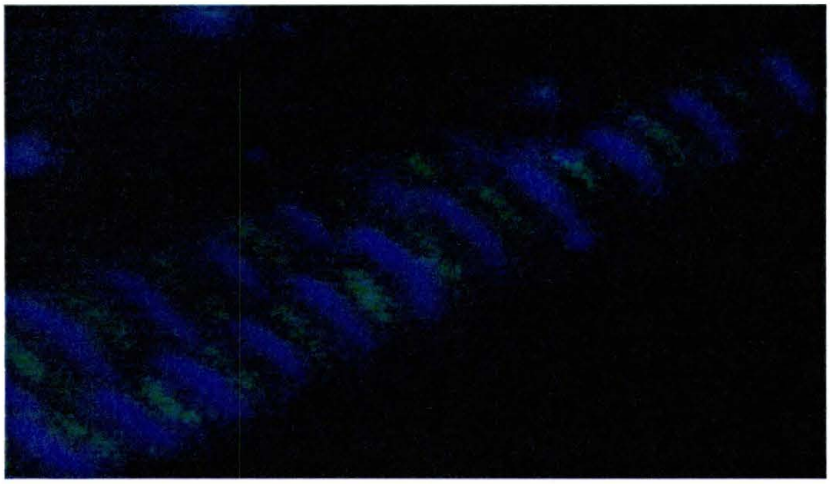


**Figure 3.4.** A confocal micrograph of a 400nm thick optical plane taken of a 25μm thick section of deer skeletal muscle labeled for the G3.5 antigen (green) and sarcomeric α-actinin (Blue). This image demonstrates that the G3.5 antigen does not co-localize with sarcomeric α-actinin.

## Discussion

The cDNA clone isolated appears to be related to α-actinin as indicated by the similarity to mouse and rat α-actinin 3. This may be indicative of a common ancestor of these two proteins. However due to the length of the clone relative to the length of α-actinin and the minimal similarity of the region in a full alignment (Figure 3.3), this provides little indication that the cloned sequence is that of an α-actinin as most α-actinin sequences show greater than 80% identity to each other (Beggs, et al., 1998). It is not likely that the cDNA from which the clone was derived will prove to be a member of the α-actinin gene family as gene family members must show at least 50% amino acid similarity (Graur and Li, 1999). However, with such a small portion of sequence considered there are few conclusions which can be drawn; instead inferences and further research questions are plentiful.

Based on nucleotide similarity, it is possible that this fragment belongs to a gene family which is a descendent of α-actinin 3. If this is the case, then the family has undergone rapid evolution in order to diverge to the point which it is no longer a member of the spectrin superfamily. This divergence would have had to occur over no more than ~273 million years as indicated by the divergence of α-actinins 2 and 3 (Figure 1.3). Or perhaps the clone is derived from a descendent of α-actinin 2 which is also an ancestor of α-actinin 3 placing the divergence of the cDNA from the α-actinins at ~311-273mya (Figure 1.3). This scenario may explain the similarity which fragments of the sequenced protein show to α-actinin 2 (Bolanos et al., 1998; Price et al., 1993) as well as the similarity between the clone and rat α-actinin 3. A phylogenetic investigation may resolve these questions; however conclusions drawn from such an analysis on a 77bp fragment would be speculative. Therefore I did not perform a phylogenetic analysis of this fragment.

The sub-cellular localization study (Figure 3.4) confirms earlier findings that the G3.5 antigen does not co-localize with sarcomeric α-actinin and that it is found in the exosarcomeric domain (Bolanos et al., 1998; Price et al., 1993). This supports a novel function due to divergent localization as well as a novel molecular classification of the G3.5 antigen. When coupled with an analysis of the sequenced clone, this finding seems to indicate that the G3.5 antigen is remotely related to α-actinin and that it is likely that the protein is unique. This conclusion may be made as both molecular and cellular evidence indicate deviation from, rather than convergence on the standard classification of α-actinin isoforms. Extension of the sequencing effort to include the entire cDNA

sequence for the clone isolated will allow for further molecular and cellular experimentation which will provide evidence to place the clone in the proper place with regard to the spectrin superfamily as well as the α-actinin gene family.

# GLOSSARY

**AMV reverse transcriptase.** Reverse transcriptase purified from the avian murine virus; used for *in vitro* conversion of RNA into a complimentary DNA molecule (cDNA).

**Big Dye.** A solution of fluor-coupled 2,3-dideoxynucleotides, 2-deoxynucleotides, *Taq* polymerase, PCR buffer and $dH_2O$. Used in cycle sequencing reactions.

**Birth-and-death evolution.** A mode of gene family evolution which involves repeated gene duplication and gene death (rapid gene turnover).

**Bootstrap value.** A reliability value which indicates the percentage of trees with the observed branching pattern which were derived from many statistically bootstrapped (rearranged) data-sets.

**Cassette style duplication.** Duplication of more than one entity due to their inclusion in a larger entity (the cassette).

**Concerted evolution.** A mode of gene family evolution in which homologous genes remain similar by repeated exchange of genetic material. See gene conversion.

**Contiguous.** A continuous series of entities which share a boundary. With regard to DNA a series of fragments which share sequence identity at the 5′ and/or 3′ end(s).

**Continuum.** A succession whose parts can only be arbitrarily divided.

**Convergence.** Obtaining the same end-point as a result of taking different paths.

**Cytoskeleton.** Matrix of proteins which provide structural support to the cell.

**Divergence.** Observed change among homologous molecules.

**Evolution.** Change through time. A change of the conformation, constitutive make-up and/or distribution of a heritable molecule.

**Evolutionary period.** A time period during which the same mode of evolution is experienced throughout.

**Gene conversion.** A non-reciprocal genetic exchange with a donor molecule and an acceptor molecule. The donor remains unchanged yet converts the acceptor so that both molecules are identical in the region of conversion.

**Gene Family.** A group of more than one gene which has a common ancestor and shares >50% amino acid similarity.

**HKY model of evolution.** A model of evolution under which unequal base frequencies and a two-parameter model of substitution are assumed.

**Homodimer.** A protein molecule consisting of two identical sub-units.

**Homogenization.** To make all components into a single form. To make all components alike.

**Homogenizing evolution.** A mode of evolution which favors homogenous gene families. Homogeneity may be observed at the amino acid or DNA level.

**Homology.** Similarity as a result of common descent. See orthology and paralogy.

**Immunostaining.** The use of fluor-coupled antibodies to label specific proteins for visualization in fluorescent microscopy.

**Inter-specific.** Between species.

**Intra-specific.** Within a species.

**Isoform.** Generally the same as paralogue. A protein molecule which is only slightly different than its homologs. Isoforms have the same general amino acid sequence, generally perform the same function and are either expressed in the same manner in one tissue or are expressed in different cell types yet perform similar functions.

**Kimura-2-parameter correction.** A distance correction which assumes that transitions and transversions occur at different rates.

**Maximum Parsimony (MP).** A method of phylogenetic reconstruction which favors the tree with a minimal number of evolutionary changes which contribute to the tree.

**Neighbor Joining (NJ).** A method of phylogenetic reconstruction which favors the shortest tree possible.

**Nemaline myopathy.** A degenerate muscle disease marked by the formation of rod-shaped protein plaques (nemaline rods) adjacent to the sarcomere.

**Oligonucleotide primer.** A short fragment of DNA (generally <30bp) which is used to induce DNA polymerase to synthesize a complimentary fragment of DNA using a template.

**Orthology.** Homology among genes as a result of speciation.

**Paralogy.** Homology among genes as a result of gene duplication.

**Parsimony.** Also known as "Ockham's Razor." The derivation of the simplest conclusion by minimizing the number of assumptions or steps needed to arrive at a conclusion.

**PCR (Polymerase Chain Reaction).** A method which uses DNA polymerase to amplify a single fragment of DNA *in vitro*.

**Phylogenetic.** The inferred separation of segregated bio-entities. Phylogenetics is the study of these relationships.

**Primer.** See **oligonucleotide primer**.

**Purifying selection.** Selection which favors paralogues which encode a single gene product. This form of selection does not restrict synonymous substitution.

**Quartet Puzzling (QP).** A method of phylogenetic reconstruction in which quartets of sequences are assembled using the maximum likelihood method of phylogenetic reconstruction followed by combination of the quartets to construct the entire tree.

**Race (Rapid Amplification of cDNA Ends).** A method used to amplify the 5′ and 3′ ends of a cDNA molecule.

**Relatedness.** The degree of spatial and temporal similarity of a group of heritable molecules.

**Residue.** An amino acid.

**RT-PCR (Reverse Transcription-Polymerase Chain Reaction).** Conversion of an RNA molecule to its complimentary DNA (cDNA) molecule using reverse transcriptase followed by amplification of a specific cDNA using the PCR.

**Saturation level.** The level at which accurate detection becomes impossible due to homoplasy.

**Selective pressure.** Pressure to maintain a beneficial mutation due to the benefit, with regard to fitness, which the resulting phenotype imparts.

**Superfamily.** A group of genes which are clearly homologous yet exhibit <50% amino acid similarity.

**Synonymous site.** A nucleotide position which may or may not exhibit divergence but is part of a single codon which does not exhibit an amino acid substitution in a pair-wise comparison.

***Taq* polymerase.** DNA polymerase used in the PCR due to its tolerance of high temperatures. Purified from the archeae *Thermus aquaticus*.

# APPENDIX



Raw data for the homogenization index tabulations (Chapter 2).

*Shaded boxes indicate that 1 was used as an approximation for 0 character change.

Spectrin Repeat data for 116 amino acids and 346 nucleotides.
The first letter in the accession # indicates the organism; the last number indicates the repeat. All others are the accession numbers.
H=human, R=rat and M=mouse.

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| HM86406 4 | MAF24864 4 | 2 | 29 | 0.17 |
| HM86407 4 | MAF24864 4 | 26 | 87 | -0.47 |
| HM86407 4 | HM86406 4 | 25 | 91 | -0.43 |
| Rat AA3 4 | MAF24864 4 | 25 | 82 | -0.48 |
| Rat AA3 4 | HM86406 4 | 25 | 88 | -0.45 |
| Rat AA3 4 | HM86407 4 | 3 | 31 | 0.02 |
| RAF19090 4 | MAF24864 4 | 28 | 86 | -0.51 |
| RAF19090 4 | HM86406 4 | 26 | 86 | -0.48 |
| RAF19090 4 | HM86407 4 | 36 | 110 | -0.51 |
| RAF19090 4 | Rat AA3 4 | 36 | 114 | -0.49 |
| HD89980 4 | MAF24864 4 | 27 | 76 | -0.55 |
| HD89980 4 | HM86406 4 | 25 | 80 | -0.49 |
| HD89980 4 | HM86407 4 | 35 | 107 | -0.51 |
| HD89980 4 | Rat AA3 4 | 35 | 109 | -0.50 |
| HD89980 4 | RAF19090 4 | 2 | 27 | 0.14 |
| RAF11538 4 | MAF24864 4 | 24 | 93 | -0.41 |
| RAF11538 4 | HM86406 4 | 22 | 91 | -0.38 |
| RAF11538 4 | HM86407 4 | 30 | 102 | -0.46 |
| RAF11538 4 | Rat AA3 4 | 30 | 106 | -0.45 |
| RAF11538 4 | RAF19090 4 | 16 | 85 | -0.27 |
| RAF11538 4 | HD89980 4 | 14 | 78 | -0.25 |
| HM95178 4 | MAF24864 4 | 24 | 80 | -0.47 |
| HM95178 4 | HM86406 4 | 22 | 84 | -0.41 |
| HM95178 4 | HM86407 4 | 30 | 96 | -0.49 |
| HM95178 4 | Rat AA3 4 | 30 | 95 | -0.49 |
| HM95178 4 | RAF19090 4 | 16 | 77 | -0.31 |
| HM95178 4 | HD89980 4 | 14 | 68 | -0.31 |
| HM95178 4 | RAF11538 4 | 1 | 30 | 0.48 |
| MAF24864 1 | MAF24864 4 | 92 | 179 | -0.71 |
| MAF24864 1 | HM86406 4 | 92 | 184 | -0.69 |
| MAF24864 1 | HM86407 4 | 93 | 182 | -0.70 |
| MAF24864 1 | Rat AA3 4 | 93 | 181 | -0.71 |
| MAF24864 1 | RAF19090 4 | 93 | 184 | -0.70 |
| MAF24864 1 | HD89980 4 | 92 | 186 | -0.69 |

| | | | | |
|---|---|---|---|---|
| MAF24864 1 | RAF11538 4 | 90 | 184 | -0.68 |
| MAF24864 1 | HM95178 4 | 90 | 179 | -0.70 |
| HM86406 1 | MAF24864 4 | 91 | 178 | -0.70 |
| HM86406 1 | HM86406 4 | 91 | 183 | -0.69 |
| HM86406 1 | HM86407 4 | 92 | 179 | -0.71 |
| HM86406 1 | Rat AA3 4 | 92 | 183 | -0.70 |
| HM86406 1 | RAF19090 4 | 92 | 186 | -0.69 |
| HM86406 1 | HD89980 4 | 91 | 190 | -0.67 |
| HM86406 1 | RAF11538 4 | 89 | 186 | -0.67 |
| HM86406 1 | HM95178 4 | 89 | 180 | -0.69 |
| HM86406 1 | MAF24864 1 | 2 | 23 | 0.07 |
| HM86407 1 | MAF24864 4 | 91 | 185 | -0.69 |
| HM86407 1 | HM86406 4 | 91 | 191 | -0.67 |
| HM86407 1 | HM86407 4 | 91 | 180 | -0.70 |
| HM86407 1 | Rat AA3 4 | 91 | 186 | -0.68 |
| HM86407 1 | RAF19090 4 | 92 | 193 | -0.67 |
| HM86407 1 | HD89980 4 | 91 | 190 | -0.67 |
| HM86407 1 | RAF11538 4 | 90 | 188 | -0.67 |
| HM86407 1 | HM95178 4 | 90 | 187 | -0.68 |
| HM86407 1 | MAF24864 1 | 22 | 69 | -0.50 |
| HM86407 1 | HM86406 1 | 20 | 69 | -0.46 |
| Rat AA3 1 | MAF24864 4 | 92 | 186 | -0.69 |
| Rat AA3 1 | HM86406 4 | 92 | 193 | -0.67 |
| Rat AA3 1 | HM86407 4 | 92 | 186 | -0.69 |
| Rat AA3 1 | Rat AA3 4 | 92 | 185 | -0.69 |
| Rat AA3 1 | RAF19090 4 | 93 | 196 | -0.67 |
| Rat AA3 1 | HD89980 4 | 92 | 193 | -0.67 |
| Rat AA3 1 | RAF11538 4 | 91 | 192 | -0.67 |
| Rat AA3 1 | HM95178 4 | 91 | 189 | -0.68 |
| Rat AA3 1 | MAF24864 1 | 21 | 72 | -0.46 |
| Rat AA3 1 | HM86406 1 | 19 | 70 | -0.43 |
| Rat AA3 1 | HM86407 1 | 1 | 32 | 0.51 |
| RAF19090 1 | MAF24864 4 | 92 | 182 | -0.70 |
| RAF19090 1 | HM86406 4 | 92 | 190 | -0.68 |
| RAF19090 1 | HM86407 4 | 92 | 184 | -0.69 |
| RAF19090 1 | Rat AA3 4 | 92 | 184 | -0.69 |
| RAF19090 1 | RAF19090 4 | 91 | 193 | -0.67 |
| RAF19090 1 | HD89980 4 | 90 | 187 | -0.68 |
| RAF19090 1 | RAF11538 4 | 90 | 196 | -0.66 |
| RAF19090 1 | HM95178 4 | 90 | 186 | -0.68 |
| RAF19090 1 | MAF24864 1 | 18 | 61 | -0.46 |
| RAF19090 1 | HM86406 1 | 16 | 62 | -0.41 |
| RAF19090 1 | HM86407 1 | 26 | 64 | -0.60 |
| RAF19090 1 | Rat AA3 1 | 25 | 70 | -0.55 |
| HD89980 1 | MAF24864 4 | 91 | 182 | -0.69 |
| HD89980 1 | HM86406 4 | 91 | 188 | -0.68 |
| HD89980 1 | HM86407 4 | 91 | 186 | -0.68 |
| HD89980 1 | Rat AA3 4 | 91 | 185 | -0.69 |

| | | | | |
|---|---|---|---|---|
| HD89980 1 | RAF19090 4 | 90 | 193 | -0.66 |
| HD89980 1 | HD89980 4 | 89 | 191 | -0.66 |
| HD89980 1 | RAF11538 4 | 89 | 197 | -0.65 |
| HD89980 1 | HM95178 4 | 89 | 188 | -0.67 |
| HD89980 1 | MAF24864 1 | 18 | 59 | -0.48 |
| HD89980 1 | HM86406 1 | 16 | 61 | -0.41 |
| HD89980 1 | HM86407 1 | 24 | 63 | -0.58 |
| HD89980 1 | Rat AA3 1 | 23 | 65 | -0.54 |
| HD89980 1 | RAF19090 1 | 2 | 25 | 0.10 |
| RAF11538 1 | MAF24864 4 | 91 | 184 | -0.69 |
| RAF11538 1 | HM86406 4 | 91 | 196 | -0.66 |
| RAF11538 1 | HM86407 4 | 92 | 185 | -0.69 |
| RAF11538 1 | Rat AA3 4 | 92 | 183 | -0.70 |
| RAF11538 1 | RAF19090 4 | 92 | 193 | -0.67 |
| RAF11538 1 | HD89980 4 | 91 | 196 | -0.66 |
| RAF11538 1 | RAF11538 4 | 90 | 198 | -0.65 |
| RAF11538 1 | HM95178 4 | 90 | 193 | -0.66 |
| RAF11538 1 | MAF24864 1 | 16 | 56 | -0.45 |
| RAF11538 1 | HM86406 1 | 14 | 59 | -0.37 |
| RAF11538 1 | HM86407 1 | 19 | 62 | -0.48 |
| RAF11538 1 | Rat AA3 1 | 18 | 65 | -0.44 |
| RAF11538 1 | RAF19090 1 | 11 | 45 | -0.38 |
| RAF11538 1 | HD89980 1 | 11 | 44 | -0.39 |
| HM95178 1 | MAF24864 4 | 92 | 180 | -0.70 |
| HM95178 1 | HM86406 4 | 92 | 185 | -0.69 |
| HM95178 1 | HM86407 4 | 93 | 175 | -0.72 |
| HM95178 1 | Rat AA3 4 | 93 | 182 | -0.70 |
| HM95178 1 | RAF19090 4 | 93 | 186 | -0.69 |
| HM95178 1 | HD89980 4 | 92 | 185 | -0.69 |
| HM95178 1 | RAF11538 4 | 91 | 191 | -0.67 |
| HM95178 1 | HM95178 4 | 91 | 186 | -0.68 |
| HM95178 1 | MAF24864 1 | 18 | 56 | -0.50 |
| HM95178 1 | HM86406 1 | 16 | 57 | -0.44 |
| HM95178 1 | HM86407 1 | 18 | 55 | -0.51 |
| HM95178 1 | Rat AA3 1 | 17 | 57 | -0.47 |
| HM95178 1 | RAF19090 1 | 13 | 45 | -0.46 |
| HM95178 1 | HD89980 1 | 13 | 42 | -0.49 |
| HM95178 1 | RAF11538 1 | 2 | 20 | 0.01 |
| MAF24864 2 | MAF24864 4 | 89 | 206 | -0.63 |
| MAF24864 2 | HM86406 4 | 89 | 210 | -0.62 |
| MAF24864 2 | HM86407 4 | 89 | 201 | -0.64 |
| MAF24864 2 | Rat AA3 4 | 88 | 207 | -0.62 |
| MAF24864 2 | RAF19090 4 | 88 | 200 | -0.64 |
| MAF24864 2 | HD89980 4 | 87 | 198 | -0.64 |
| MAF24864 2 | RAF11538 4 | 87 | 200 | -0.63 |
| MAF24864 2 | HM95178 4 | 87 | 199 | -0.64 |
| MAF24864 2 | MAF24864 1 | 91 | 217 | -0.62 |
| MAF24864 2 | HM86406 1 | 90 | 221 | -0.60 |

| | | | | |
|---|---|---|---|---|
| MAF24864 2 | HM86407 1 | 90 | 210 | -0.63 |
| MAF24864 2 | Rat AA3 1 | 91 | 214 | -0.62 |
| MAF24864 2 | RAF19090 1 | 89 | 221 | -0.60 |
| MAF24864 2 | HD89980 1 | 88 | 217 | -0.60 |
| MAF24864 2 | RAF11538 1 | 90 | 220 | -0.61 |
| MAF24864 2 | HM95178 1 | 91 | 214 | -0.62 |
| HM86406 2 | MAF24864 4 | 89 | 202 | -0.64 |
| HM86406 2 | HM86406 4 | 89 | 204 | -0.63 |
| HM86406 2 | HM86407 4 | 89 | 195 | -0.65 |
| HM86406 2 | Rat AA3 4 | 88 | 201 | -0.64 |
| HM86406 2 | RAF19090 4 | 88 | 192 | -0.66 |
| HM86406 2 | HD89980 4 | 87 | 195 | -0.64 |
| HM86406 2 | RAF11538 4 | 87 | 200 | -0.63 |
| HM86406 2 | HM95178 4 | 87 | 196 | -0.64 |
| HM86406 2 | MAF24864 1 | 91 | 215 | -0.62 |
| HM86406 2 | HM86406 1 | 90 | 220 | -0.61 |
| HM86406 2 | HM86407 1 | 90 | 214 | -0.62 |
| HM86406 2 | Rat AA3 1 | 91 | 215 | -0.62 |
| HM86406 2 | RAF19090 1 | 89 | 219 | -0.60 |
| HM86406 2 | HD89980 1 | 88 | 215 | -0.61 |
| HM86406 2 | RAF11538 1 | 90 | 220 | -0.61 |
| HM86406 2 | HM95178 1 | 91 | 212 | -0.63 |
| HM86406 2 | MAF24864 2 | 1 | 32 | 0.51 |
| HM86407 2 | MAF24864 4 | 90 | 191 | -0.67 |
| HM86407 2 | HM86406 4 | 90 | 196 | -0.66 |
| HM86407 2 | HM86407 4 | 89 | 181 | -0.69 |
| HM86407 2 | Rat AA3 4 | 88 | 185 | -0.67 |
| HM86407 2 | RAF19090 4 | 89 | 189 | -0.67 |
| HM86407 2 | HD89980 4 | 88 | 188 | -0.66 |
| HM86407 2 | RAF11538 4 | 89 | 195 | -0.65 |
| HM86407 2 | HM95178 4 | 89 | 193 | -0.66 |
| HM86407 2 | MAF24864 1 | 91 | 216 | -0.62 |
| HM86407 2 | HM86406 1 | 91 | 217 | -0.62 |
| HM86407 2 | HM86407 1 | 90 | 207 | -0.63 |
| HM86407 2 | Rat AA3 1 | 91 | 208 | -0.64 |
| HM86407 2 | RAF19090 1 | 89 | 211 | -0.62 |
| HM86407 2 | HD89980 1 | 88 | 204 | -0.63 |
| HM86407 2 | RAF11538 1 | 90 | 211 | -0.62 |
| HM86407 2 | HM95178 1 | 92 | 201 | -0.66 |
| HM86407 2 | MAF24864 2 | 22 | 92 | -0.37 |
| HM86407 2 | HM86406 2 | 22 | 85 | -0.41 |
| Rat AA3 2 | MAF24864 4 | 90 | 201 | -0.65 |
| Rat AA3 2 | HM86406 4 | 90 | 205 | -0.64 |
| Rat AA3 2 | HM86407 4 | 90 | 196 | -0.66 |
| Rat AA3 2 | Rat AA3 4 | 89 | 190 | -0.67 |
| Rat AA3 2 | RAF19090 4 | 89 | 195 | -0.65 |
| Rat AA3 2 | HD89980 4 | 88 | 195 | -0.65 |
| Rat AA3 2 | RAF11538 4 | 89 | 199 | -0.65 |

| | | | | |
|---|---|---|---|---|
| Rat AA3 2 | HM95178 4 | 89 | 200 | -0.64 |
| Rat AA3 2 | MAF24864 1 | 91 | 208 | -0.64 |
| Rat AA3 2 | HM86406 1 | 91 | 214 | -0.62 |
| Rat AA3 2 | HM86407 1 | 89 | 208 | -0.63 |
| Rat AA3 2 | Rat AA3 1 | 90 | 203 | -0.64 |
| Rat AA3 2 | RAF19090 1 | 89 | 205 | -0.63 |
| Rat AA3 2 | HD89980 1 | 88 | 209 | -0.62 |
| Rat AA3 2 | RAF11538 1 | 90 | 208 | -0.63 |
| Rat AA3 2 | HM95178 1 | 92 | 205 | -0.65 |
| Rat AA3 2 | MAF24864 2 | 21 | 98 | -0.33 |
| Rat AA3 2 | HM86406 2 | 21 | 96 | -0.33 |
| Rat AA3 2 | HM86407 2 | 5 | 46 | -0.03 |
| RAF19090 2 | MAF24864 4 | 91 | 198 | -0.66 |
| RAF19090 2 | HM86406 4 | 91 | 199 | -0.65 |
| RAF19090 2 | HM86407 4 | 91 | 197 | -0.66 |
| RAF19090 2 | Rat AA3 4 | 90 | 200 | -0.65 |
| RAF19090 2 | RAF19090 4 | 91 | 190 | -0.67 |
| RAF19090 2 | HD89980 4 | 90 | 192 | -0.67 |
| RAF19090 2 | RAF11538 4 | 90 | 200 | -0.65 |
| RAF19090 2 | HM95178 4 | 90 | 193 | -0.66 |
| RAF19090 2 | MAF24864 1 | 93 | 216 | -0.63 |
| RAF19090 2 | HM86406 1 | 92 | 215 | -0.63 |
| RAF19090 2 | HM86407 1 | 92 | 213 | -0.63 |
| RAF19090 2 | Rat AA3 1 | 93 | 213 | -0.63 |
| RAF19090 2 | RAF19090 1 | 91 | 217 | -0.62 |
| RAF19090 2 | HD89980 1 | 90 | 213 | -0.62 |
| RAF19090 2 | RAF11538 1 | 91 | 217 | -0.62 |
| RAF19090 2 | HM95178 1 | 92 | 216 | -0.62 |
| RAF19090 2 | MAF24864 2 | 25 | 86 | -0.46 |
| RAF19090 2 | HM86406 2 | 25 | 87 | -0.45 |
| RAF19090 2 | HM86407 2 | 28 | 91 | -0.48 |
| RAF19090 2 | Rat AA3 2 | 29 | 92 | -0.49 |
| HD89980 2 | MAF24864 4 | 91 | 200 | -0.65 |
| HD89980 2 | HM86406 4 | 91 | 205 | -0.64 |
| HD89980 2 | HM86407 4 | 91 | 199 | -0.65 |
| HD89980 2 | Rat AA3 4 | 90 | 202 | -0.64 |
| HD89980 2 | RAF19090 4 | 91 | 191 | -0.67 |
| HD89980 2 | HD89980 4 | 90 | 189 | -0.67 |
| HD89980 2 | RAF11538 4 | 90 | 199 | -0.65 |
| HD89980 2 | HM95178 4 | 90 | 194 | -0.66 |
| HD89980 2 | MAF24864 1 | 93 | 215 | -0.63 |
| HD89980 2 | HM86406 1 | 92 | 216 | -0.62 |
| HD89980 2 | HM86407 1 | 92 | 204 | -0.65 |
| HD89980 2 | Rat AA3 1 | 93 | 209 | -0.64 |
| HD89980 2 | RAF19090 1 | 91 | 216 | -0.62 |
| HD89980 2 | HD89980 1 | 90 | 208 | -0.63 |
| HD89980 2 | RAF11538 1 | 91 | 214 | -0.62 |
| HD89980 2 | HM95178 1 | 92 | 210 | -0.64 |

| | | | | |
|---|---|---|---|---|
| HD89980 2 | MAF24864 2 | 25 | 75 | -0.52 |
| HD89980 2 | HM86406 2 | 25 | 79 | -0.49 |
| HD89980 2 | HM86407 2 | 29 | 84 | -0.53 |
| HD89980 2 | Rat AA3 2 | 30 | 98 | -0.48 |
| HD89980 2 | RAF19090 2 | 1 | 32 | 0.51 |
| RAF11538 2 | MAF24864 4 | 91 | 199 | -0.65 |
| RAF11538 2 | HM86406 4 | 91 | 204 | -0.64 |
| RAF11538 2 | HM86407 4 | 91 | 199 | -0.65 |
| RAF11538 2 | Rat AA3 4 | 90 | 202 | -0.64 |
| RAF11538 2 | RAF19090 4 | 90 | 198 | -0.65 |
| RAF11538 2 | HD89980 4 | 89 | 196 | -0.65 |
| RAF11538 2 | RAF11538 4 | 89 | 204 | -0.63 |
| RAF11538 2 | HM95178 4 | 89 | 206 | -0.63 |
| RAF11538 2 | MAF24864 1 | 92 | 213 | -0.63 |
| RAF11538 2 | HM86406 1 | 90 | 216 | -0.61 |
| RAF11538 2 | HM86407 1 | 90 | 209 | -0.63 |
| RAF11538 2 | Rat AA3 1 | 91 | 212 | -0.63 |
| RAF11538 2 | RAF19090 1 | 88 | 216 | -0.60 |
| RAF11538 2 | HD89980 1 | 87 | 214 | -0.60 |
| RAF11538 2 | RAF11538 1 | 89 | 215 | -0.61 |
| RAF11538 2 | HM95178 1 | 90 | 211 | -0.62 |
| RAF11538 2 | MAF24864 2 | 21 | 79 | -0.42 |
| RAF11538 2 | HM86406 2 | 21 | 76 | -0.44 |
| RAF11538 2 | HM86407 2 | 24 | 96 | -0.39 |
| RAF11538 2 | Rat AA3 2 | 24 | 100 | -0.37 |
| RAF11538 2 | RAF19090 2 | 12 | 61 | -0.29 |
| RAF11538 2 | HD89980 2 | 13 | 55 | -0.37 |
| HM95178 2 | MAF24864 4 | 91 | 193 | -0.67 |
| HM95178 2 | HM86406 4 | 91 | 199 | -0.65 |
| HM95178 2 | HM86407 4 | 91 | 193 | -0.67 |
| HM95178 2 | Rat AA3 4 | 90 | 198 | -0.65 |
| HM95178 2 | RAF19090 4 | 90 | 192 | -0.67 |
| HM95178 2 | HD89980 4 | 89 | 187 | -0.67 |
| HM95178 2 | RAF11538 4 | 89 | 197 | -0.65 |
| HM95178 2 | HM95178 4 | 89 | 197 | -0.65 |
| HM95178 2 | MAF24864 1 | 92 | 215 | -0.63 |
| HM95178 2 | HM86406 1 | 90 | 219 | -0.61 |
| HM95178 2 | HM86407 1 | 90 | 211 | -0.62 |
| HM95178 2 | Rat AA3 1 | 91 | 214 | -0.62 |
| HM95178 2 | RAF19090 1 | 88 | 213 | -0.61 |
| HM95178 2 | HD89980 1 | 87 | 214 | -0.60 |
| HM95178 2 | RAF11538 1 | 89 | 219 | -0.60 |
| HM95178 2 | HM95178 1 | 90 | 214 | -0.62 |
| HM95178 2 | MAF24864 2 | 21 | 81 | -0.41 |
| HM95178 2 | HM86406 2 | 21 | 76 | -0.44 |
| HM95178 2 | HM86407 2 | 24 | 82 | -0.46 |
| HM95178 2 | Rat AA3 2 | 24 | 93 | -0.41 |
| HM95178 2 | RAF19090 2 | 12 | 58 | -0.31 |

| | | | | |
|---|---|---|---|---|
| HM95178 2 | HD89980 2 | 13 | 52 | -0.39 |
| HM95178 2 | RAF11538 2 | 1 | 31 | 0.50 |
| MAF24864 3 | MAF24864 4 | 94 | 194 | -0.68 |
| MAF24864 3 | HM86406 4 | 94 | 196 | -0.68 |
| MAF24864 3 | HM86407 4 | 94 | 194 | -0.68 |
| MAF24864 3 | Rat AA3 4 | 93 | 201 | -0.66 |
| MAF24864 3 | RAF19090 4 | 93 | 188 | -0.69 |
| MAF24864 3 | HD89980 4 | 94 | 192 | -0.68 |
| MAF24864 3 | RAF11538 4 | 95 | 200 | -0.67 |
| MAF24864 3 | HM95178 4 | 95 | 197 | -0.68 |
| MAF24864 3 | MAF24864 1 | 97 | 212 | -0.66 |
| MAF24864 3 | HM86406 1 | 96 | 213 | -0.65 |
| MAF24864 3 | HM86407 1 | 92 | 213 | -0.63 |
| MAF24864 3 | Rat AA3 1 | 93 | 217 | -0.63 |
| MAF24864 3 | RAF19090 1 | 93 | 211 | -0.64 |
| MAF24864 3 | HD89980 1 | 91 | 209 | -0.63 |
| MAF24864 3 | RAF11538 1 | 93 | 211 | -0.64 |
| MAF24864 3 | HM95178 1 | 94 | 211 | -0.64 |
| MAF24864 3 | MAF24864 2 | 84 | 191 | -0.64 |
| MAF24864 3 | HM86406 2 | 84 | 180 | -0.66 |
| MAF24864 3 | HM86407 2 | 84 | 183 | -0.66 |
| MAF24864 3 | Rat AA3 2 | 83 | 194 | -0.63 |
| MAF24864 3 | RAF19090 2 | 82 | 177 | -0.66 |
| MAF24864 3 | HD89980 2 | 82 | 178 | -0.66 |
| MAF24864 3 | RAF11538 2 | 82 | 180 | -0.65 |
| MAF24864 3 | HM95178 2 | 82 | 172 | -0.67 |
| HM86406 3 | MAF24864 4 | 93 | 197 | -0.67 |
| HM86406 3 | HM86406 4 | 93 | 204 | -0.65 |
| HM86406 3 | HM86407 4 | 93 | 195 | -0.67 |
| HM86406 3 | Rat AA3 4 | 92 | 198 | -0.66 |
| HM86406 3 | RAF19090 4 | 92 | 193 | -0.67 |
| HM86406 3 | HD89980 4 | 93 | 197 | -0.67 |
| HM86406 3 | RAF11538 4 | 94 | 200 | -0.67 |
| HM86406 3 | HM95178 4 | 94 | 192 | -0.68 |
| HM86406 3 | MAF24864 1 | 97 | 212 | -0.66 |
| HM86406 3 | HM86406 1 | 96 | 213 | -0.65 |
| HM86406 3 | HM86407 1 | 93 | 211 | -0.64 |
| HM86406 3 | Rat AA3 1 | 94 | 216 | -0.63 |
| HM86406 3 | RAF19090 1 | 94 | 210 | -0.65 |
| HM86406 3 | HD89980 1 | 92 | 212 | -0.63 |
| HM86406 3 | RAF11538 1 | 94 | 210 | -0.65 |
| HM86406 3 | HM95178 1 | 95 | 210 | -0.65 |
| HM86406 3 | MAF24864 2 | 84 | 195 | -0.63 |
| HM86406 3 | HM86406 2 | 84 | 183 | -0.66 |
| HM86406 3 | HM86407 2 | 84 | 196 | -0.63 |
| HM86406 3 | Rat AA3 2 | 83 | 199 | -0.61 |
| HM86406 3 | RAF19090 2 | 83 | 179 | -0.66 |
| HM86406 3 | HD89980 2 | 83 | 185 | -0.65 |

| | | | | |
|---|---|---|---|---|
| HM86406 3 | RAF11538 2 | 83 | 183 | -0.65 |
| HM86406 3 | HM95178 2 | 83 | 182 | -0.65 |
| HM86406 3 | MAF24864 3 | 4 | 40 | 0.01 |
| RAF19090 3 | MAF24864 4 | 92 | 186 | -0.69 |
| RAF19090 3 | HM86406 4 | 92 | 190 | -0.68 |
| RAF19090 3 | HM86407 4 | 92 | 184 | -0.69 |
| RAF19090 3 | Rat AA3 4 | 92 | 190 | -0.68 |
| RAF19090 3 | RAF19090 4 | 91 | 186 | -0.68 |
| RAF19090 3 | HD89980 4 | 90 | 185 | -0.68 |
| RAF19090 3 | RAF11538 4 | 92 | 187 | -0.69 |
| RAF19090 3 | HM95178 4 | 92 | 181 | -0.70 |
| RAF19090 3 | MAF24864 1 | 95 | 207 | -0.66 |
| RAF19090 3 | HM86406 1 | 94 | 206 | -0.65 |
| RAF19090 3 | HM86407 1 | 95 | 204 | -0.66 |
| RAF19090 3 | Rat AA3 1 | 95 | 208 | -0.65 |
| RAF19090 3 | RAF19090 1 | 91 | 204 | -0.64 |
| RAF19090 3 | HD89980 1 | 90 | 205 | -0.64 |
| RAF19090 3 | RAF11538 1 | 92 | 201 | -0.66 |
| RAF19090 3 | HM95178 1 | 92 | 201 | -0.66 |
| RAF19090 3 | MAF24864 2 | 83 | 192 | -0.63 |
| RAF19090 3 | HM86406 2 | 83 | 185 | -0.65 |
| RAF19090 3 | HM86407 2 | 83 | 187 | -0.64 |
| RAF19090 3 | Rat AA3 2 | 83 | 194 | -0.63 |
| RAF19090 3 | RAF19090 2 | 81 | 181 | -0.65 |
| RAF19090 3 | HD89980 2 | 81 | 179 | -0.65 |
| RAF19090 3 | RAF11538 2 | 80 | 185 | -0.63 |
| RAF19090 3 | HM95178 2 | 80 | 180 | -0.64 |
| RAF19090 3 | MAF24864 3 | 39 | 74 | -0.72 |
| RAF19090 3 | HM86406 3 | 40 | 86 | -0.66 |
| HD89980 3 | MAF24864 4 | 92 | 189 | -0.68 |
| HD89980 3 | HM86406 4 | 92 | 194 | -0.67 |
| HD89980 3 | HM86407 4 | 92 | 190 | -0.68 |
| HD89980 3 | Rat AA3 4 | 92 | 194 | -0.67 |
| HD89980 3 | RAF19090 4 | 91 | 191 | -0.67 |
| HD89980 3 | HD89980 4 | 90 | 187 | -0.68 |
| HD89980 3 | RAF11538 4 | 92 | 188 | -0.68 |
| HD89980 3 | HM95178 4 | 92 | 185 | -0.69 |
| HD89980 3 | MAF24864 1 | 95 | 203 | -0.66 |
| HD89980 3 | HM86406 1 | 94 | 203 | -0.66 |
| HD89980 3 | HM86407 1 | 95 | 205 | -0.66 |
| HD89980 3 | Rat AA3 1 | 95 | 208 | -0.65 |
| HD89980 3 | RAF19090 1 | 91 | 203 | -0.65 |
| HD89980 3 | HD89980 1 | 90 | 201 | -0.65 |
| HD89980 3 | RAF11538 1 | 92 | 201 | -0.66 |
| HD89980 3 | HM95178 1 | 92 | 198 | -0.66 |
| HD89980 3 | MAF24864 2 | 83 | 196 | -0.62 |
| HD89980 3 | HM86406 2 | 83 | 188 | -0.64 |
| HD89980 3 | HM86407 2 | 83 | 186 | -0.64 |

| | | | | |
|---|---|---|---|---|
| HD89980 3 | Rat AA3 2 | 83 | 192 | -0.63 |
| HD89980 3 | RAF19090 2 | 81 | 184 | -0.64 |
| HD89980 3 | HD89980 2 | 81 | 184 | -0.64 |
| HD89980 3 | RAF11538 2 | 80 | 187 | -0.63 |
| HD89980 3 | HM95178 2 | 80 | 181 | -0.64 |
| HD89980 3 | MAF24864 3 | 38 | 82 | -0.66 |
| HD89980 3 | HM86406 3 | 39 | 90 | -0.63 |
| HD89980 3 | RAF19090 3 | 1 | 25 | 0.40 |
| RAF11538 3 | MAF24864 4 | 92 | 198 | -0.66 |
| RAF11538 3 | HM86406 4 | 92 | 200 | -0.66 |
| RAF11538 3 | HM86407 4 | 94 | 199 | -0.67 |
| RAF11538 3 | Rat AA3 4 | 93 | 197 | -0.67 |
| RAF11538 3 | RAF19090 4 | 93 | 197 | -0.67 |
| RAF11538 3 | HD89980 4 | 92 | 196 | -0.67 |
| RAF11538 3 | RAF11538 4 | 93 | 194 | -0.68 |
| RAF11538 3 | HM95178 4 | 93 | 195 | -0.67 |
| RAF11538 3 | MAF24864 1 | 96 | 219 | -0.64 |
| RAF11538 3 | HM86406 1 | 95 | 219 | -0.63 |
| RAF11538 3 | HM86407 1 | 95 | 212 | -0.65 |
| RAF11538 3 | Rat AA3 1 | 96 | 211 | -0.65 |
| RAF11538 3 | RAF19090 1 | 92 | 214 | -0.63 |
| RAF11538 3 | HD89980 1 | 90 | 215 | -0.62 |
| RAF11538 3 | RAF11538 1 | 93 | 214 | -0.63 |
| RAF11538 3 | HM95178 1 | 94 | 209 | -0.65 |
| RAF11538 3 | MAF24864 2 | 82 | 194 | -0.62 |
| RAF11538 3 | HM86406 2 | 82 | 184 | -0.64 |
| RAF11538 3 | HM86407 2 | 84 | 187 | -0.65 |
| RAF11538 3 | Rat AA3 2 | 84 | 190 | -0.64 |
| RAF11538 3 | RAF19090 2 | 81 | 178 | -0.65 |
| RAF11538 3 | HD89980 2 | 81 | 186 | -0.63 |
| RAF11538 3 | RAF11538 2 | 81 | 189 | -0.63 |
| RAF11538 3 | HM95178 2 | 81 | 184 | -0.64 |
| RAF11538 3 | MAF24864 3 | 35 | 80 | -0.64 |
| RAF11538 3 | HM86406 3 | 36 | 84 | -0.63 |
| RAF11538 3 | RAF19090 3 | 34 | 67 | -0.70 |
| RAF11538 3 | HD89980 3 | 33 | 73 | -0.65 |
| HM95178 3 | MAF24864 4 | 92 | 194 | -0.67 |
| HM95178 3 | HM86406 4 | 92 | 196 | -0.67 |
| HM95178 3 | HM86407 4 | 94 | 193 | -0.68 |
| HM95178 3 | Rat AA3 4 | 93 | 194 | -0.68 |
| HM95178 3 | RAF19090 4 | 93 | 196 | -0.67 |
| HM95178 3 | HD89980 4 | 92 | 193 | -0.67 |
| HM95178 3 | RAF11538 4 | 93 | 202 | -0.66 |
| HM95178 3 | HM95178 4 | 93 | 190 | -0.68 |
| HM95178 3 | MAF24864 1 | 96 | 212 | -0.65 |
| HM95178 3 | HM86406 1 | 95 | 215 | -0.64 |
| HM95178 3 | HM86407 1 | 95 | 209 | -0.65 |
| HM95178 3 | Rat AA3 1 | 96 | 212 | -0.65 |

| | | | | |
|---|---|---|---|---|
| HM95178 3 | RAF19090 1 | 92 | 212 | -0.63 |
| HM95178 3 | HD89980 1 | 90 | 211 | -0.62 |
| HM95178 3 | RAF11538 1 | 93 | 211 | -0.64 |
| HM95178 3 | HM95178 1 | 94 | 210 | -0.65 |
| HM95178 3 | MAF24864 2 | 82 | 192 | -0.63 |
| HM95178 3 | HM86406 2 | 82 | 181 | -0.65 |
| HM95178 3 | HM86407 2 | 84 | 188 | -0.64 |
| HM95178 3 | Rat AA3 2 | 84 | 193 | -0.63 |
| HM95178 3 | RAF19090 2 | 81 | 175 | -0.66 |
| HM95178 3 | HD89980 2 | 81 | 178 | -0.65 |
| HM95178 3 | RAF11538 2 | 81 | 184 | -0.64 |
| HM95178 3 | HM95178 2 | 81 | 180 | -0.65 |
| HM95178 3 | MAF24864 3 | 35 | 74 | -0.67 |
| HM95178 3 | HM86406 3 | 36 | 77 | -0.66 |
| HM95178 3 | RAF19090 3 | 34 | 62 | -0.73 |
| HM95178 3 | HD89980 3 | 33 | 71 | -0.66 |
| HM95178 3 | RAF11538 3 | 1 | 26 | 0.42 |
| HM86407 3 | MAF24864 4 | 92 | 182 | -0.70 |
| HM86407 3 | HM86406 4 | 92 | 193 | -0.67 |
| HM86407 3 | HM86407 4 | 93 | 182 | -0.70 |
| HM86407 3 | Rat AA3 4 | 92 | 180 | -0.70 |
| HM86407 3 | RAF19090 4 | 93 | 189 | -0.69 |
| HM86407 3 | HD89980 4 | 92 | 190 | -0.68 |
| HM86407 3 | RAF11538 4 | 92 | 200 | -0.66 |
| HM86407 3 | HM95178 4 | 92 | 191 | -0.68 |
| HM86407 3 | MAF24864 1 | 97 | 209 | -0.66 |
| HM86407 3 | HM86406 1 | 96 | 205 | -0.67 |
| HM86407 3 | HM86407 1 | 94 | 208 | -0.65 |
| HM86407 3 | Rat AA3 1 | 95 | 215 | -0.64 |
| HM86407 3 | RAF19090 1 | 94 | 207 | -0.65 |
| HM86407 3 | HD89980 1 | 94 | 210 | -0.65 |
| HM86407 3 | RAF11538 1 | 95 | 209 | -0.65 |
| HM86407 3 | HM95178 1 | 96 | 210 | -0.65 |
| HM86407 3 | MAF24864 2 | 83 | 187 | -0.64 |
| HM86407 3 | HM86406 2 | 83 | 176 | -0.67 |
| HM86407 3 | HM86407 2 | 83 | 192 | -0.63 |
| HM86407 3 | Rat AA3 2 | 82 | 196 | -0.62 |
| HM86407 3 | RAF19090 2 | 83 | 181 | -0.66 |
| HM86407 3 | HD89980 2 | 83 | 182 | -0.65 |
| HM86407 3 | RAF11538 2 | 82 | 176 | -0.66 |
| HM86407 3 | HM95178 2 | 82 | 175 | -0.67 |
| HM86407 3 | MAF24864 3 | 42 | 85 | -0.69 |
| HM86407 3 | HM86406 3 | 40 | 93 | -0.63 |
| HM86407 3 | RAF19090 3 | 51 | 97 | -0.72 |
| HM86407 3 | HD89980 3 | 50 | 106 | -0.67 |
| HM86407 3 | RAF11538 3 | 40 | 99 | -0.60 |
| HM86407 3 | HM95178 3 | 40 | 94 | -0.62 |
| Rat AA3 3 | MAF24864 4 | 93 | 195 | -0.67 |

| | | | | |
|---|---|---|---|---|
| Rat AA3 3 | HM86406 4 | 93 | 200 | -0.66 |
| Rat AA3 3 | HM86407 4 | 94 | 191 | -0.69 |
| Rat AA3 3 | Rat AA3 4 | 93 | 194 | -0.68 |
| Rat AA3 3 | RAF19090 4 | 94 | 198 | -0.67 |
| Rat AA3 3 | HD89980 4 | 93 | 200 | -0.66 |
| Rat AA3 3 | RAF11538 4 | 93 | 201 | -0.66 |
| Rat AA3 3 | HM95178 4 | 93 | 195 | -0.67 |
| Rat AA3 3 | MAF24864 1 | 96 | 211 | -0.65 |
| Rat AA3 3 | HM86406 1 | 95 | 208 | -0.65 |
| Rat AA3 3 | HM86407 1 | 94 | 217 | -0.63 |
| Rat AA3 3 | Rat AA3 1 | 95 | 218 | -0.63 |
| Rat AA3 3 | RAF19090 1 | 92 | 213 | -0.63 |
| Rat AA3 3 | HD89980 1 | 92 | 217 | -0.62 |
| Rat AA3 3 | RAF11538 1 | 93 | 211 | -0.64 |
| Rat AA3 3 | HM95178 1 | 94 | 210 | -0.65 |
| Rat AA3 3 | MAF24864 2 | 82 | 195 | -0.62 |
| Rat AA3 3 | HM86406 2 | 82 | 187 | -0.64 |
| Rat AA3 3 | HM86407 2 | 82 | 197 | -0.61 |
| Rat AA3 3 | Rat AA3 2 | 81 | 200 | -0.60 |
| Rat AA3 3 | RAF19090 2 | 82 | 185 | -0.64 |
| Rat AA3 3 | HD89980 2 | 82 | 189 | -0.63 |
| Rat AA3 3 | RAF11538 2 | 81 | 185 | -0.64 |
| Rat AA3 3 | HM95178 2 | 81 | 189 | -0.63 |
| Rat AA3 3 | MAF24864 3 | 39 | 90 | -0.63 |
| Rat AA3 3 | HM86406 3 | 41 | 95 | -0.63 |
| Rat AA3 3 | RAF19090 3 | 52 | 103 | -0.70 |
| Rat AA3 3 | HD89980 3 | 51 | 109 | -0.66 |
| Rat AA3 3 | RAF11538 3 | 40 | 112 | -0.55 |
| Rat AA3 3 | HM95178 3 | 40 | 104 | -0.58 |
| Rat AA3 3 | HM86407 3 | 8 | 47 | -0.23 |
| | | | Mean: | -0.59 |
| | | Standard Deviation: | | 0.18 |

Extreme ubiquitin data for 915 nucleotides and 305 amino acids
First letter in accession number indicates species: H=human, R=rat and M=mouse

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| HM26880 | RD17296 | 1 | 101 | 1.01 |
| RD16554 | RD17296 | 1 | 90 | 0.96 |
| RD16554 | HM26880 | 1 | 121 | 1.09 |
| MX51703 | RD17296 | 1 | 84 | 0.93 |
| MX51703 | HM26880 | 1 | 114 | 1.06 |
| MX51703 | RD16554 | 1 | 25 | 0.40 |
| | | | Mean: | 1.01 |
| | | Standard Deviation: | | 0.08 |

Intra-specific (Corn) ubiquitin data for 76 amino acids and, 228 nucleotides. The last number in the accession # indicates the polyubiquitin monomer.

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| U291621 | U291581 | 1 | 27 | 0.43 |
| U291592 | U291581 | 1 | 20 | 0.30 |
| U291592 | U291621 | 1 | 21 | 0.33 |
| U291593 | U291581 | 1 | 23 | 0.36 |
| U291593 | U291621 | 1 | 22 | 0.35 |
| U291593 | U291592 | 1 | 15 | 0.18 |
| U291594 | U291581 | 1 | 27 | 0.43 |
| U291594 | U291621 | 1 | 26 | 0.42 |
| U291594 | U291592 | 1 | 21 | 0.33 |
| U291594 | U291593 | 1 | 23 | 0.36 |
| U291595 | U291581 | 1 | 31 | 0.49 |
| U291595 | U291621 | 1 | 28 | 0.45 |
| U291595 | U291592 | 1 | 31 | 0.49 |
| U291595 | U291593 | 1 | 30 | 0.48 |
| U291595 | U291594 | 1 | 20 | 0.30 |
| U291596 | U291581 | 1 | 38 | 0.58 |
| U291596 | U291621 | 1 | 38 | 0.58 |
| U291596 | U291592 | 1 | 32 | 0.51 |
| U291596 | U291593 | 1 | 30 | 0.48 |
| U291596 | U291594 | 1 | 28 | 0.45 |
| U291596 | U291595 | 1 | 29 | 0.47 |
| U291597 | U291581 | 1 | 38 | 0.58 |
| U291597 | U291621 | 1 | 40 | 0.60 |
| U291597 | U291592 | 1 | 40 | 0.60 |
| U291597 | U291593 | 1 | 33 | 0.52 |
| U291597 | U291594 | 1 | 26 | 0.42 |
| U291597 | U291595 | 1 | 33 | 0.52 |
| U291597 | U291596 | 1 | 26 | 0.42 |
| U291582 | U291581 | 1 | 18 | 0.26 |
| U291582 | U291621 | 1 | 23 | 0.36 |
| U291582 | U291592 | 1 | 21 | 0.33 |
| U291582 | U291593 | 1 | 20 | 0.30 |
| U291582 | U291594 | 1 | 26 | 0.42 |
| U291582 | U291595 | 1 | 25 | 0.40 |
| U291582 | U291596 | 1 | 35 | 0.55 |
| U291582 | U291597 | 1 | 37 | 0.57 |
| U291583 | U291581 | 1 | 29 | 0.47 |
| U291583 | U291621 | 1 | 37 | 0.57 |
| U291583 | U291592 | 1 | 32 | 0.51 |
| U291583 | U291593 | 1 | 33 | 0.52 |
| U291583 | U291594 | 1 | 24 | 0.38 |
| U291583 | U291595 | 1 | 21 | 0.33 |
| U291583 | U291596 | 1 | 26 | 0.42 |

| | | | | |
|---|---|---|---|---|
| U291583 | U291597 | 1 | 31 | 0.49 |
| U291583 | U291582 | 1 | 31 | 0.49 |
| U291584 | U291581 | 1 | 37 | 0.57 |
| U291584 | U291621 | 1 | 37 | 0.57 |
| U291584 | U291592 | 1 | 32 | 0.51 |
| U291584 | U291593 | 1 | 33 | 0.52 |
| U291584 | U291594 | 1 | 27 | 0.43 |
| U291584 | U291595 | 1 | 30 | 0.48 |
| U291584 | U291596 | 1 | 22 | 0.35 |
| U291584 | U291597 | 1 | 16 | 0.21 |
| U291584 | U291582 | 1 | 35 | 0.55 |
| U291584 | U291583 | 1 | 29 | 0.47 |
| U291622 | U291581 | 1 | 32 | 0.51 |
| U291622 | U291621 | 1 | 27 | 0.43 |
| U291622 | U291592 | 1 | 27 | 0.43 |
| U291622 | U291593 | 1 | 29 | 0.47 |
| U291622 | U291594 | 1 | 24 | 0.38 |
| U291622 | U291595 | 1 | 27 | 0.43 |
| U291622 | U291596 | 1 | 32 | 0.51 |
| U291622 | U291597 | 1 | 33 | 0.52 |
| U291622 | U291582 | 1 | 25 | 0.40 |
| U291622 | U291583 | 1 | 30 | 0.48 |
| U291622 | U291584 | 1 | 30 | 0.48 |
| U291623 | U291581 | 1 | 30 | 0.48 |
| U291623 | U291621 | 1 | 25 | 0.40 |
| U291623 | U291592 | 1 | 25 | 0.40 |
| U291623 | U291593 | 1 | 29 | 0.47 |
| U291623 | U291594 | 1 | 18 | 0.26 |
| U291623 | U291595 | 1 | 20 | 0.30 |
| U291623 | U291596 | 1 | 31 | 0.49 |
| U291623 | U291597 | 1 | 34 | 0.53 |
| U291623 | U291582 | 1 | 22 | 0.35 |
| U291623 | U291583 | 1 | 29 | 0.47 |
| U291623 | U291584 | 1 | 31 | 0.49 |
| U291623 | U291622 | 1 | 25 | 0.40 |
| U291624 | U291581 | 1 | 36 | 0.56 |
| U291624 | U291621 | 1 | 22 | 0.35 |
| U291624 | U291592 | 1 | 30 | 0.48 |
| U291624 | U291593 | 1 | 30 | 0.48 |
| U291624 | U291594 | 1 | 35 | 0.55 |
| U291624 | U291595 | 1 | 33 | 0.52 |
| U291624 | U291596 | 1 | 30 | 0.48 |
| U291624 | U291597 | 1 | 36 | 0.56 |
| U291624 | U291582 | 1 | 27 | 0.43 |
| U291624 | U291583 | 1 | 39 | 0.59 |
| U291624 | U291584 | 1 | 28 | 0.45 |
| U291624 | U291622 | 1 | 31 | 0.49 |
| U291624 | U291623 | 1 | 29 | 0.47 |

| | | | | |
|---|---|---|---|---|
| U291625 | U291581 | 1 | 35 | 0.55 |
| U291625 | U291621 | 1 | 35 | 0.55 |
| U291625 | U291592 | 1 | 32 | 0.51 |
| U291625 | U291593 | 1 | 32 | 0.51 |
| U291625 | U291594 | 1 | 27 | 0.43 |
| U291625 | U291595 | 1 | 35 | 0.55 |
| U291625 | U291596 | 1 | 23 | 0.36 |
| U291625 | U291597 | 1 | 18 | 0.26 |
| U291625 | U291582 | 1 | 36 | 0.56 |
| U291625 | U291583 | 1 | 31 | 0.49 |
| U291625 | U291584 | 1 | 13 | 0.12 |
| U291625 | U291622 | 1 | 29 | 0.47 |
| U291625 | U291623 | 1 | 34 | 0.53 |
| U291625 | U291624 | 1 | 31 | 0.49 |
| | | | Mean: | 0.45 |
| | | Standard Deviation: | | 0.10 |

Inter-specific (Corn and Chicken) ubiquitin data for 71 amino acids and 228 nucleotides. The last character in the accession # indicates the polyubiquitin monomer.

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| U291583 | U291582 | 31 | 1 | 0.64 |
| M11100B | U291582 | 40 | 3 | 0.81 |
| M11100B | U291583 | 43 | 3 | 0.89 |
| M11100C | U291582 | 35 | 3 | 0.69 |
| M11100C | U291583 | 41 | 3 | 0.84 |
| M11100C | M11100B | 19 | 1 | 0.32 |
| | | | Mean: | 0.70 |
| | | Standard Deviation: | | 0.21 |

Human and chimpanzee comparison of HLA-A, B and C antigens from the MHC-complex. 364 amino acids, 1094 nucleotides

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| AF115460 | L19923 | 51 | 26 | -0.71 |
| D83031 | L19923 | 78 | 45 | -0.76 |
| D83031 | AF115460 | 86 | 43 | -0.70 |
| AF115464 | L19923 | 81 | 44 | -0.73 |
| AF115464 | AF115460 | 91 | 46 | -0.70 |
| AF115464 | D83031 | 40 | 26 | -0.81 |
| L77702 | L19923 | 97 | 50 | -0.71 |

| | | | | |
|---|---|---|---|---|
| L77702 | AF115460 | 98 | 48 | -0.69 |
| L77702 | D83031 | 119 | 55 | -0.66 |
| L77702 | AF115464 | 115 | 60 | -0.72 |
| AF115463 | L19923 | 103 | 51 | -0.69 |
| AF115463 | AF115460 | 101 | 48 | -0.67 |
| AF115463 | D83031 | 123 | 58 | -0.67 |
| AF115463 | AF115464 | 116 | 57 | -0.69 |
| AF115463 | L77702 | 34 | 20 | -0.77 |
| | | | Mean: | -0.71 |
| | | Standard Deviation: | | 0.04 |

Inter-specific human comparison of HLA-A, B and C antigens from the MHC-complex. 364 amino acids, 1094 nucleotides

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| D83031 | L19923 | 78 | 45 | -0.759033849 |
| L77702 | L19923 | 97 | 50 | -0.710114208 |
| L77702 | D83031 | 119 | 55 | -0.662731666 |
| | | | Mean: | -0.71 |
| | | Standard Deviation: | | 0.05 |

Inter-specific chimpanzee comparison of HLA-A, B and C antigens from the MHC-complex. 364 amino acids, 1094 nucleotides

| Pair-wise Comparison | | Amino | Nucs. | HI |
|---|---|---|---|---|
| AF115464 | AF115460 | 91 | 46 | -0.701632378 |
| AF115463 | AF115460 | 101 | 48 | -0.674835802 |
| AF115463 | AF115464 | 116 | 57 | -0.689332805 |
| | | | Mean: | -0.69 |
| | | Standard Deviation: | | 0.01 |

## REFERENCES

Addis M. F. 1998. Cloning and molecular characterization of a cDNA clone coding for *Trichomonas vaginalis* alpha-actinin and intracellular localization of the protein. Infect. Immun. **66**(10):4924-4931.

Arimura,C., T. Suzuki, M. Yanagisawa, M. Imamura, Y. Hamada and T Masaki. 1988 Primary structure of chicken skeletal muscle and fibroblast alpha-actinins deduced from cDNA sequences. Eur. J. Biochem. **177**(3):649-655.

Baron, M. D., M. D. Davison, P. Jones and D.R. Critchley. 1987. The sequence of chick α-actinin reveals homologies to spectrin and calmodulin. J. Biol. Chem. **262**(36):17623-17629

Barstead, R. J., L. Kleiman and R. H. Waterston. 1991. Cloning, sequencing, and mapping of an α-actinin gene from the nematode *Caenorhabitis elegans*. Cell Motility and Cytoskeleton **20**:69-78.

Beggs, A. H., T. J. Byers, J.H.M. Knoll, F. M .Boyce, G. A. P. Bruns and L. M. Kunkel 1992. Cloning and characterization of two human skeletal muscle α-actinin genes located on chromosomes 1 and 11. J. Biol. Chem. **267**(13):9281-9288.

Blanchard, A., V. Ohanian and D. Critchley. 1989. The structure and function of α-actinin. J. Mus. Res. and Cell Motil. **10**:280-289.

Bolanos, Sandra H., D. O. Zamora, D. M. García and J. R. Koke. 1998. An α-actinin isoform which may cross-link intermediate filaments and microfilaments. Cytobios **94**:39-61.

Bond, U. and M. J. Schlesinger. 1985. Ubiquitin is a heat shock protein in chicken embryo fibroblasts. Mol. Cell. Biol. **5**:949-956.

Bricheux G., G. Coffe, N. Pradel and G. Brugerolle. 1998. Evidence for an uncommon alpha-actinin protein in *Trichomonas vaginalis*. Mol. Biochem. Parasitol. **95**(2): 241-249.

Burridge, K. and J. Feramisco. 1981. Non-muscle α-actinins are calcium-sensitive actin-binding proteins. Nature **294**:565-567.

Chan Y. 1998. Human skeletal muscle-specific alpha-actinin-2 and -3 isoforms form homodimers and heterodimers *in vitro* and *in vivo*. Biochem. Biophys. Res. Commun. **248** (1):134-139.

Cowell, I. G. and C. A. Austin. 1997 cDNA Library Protocols. Humana Press Totowa, New Jersey.

Dear, T. N., N. T. Meier, M. Hunn and T. Boehm, 2000. Gene structure, chromosomal localization, and expression pattern of capn12, a new member of the calpain large subunit gene family. Genomics **68**(2):152-160.

Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. Nature **299** 111-117.

Dubreil, R. R. 1991. Structure and evolution of the actin cross-linking proteins. BioEssays **13**(5):219-226.

El-Husseini A. E., D. Kwasnicka, T Yamada, S. Hirohashi and S. R. Vincent. 2000. BERP, a novel ring finger protein, binds to alpha-actinin-4. Biochem Biophys Res Commun. **267**(3):906-911.

Endo, T. and T. Masaki. 1984. Differential expression and distribution of chicken skeletal- and smooth-muscle-type α-actinins during myogenesis in culture. J. Cell Biol. **99**:2322-2332.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. *Distributed by the author. Department of Genetics, University of Washington, Seattle.*

Finch, J. S., K. Bonham, P. Krieg and G. T. Bowden. 1990. Murine polyubiquitin mRNA sequence. Nucleic Acids Res. **18**(7):1907.

Fyrberg, E., M. Kelly, E. Ball, C. Fyrberg and M. C. Reedy. 1990. Molecular genetics of Drosophila alpha-actinin: mutant alleles disrupt Z disc integrity and muscle insertions. J. Cell Biol. **110**:1999-2011.

Geisert, E. E., T. P. Murphy, M. H. Irwin, and H. Larhava. 1996. A novel cell adhesion molecule G-CAM, found on cultured rat glia. Neurosci. Letters **133**:262-266.

Gojobori, T. and M. Nei. 1984. Concerted evolution and the immunoglobulin $v_h$ gene family. Mol. Biol. Evol. **1**(2):195-212.

Graur, D. and W-H. Li. 1999. Fundamentals of Molecular Evolution. Sinauer, Sunderland, MA.

Hammonds, R. G. 1987. Protein sequence of DMD gene is related to actin-binding domain of α-actinin. Cell **51**(1):321.

Hasegawa, M., H. Kishino and K. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol. Evol. **22**:160-174.

Högstrand, K. and J. Böhme. 1997. Gene conversion of major histocompatibility complex genes in the mouse spermatogenesis is a premeiotic event. Mol. Biol. of the Cell **8**:2511-2517.

Honda, K., T. Yamada, R. Endo, Y. Ino, M. Gotoh, H. Tsuda, Y Yamada, H. Chiba and S. Hirohashi. 1998. Actinin 4, a novel actin-bundling protein associated with cell motility and cancer invasion. J. Cell Biol. **140**(6):1383-1393.

Hughes, A. L. 1994. Concerted evolution of exons and introns in the MHC-linked Tenascin-X gene of mammals. Mol. Biol. Evol. **16**(11):1558-1567.

Hughes, A. L. 1999. Adaptive evolution of genes and genomes. Oxford University Press, New York.

Hyashi, T., M. Noga and M. Matsuda. 1994. Nucleotide sequence and expression of the rat polyubiquitin mRNA. Biophys. Acta **1218**(2):232-234.

Imamura, M., T. Sakurai, Y. Ogawa, T. Ihikawa, K. Goto and T. Masaki. 1994. Molecular cloning of low-$Ca^{2+}$-sensitive-type non-muscle α-actinin. Eur. J. Biochem. **223**:395-401.

Jackson, M. S., T. Strachan and G. Dover. 1996. Human Genome Evolution. BIOS Scientific Publishers, Oxford, UK.

Jockusch, B. M., H. Veldman, G. W. Griffiths, B. A. Van Oost and F. G. I. Jennekens 1980. Immunofluorescence microscopy of a myopathy: α-actinin is a major constituent of nemaline rods. Experimental Cell Research **127**:409-420

Kitano, T., S. Oota and N. Saitou. 1999. Molecular evolutionary analyses of the rh blood group genes and rh50 genes in mammals. Zoological Studies **38**(4):379-386.

Klein, J., C. O'hUigin, F. Figueroa, W. E. Mayer and D. Klein. 1993. Different modes of MHC evolution in primates. Mol. Biol. Evol. **10**(1):48-59.

Kumar, S. and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917-918.

Lienert, K., G. Russ, S. Lester, G. Bennett, X. Gao and J. McCluskey. 1996. Stable inheritance of an HLA-_ phenotype associated with a structural mutation in HLA-A*0301 gene. Tissue Antigens **48**(3):187-191.

Liu, L., D. S. Maillet, J. R. Frappier, D. B. Walden and B. G. Atkinson. 1995. Characterization, chromosomal mapping and expression of different polyubiquitin

genes in tissues from control and heat-shocked maize seedlings. Biochem. Cell Biol. **73**(1):19-30.

Malhotra, S. K., F. Wong, N. Sinha, V. Manickeval, T. K. Shnitka and K. G. Warren. 1984. A monoclonal antibody for astrocytes. Ann. N. Y. Academy of Science **436**:492-494.

Maniatis, T., E. F. Fritsch and J. Sambrook. 1989. Molecular Cloning. A Laboratory Manual. Cold Spring Harbor, NY.

Matsui, M., S. Machida, S. M. Feinstone and T Akatsuka. 1999. Molecular analyses of five new chimpanzee MHC class I alleles: implications for differences between evolutional mechanisms of HLA-A, B and C loci. Biochem. Biophys. Res. Comm. **261**(1)46-52.

Muse, S. V., A. G. Clark and G.H. Thomas. 1997. Comparisons of the nucleotide substitution process among repetitive segments of the α- and β-spectrin genes. J. Mol. Evol. **44**:492-500.

Nei, M., X. Gu and T. Sitnikova. 1997. Evolution by the birth-and-death process in the multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci **94**:7799-7806.

Nei, M., I. B. Rogozin and H. Piontkivska. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc. Natl. Acad. Sci. **97**(20):10866-10871.

Nenoi, M., K. Mita, S. Ichimura and A.Kawano 1998. Higher frequency of concerted evolutionary events in rodents than in man at the polyubiquitin gene VNTR locus. Genetics. **148**:867-876.

Nikolopoulos, S. N., B. A. Spengler, K. Kisselbach, A.E. Evans, J.L. Bielder and R A. Ross. 2000. The human non-muscle α-actinin protein encoded by the ACTN4 gene suppresses tumorigenicity of human neuroblastoma cells. Oncogene **19**:380-386.

Noegel, A. 1987. Calcium-sensitive non-muscle α-actinin contains EF-hand structures and highly conserved regions. FEBS. **221**(2):391-396.

North, K. N., N. Yang, D. Wattanasirichaigoon, M. Mills, S. Easteal and A. H. Beggs. 1999. A common nonsense mutation results in α-actinin-3 deficiency in the general population. Nature Genetics **21**:353-354.

Ohta, T. 1983. On the evolution of multigene families. Theor. Popul. Biol. 23:216-240.

Ota, T. and M. Nei, 1994. Divergent evolution by the birth-and death process in the immunoglobulin $v_h$ gene family. Mol. Biol. Evol. **11**(3):469-482.

Parr, T., G. T. Waites, B. Patel, D. B. Millake and D. R. Critchley. 1992. A chick skeletal muscle α-actinin gene gives rise to two alternatively spliced isoforms which differ in the EF-hand $Ca^{2+}$-binding domain. Eur. J. Biochem. **210**:801-809.

Price, L. A., S. K. Malhotra and J. R. Koke. 1993. Localization and characterization of an intermediate filament-associated protein. Cytobios **76**:157-173.

S. A. Sawyer. (1999) GENECONV: A computer package for the statistical detection of gene conversion. *Distributed by the author, Department of Mathematics, Washington University in St. Louis, available at* http://www.math.wustl.edu/~sawyer.

Sharp, P.M. and W. H. Li. 1987. Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. Mol. Evol. **25**:58-64.

Strimmer, K., and A. Haesler. 1996. Quartet puzzling A quartet maximum-likelihood method for reconstructing tree topologies Mol. Biol. Evol. **13**(7):964-969.

Swofford, D. L. 1999. PAUP, Phylogenetic Analysis Using Parsimony (and other methods) version 4.0. Sinauer, Sunderland, MA.

Tan, Y., S. T. Bishoff and M. A. Riley. 1993. Ubiquitins revisited: further examples of within- and between-locus concerted evolution. Mol. Phylogenet. Evol. **2**:351-360.

Thomas, G. H., E. C. Newbern, C. C. Korte, M. A. Bales, S V. Muse, A. G. Clark and D P. Kiehart. 1997. Intragenic duplication and divergence in the spectrin superfamily of proteins. Mol. Biol. Evol. **14**(12):1285-1295.

Vrana, P.B. and W. C. Wheeler. 1996. Molecular evolution and phylogenetic utility of the polyubiquitin locus in mammals and higher vertebrates. Mol. Phylogenet Evol. **6**:259-269.

Wiborg, O. M. S. Pedersen, A. Wind, L. E. Berglund, K. A. Marcker and J. Vuust. 1985. The human ubiquitin multigene family: Some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J. **4**:755-759.

Wittzell, H., A. Bernot, C. Auffray and R. Zoorob. 1999. Concerted evolution of two MHC class II loci in pheasants and domestic chickens. Mol. Biol. Evol. **16**(4):479-490.

Youssoufian, H., M. McAfee and D. J. Kwiatkowski. 1990. Cloning and chromosomal localization of the human cytoskeletal alpha-actinin gene revveals linkage to the beta-spectrin gene. Am. J. Human Genetics **47**(1):62-71.

Zimmer, E. A., S. L. Martin, S. M. Beverley, Y. M Kan and A. C. Wilson. 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin Proc. Natl Acad. Sci. 77:2158-2162.