Isolation, Characterisation and Molecular Evolution of the Actin Gene Family of the New Zealand Black-Footed Abalone, *Haliotis iris*

A thesis submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy in
Cellular and Molecular Biology

by

Maxine Joyce Bryant

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

School of Biological Sciences

2004

For my parents,
David Kempthorne & Ivy Te Paea

TABLE OF CONTENTS

ABSTRACT

Molluscs are biologically important; they form a diverse taxon and the most well known members of the Lophotrochozoa, the most understudied Bilaterian group. Despite this importance, few studies have characterised molluscs at the genetic level. The New Zealand black-footed abalone *Haliotis iris*, an economically and culturally valued species, was chosen as a model for genetic characterisation of a molluscan actin gene family. In *H. iris*, actin is essential for the production of a large muscular foot, which forms the bulk of the body mass. The structure, expression and evolution of actin genes were investigated to elucidate the function of the actin gene family in *H. iris*.

H. iris actin genes were isolated by PCR using gene subtype-specific primers designed from previously characterised partial *H. iris* actin sequences and generic primers derived from *H. rufescens* (Californian red abalone) and *Cyprinus carpio* (common carp). Three full length genes, *H.irisA1*, *H.irisA2* and *H.irisA3*, and three partial genes, *H.irisA1a*, *H.irisA1b* and *H.irisA1c*, were isolated. The full length genes showed 82-95% sequence similarity to mollusc actin gene sequences deposited in GenBank. Sequence conservation confirmed the identity of the putative actin genes. The six genes contained a single variable length intron between codons 41 and 42. Intron lengths were: 174 nt, *H.irisA1*; 1,078 nt, *H.irisA2*; 581 nt, *H.irisA3*; 301 nt, *H.irisA1a*; 282 nt, *H.irisA1* and 229 nt, *H.irisA1c*. The predicted proteins of the full length genes contained 375 aa and lacked the second amino acid usually found in invertebrate actin proteins. Southern hybridisation of genomic DNA suggested there was a large gene family composed of at least eight members.

The expression of *H.irisA1*, *H.irisA2* and *H.irisA3* in developmental stages and adult tissues was investigated by RT-PCR. RT-PCR demonstrated differential expression of *H. iris* actin genes during development and in adult tissues. *H.irisA1* and *H.irisA2* were expressed at low levels in fertilised eggs and

blastula, with expression increasing in trochophore and veliger larvae. *H.irisA3* was not expressed in eggs, but was faintly detected in blastula and highly expressed in trochophore and veliger larvae. *H.irisA1* was ubiquitously expressed in adult gill, gonad, hepatopancreas, foot and mantle tissue, suggesting it may be a cytoplasmic-type actin. *H.irisA2* was expressed in all tissues except the hepatopancreas, although low expression may not have been detectable by electrophoresis of RT-PCR products. Further characterisation is required to confirm whether *H.irisA2* encodes a cytoplasmic-type actin. *H.irisA3* was expressed at high levels in the muscular foot and mantle, and was faintly detected in gonad, suggesting it may be a muscle type actin.

Phylogenetic analyses of *H. iris* actin genes and other molluscan actin genes available on GenBank were performed using maximum parsimony and maximum likelihood methods. Analyses suggested that haliotid actins can be divided into two orthologous clades, the first clade containing *H.irisA1, H.irisA1a, H.irisA1b, H.irisA1c, H.virgA1a, H.virgA1b, H.virgA1c* and *H. rufescens* actin, the second clade containing *H.irisA2, H.irisA3* and *H. discus hannai actin.* Orthology indicated that the last common ancestor of haliotids had at least two actin genes. Clustering of actin genes from individual haliotid species within orthologous actin gene clades suggests paralogy resulting from duplication of actin genes within species. Evidence for gene orthology between mollusc actin genes was found, but further characterisation of actin genes from other mollusc species is required to infer the evolutionary significance of orthology.

ABBREVIATIONS

α	shape parameter value (Gamma distribution)
A	adenine
aa	amino acid(s)
α^{32}P-dCTP	deoxycytdine 5'-[α-^{32}P]-triphosphate
APS	ammonium persulfate
A+T	adenine and thymine content
BC-KA P-value	Bonferroni-corrected Karlin Altshcul P-value
C	cytosine
cDNA	complementary DNA
Ci	Curie
ddNTPs	dideoxynucleotides
d_N	number of nonsynonymous substitutions/nonsynonymous site
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotides
d_S	number of synonymous substitutions/ synonymous site
EtOH	ethanol
fmoles	femtomoles
g	gravitational force
G	guanine
G+C	guanine and cytosine content
γ^{33}P-dATP	deoxyadenosine 5'-[γ-^{33}P]-triphosphate
h	hour(s)
HCl	hydrochloric acid
KCl	potassium chloride
Kb	kilobase(s)
M	molar
$Mg(OAc)_2$	magnesium acetate
$MgCl_2$	magnesium chloride
min	minute(s)
ml	millilitre(s)
ML	maximum likelihood
mM	millimolar
MP	maximum parsimony
mRNA	messenger RNA
mw	molecular weight
NaCl	sodium chloride
Na_2EDTA	disodium ethylenediaminetetraacetic acid
NaI	sodium iodide
NaOAc	sodium acetate
NH_4OAc	ammonium acetate
ng	nanogram(s)
nt	nucleotide(s)

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pmoles	picomoles
PNK	T4 polynucleotide kinase
R	transition/transversion ratio
RNA	ribonucleic acid
s	second(s)
SDS	sodium dodecyl sulfate
0.1 × SSC	15 mM sodium chloride-1.5 mM sodium citrate
1 × SSC	150 mM sodium chloride-15 mM sodium citrate
2 × SSC	300 mM sodium chloride-30 mM sodium citrate
6 × SSC	0.9 M sodium chloride-90 mM sodium citrate
T	thymine
TBE	9 mM Tris-borate/1 mM Na_2EDTA Electrophoresis Buffer
TE7.6	10 mM Tris-Na_2/1 mM EDTA Buffer at pH 7.6
TE8	10 mM Tris-Na_2/1 mM EDTA Buffer at pH 8.0
TEMED	tetramethylethylenediamine
Tm	melting temperature of primer
Tris	tris(hydroxymethyl)aminomethane
Tris-HCl	tris buffer adjusted with HCl
µl	microlitre(s)
µg	microgram(s)
µM	micromolar
U	unit(s)
UTR	untranslated region
v	volume
w	weight
w	ratio of d_N/d_S

ACTIN GENE ABBREVIATIONS (FIGURES)

Aplysia Cy	*Aplysia californica* cytoplasmic actin gene
Aplysia M	*Aplysia californica* muscle actin gene
B. alex	*Biomphalaria alexandrina*
B. glabrata	*Biomphalaria glabrata* M line
B. obstructa	*Biomphalaria obstructa*
B. pfeifferi	*Biomphalaria pfeifferi*
B. tena	*Biomphalaria tenagophila*
Chtenopteryx Act I	*Chtenopteryx sicula Actin 1*
Chtenopteryx Act II	*Chtenopteryx sicula Actin 2*
Chtenopteryx Act III	*Chtenopteryx sicula Actin 3*
Crass 2	*Crassostrea gigas actin 2*
Crass GIA	*Crassostrea gigas GIA*
Dreissena	*Dreissena polymorpha*
H. disc	*Haliotis discus hannai*
Helisoma	*Helisoma trivolvis*
H.irisA1	*Haliotis iris actin 1*
H.irisA1a	*Haliotis iris actin 1 subtype a*
H.irisA1b	*Haliotis iris actin 1 subtype b*
H.irisA1c	*Haliotis iris actin 1 subtype c*
H.irisA2	*Haliotis iris actin 2*
H.irisA3	*Haliotis iris actin 3*
Histioteuthis Act I	*Histioteuthis hoylei Actin 1*
Histioteuthis Act II	*Histioteuthis hoylei Actin 2*
Histioteuthis Act III	*Histioteuthis hoylei Actin 3*
H.virgA1a	*Haliotis virginea actin 1 subtype a*
H.virgA1b	*Haliotis virginea actin 1 subtype b*
H.virgA1c	*Haliotis virginea actin 1 subtype b*
H. rufes	*Haliotis rufescens*
Idiosepius Act I	*Idiosepius pygmaeus Actin 1*
Idiosepius Act II	*Idiosepius pygmaeus Actin 2*
Idiosepius Act III	*Idiosepius pygmaeus Actin 3*
Mytilus	*Mytilus galloprovincialis*
Patella 1	*Patella vulgata pPA1* subtype
Patella 2	*Patella vulgata pPA2* subtype
Patella 3	*Patella vulgata pPA3* subtype
Pisaster Cy	*Pisaster ochraceus* cytoplasmic actin gene
Pisaster M	*Pisaster ochraceus* muscle actin gene
Placopecten	*Placopecten magellanicus*
Se. officinalis Act I	*Sepia officinalis Actin 1*
Se. officinalis Act II	*Sepia officinalis Actin 2*
Se. officinalis Act III	*Sepia officinalis Actin 3*
Se. opipara Act I	*Sepia opipara Actin 1*
Se. opipara Act II	*Sepia opipara Actin 2*

Se. opipara Act III	*Sepia opipara Actin 3*
Spirula Act I	*Spirula spirula Actin 1*
Spirula Act II	*Spirula spirula Actin 2*
Spirula Act III	*Spirula spirula Actin 3*
Vampyroteuthis Act I	*Vampyroteuthis infernalis Actin 1*
Vampyroteuthis Act II	*Vampyroteuthis infernalis Actin 2*
Vampyroteuthis Act III	*Vampyroteuthis infernalis Actin 3*

Chapter 1:

GENERAL INTRODUCTION

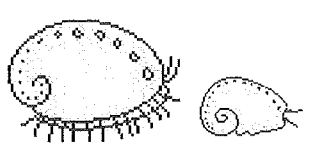


TABLE OF CONTENTS

LIST OF FIGURES

ABALONE

WHY MOLLUSCS?

Molecular analyses show that Bilaterians can be divided into three main groups: ecdysozoans, deuterostomes and lophotrochozoans. The first two groups contain some of the best understood molecular model systems, the ecdysozoans contain the arthropods and nematodes, the deuterostomes contain the vertebrates, ascidians and sea urchins. The third division is a large group which contains a diversity of taxa, including the annelids, platyhelminths, nemertines, bryozoa, phoronida, brachiopoda and molluscs, but it is poorly studied at the molecular level (Tessmar-Raible and Arendt, 2003). Recent developmental genetic studies have realised the importance of lophotrochozoan models and have demonstrated that they share many features previously considered to be vertebrate traits (O'Brien, 2000; Lespinet *et al.*, 2002; Nederbragt *et al.*, 2002a; Nederbragt *et al.*, 2002b). Consequently, lophotrochozoans, such as molluscs, may play an important role in understanding the evolution of Bilateria.

The Phylum Mollusca is divided into seven classes: Aplacophora, Polyplacophora (chitons), Monoplacophora, Scaphopoda (tusk shells), Bivalvia (bivalves), Gastropoda (gastropods, slugs and snails) and the Cephalopoda (octopuses and squid). Gastropods are the largest group, containing over 30,000 species (Ruppert and Barnes, 1994). Within the Gastropoda there are three subclasses: Prosobranchia (marine, freshwater and terrestrial forms with a single assymetrical shell forming a mantle cavity, cephalisation and torsion), Opistobranchia (mostly marine forms with detorsion and reduction of the shell and mantle cavity, often with secondary bilateral symmetry) and the Pulmonata (gill-less forms with mantle converted to a vascular chamber for gas exchange). Within the Prosobranchia is the Order Archeogastropoda, which contains the Family Haliotidae, the abalone.

ABALONE BIOLOGY

Abalone are intertidal archeogastropods found internationally in tropical and temperate oceans, often in the subtidal zone (Lindberg, 1992). Abalone are characterised by an auriform shell, no adult operculum, enlargement of the right shell muscle, displacement of the mantle cavity to the left and a row of perforations (tremata) in the shell. There are three abalone species in New Zealand waters, *Haliotis iris* (black-footed paua), *H. australis* (yellow-footed or queen paua) and *H. virginea* (white-footed or virgin paua). *H. iris* is the largest and most common of the New Zealand species, it reaches 200 mm in length and is distributed along the coast of the North Island, South Island, Stewart Island, Chatham Islands and Snares Islands (Sainsbury, 1982; Schiel, 1992). *H. australis* and *H. virginea* also occur around New Zealand's coasts, but are much rarer than *H. iris* (Poore, 1969).

LIFE CYCLE

Abalone are long-lived, slow growing, sedentary snails (Poore, 1972b; Poore, 1972c; Sainsbury, 1982) that use broadcast spawning strategies to aid in dispersal (Tong *et al.*, 1992). Abalone males and females spawn simultaneously during late summer to early autumn in response to climatic triggers, such as shortening day length and decreases in water temperature (Poore, 1973). An overview of the abalone life cycle is shown in Figure 1.1. Development is dependent on water temperature, spawning may often occur earlier in cooler southern waters (Tong *et al.*, 1992).

Fertilised embryos undergo spiral cleavage to produce a blastula which develops into a trochophore larva within 24 h at 15°C (Verdonk and Van Den Biggelaar, 1883). Trochophore are distinguished by a band of cilia called a prototroch which provides propulsion for a free swimming lifestyle that aids dispersal. Trochophore metamorphose into veliger larvae within a further 24 h (Tong *et al.*, 1992). Following dispersal larvae move to the substrate and settle in response to algal chemical cues, this is usually between seven and ten days

(McShane, 1992; Morse, 1992; Tong and Moss, 1992; Tong *et al.*, 1992). Veliger larvae loose their ability to swim and begin to feed and develop the body plan of the adult. Veligers grow to form juveniles which live in the subtidal zone. Once juveniles are approximately 10-15 mm in length they move to the intertidal zone. At maturity abalone occupy deeper subtidal waters, up to 20-60 m depth (Poore, 1972a; Lindberg, 1992).

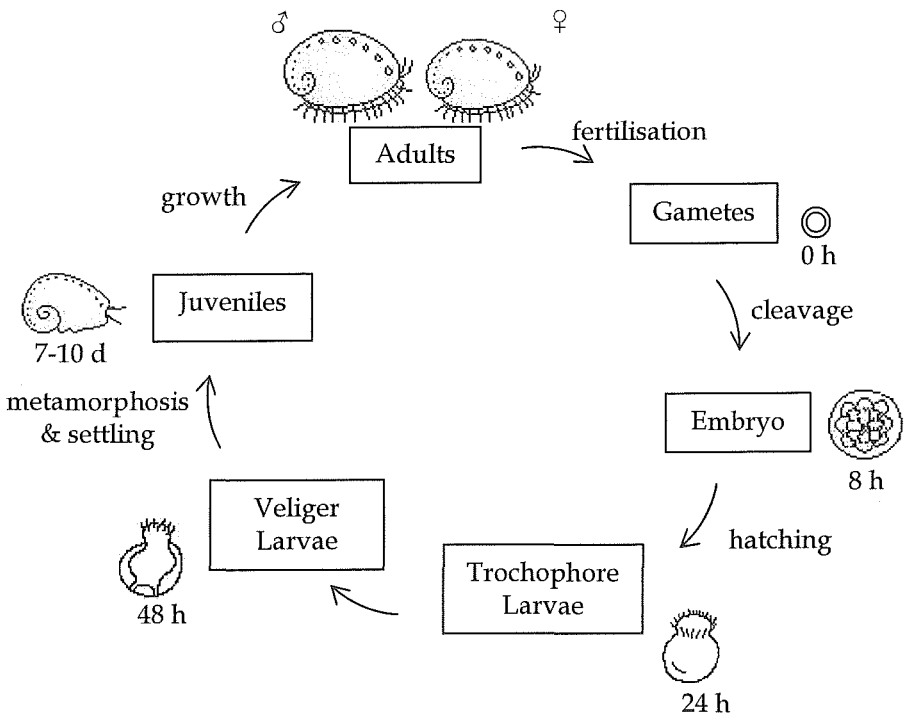


Figure 1.1. Abalone life cycle. Male and female adults spawn in response to environmental triggers such as falling water temperature. Fertilisation results in a zygote that undergoes spiral cleavage to produce a blastula. A blastula hatches into a swimming trochophore larvae. Trochophores metamorphose into veliger larvae which settle on the substrate in response to algal chemical cues. Veligers feed and develop into juveniles which continue to grow and mature into adults. The approximate lengths of time for development at 15°C are shown.

DISTRIBUTION

Extant abalone are distributed globally, though fossils records show that distributions have changed over geological time due to climatic changes and tectonic activity (Lindberg, 1992).

Contemporary abalone populations are influenced by temperature gradients and depth. Most species are found in warm temperate, subtropical

and tropical regions. Temperature is also associated with abalone size, larger abalone are found in cooler waters. Abalone may be found at depths ranging from the intertidal to 400 m. Most North American, Australian, South African and New Zealand abalone are found in waters less than 20 m in depth. In tropical regions abalone are found at deeper depths (Lindberg, 1992).

Research indicates that abalone populations are distributed due to habitat heterogeneity and the stochastic process of settlement and recruitment and that population structure may result from the accumulation of biomass (Sainsbury, 1982). In New Zealand abalone are located on shallow rocky coastlines, but are most abundant in cooler waters below latitude 41°S.

SYSTEMATICS

All abalone are grouped within one family, Haliotidae. The Haliotidae contains two genera; the genus *Haliotis,* which contains approximately 70 species, and the genus *Sulculus*, which contains one species (*S. diversicolor*). *Haliotis* has been divided into more than 15 subgenera (Lindberg, 1992). The division of genera is based on morphological characters, including the ratio of shell to body size, shell sculpture and the morphology of the tremata. The choice of these characters for classification is questionable, they change easily (reduction or change in shell size or sculpture is common amongst gastropods and tremata change due to habitat conditions and during development) and can therefore not be used as a guide to taxonomic relationships with confidence. In some studies of localised abalone groups epipodial structures have been used successfully to distinguish species. Epipodial characters are conservative within a species, with hybrids showing intermediate characters. Taxonomy is also confused by the existence of hybrids which may show character states that are intermediate to the states found in the parent species. Hybridisation has been described along the Californian coast, where seven abalone species coexist (Owen *et al.*, 1971; Lindberg, 1992; Geiger, 1998).

The ribosomal DNA internal transcribed spacers (rDNA ITS1 and ITS2) have been analysed to clarify haliotid taxonomy (Coleman and Vacquier, 2002).

Eighteen haliotids, representing North Pacific, Caribbean, European, South African, Australian, Taiwanese and New Zealand species, were analysed. Only one New Zealand species, *H. iris*, was studied. This approach separated *H. iris* from all other haliotids (100% bootstrap support). Of the remaining haliotids, three distinct subclades were supported, a North Pacific/Caribbean, a European and an Australian subclade (to which *H. midae* (South African) and *H. diversicolor* (Taiwanese) are basal). Previous studies of the cDNA sequence of lysin and its 3' untranslated flanking region and mitochondrial DNA also support these groupings (Lee *et al.*, 1995; Lee and Vacquier, 1995; Metz *et al.*, 1998).

ACTIN

THE ACTIN PROTEIN

Actins are a ubiquitous group of eukaryotic proteins. The actin protein is highly conserved, showing differences of only 3.5%-6.7% between species as divergent as *D. melanogaster* and humans (Bray, 1973; Hightower and Meagher, 1986).

STRUCTURE

The actin polypeptide is 376 aa in size (including the NH_2-terminal methionine), with a molecular mass of approximately 43,000 Daltons. The most notable feature of actin is a highly acidic NH_2-terminal sequence. Actin exists in two forms: G-actin, the globular monomeric molecule, and F-actin, the filamentous polymer (Pollard and Weihing, 1974; Korn, 1982).

G-actin has a clam shell-like three dimensional structure, it consists of two halves (domains) which can open and close. Within the crevice formed by the two halves of the protein is a tightly bound ATP or ADP molecule (Figure 1.2A) (Kabsch and Vandekerckhove, 1992). The actin molecule has a single high affinity and several low affinity binding sites for divalent cations. Within the cell the bound cation is usually Mg^{2+}, which associates with the phosphates of the bound nucleotide (Pollard and Weihing, 1974). G-actin will polymerise to form F-actin in the presence of physiological concentrations of salts. High concentrations of G-actin are maintained in the cell by specialised actin-binding proteins, such as profilin and thymosin β_4, which bind G-actin and inhibit its incorporation into F-actin (Pollard and Cooper, 1986; Kabsch and Vandekerckhove, 1992).

F-actin is a flexible two-stranded helical polymer with a diameter of approximately 5-9 nm (Kabsch and Vandekerckhove, 1992). F-actin is a polar molecule, it has a plus (or barbed) end and a minus (or pointed) end (Figure 1.2B). F-actin is organised into linear bundles, two-dimensional networks or

three-dimensional gels which are stronger than filaments (Korn, 1982; Pollard and Cooper, 1986).

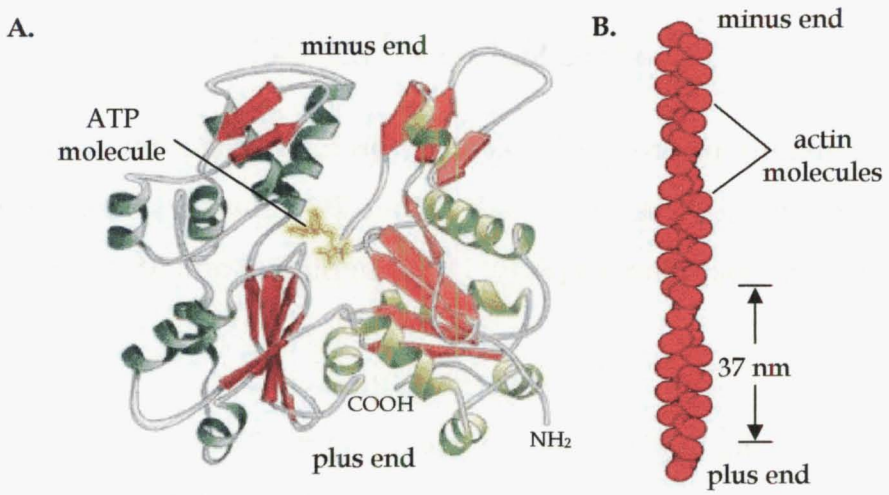


Figure 1.2. The structure of actin. A. Three-dimensional structure of G-actin. G-actin forms a clam shell-like shape, the two domains are hinged, but can open and close, forming a crevice. An ATP molecule (yellow) is tightly bound within the crevice. **B. Schematic diagram of F-actin.** G-actin molecules polymerise to form an F-actin polymer, a polar two-stranded helix. (From Alberts *et al.*, 1994).

POLYMERISATION

Polymerisation of G-actin to form F-actin occurs in the presence of K^+ and Mg^{2+}. Initially there is a slow lag phase as nucleation (molecule aggregation to form an arrangement amenable to polymerisation) occurs. Nucleation is proportional to the cube of the G-actin concentration, suggesting that a trimer of G-actin is required for polymerisation. Following the lag phase there is a rapid polymerisation phase. Polymerisation is proportional to the G-actin concentration, indicating that elongation occurs by the addition of one G-actin molecule at a time. Polymerisation stops when G-actin and F-actin reach an equilibrium. Following polymerisation the ATP is hydrolysed, causing the two halves of the G-actin molecule to be clamped together, trapping the ADP molecule. The addition of G-actin to a F-actin polymer may occur faster than the rate of ATP hydrolysis, causing an ATP cap to form (Pollard and Cooper, 1986; Carlier, 1990; Pollard, 1990; Kabsch and Vandekerckhove, 1992).

Actin filaments are dynamic, as actin molecules are being added to the plus end actin molecules are being lost from the minus end (Reisler, 1993). Addition usually occurs at the plus end as the rate of polymerisation is 10 times faster at the plus end than the minus end. Polymerisation results in a conformational change in actin which kinetically favours the incorporation of G-actin at the plus end. ATP hydrolysis reduces the binding affinity of actin molecules, promoting depolymerisation, so actin molecules are usually lost from the minus end of F-actin (Carlier, 1990). When the addition and loss of actin molecules is equal the filament maintains its length, this process is called treadmilling. Changes in the relative rates of actin addition and loss allows the regulation of filament length (Carlier and Pantaloni, 1997).

FUNCTION

The functions of actin can be divided into two main types, cytoplasmic functions (non-muscle) and muscle functions. Differences in function lead to differences in the structure of actin found in muscle and non-muscle cells, muscle cell actin is polymerised, whereas non-muscle cell actin polymers occur at lower levels and polymerisation is temporally and spatially regulated (Korn, 1982).

In vertebrates, actins can be divided into cytoplasmic and muscle forms based on characteristic aa differences (Vandekerckhove and Weber, 1978). Muscle-specific amino acids have also been identified in insect muscle actins, but not in other invertebrates (Mounier *et al.*, 1992; White and Crother, 1999). In molluscs and echinoderms functional analyses are required to distinguish the different types of actins.

CYTOPLASMIC TYPE ACTINS

Cytoplasmic actin filaments are organised into a network, the cell cortex, below the cell membrane. This cortex forms part of the cytoskeleton and is integral to the diverse cellular functions of actin. Some of the key functions of actin in non-

muscle cells are the maintenance of cell shape and motility and the creation of cell polarity. The ability of cytoplasmic actin to fulfil these functions is dependent on the sequestration and localisation of G-actin to active sites (i.e., near secondary messenger systems or at the base of pseuodpodia) for rapid polymerisation and the presence of a suite of accessory proteins which regulate actin polymerisation (Cao *et al.*, 1993; Pollard *et al.*, 1994).

Actin's role in facilitating cell motility is due to its ability to form a solid cytoskeletal structure which can influence the shape of the cytoplasm and the cell membrane (Cooper, 1991). Studies in amoeba show that actin and other regulatory proteins form a cytoplasmic gel that can be contracted and extended to cause cell movement (Pollard, 1976). Studies in *Dictyostelium* suggest that actin and myosin filament interaction, via the 'sliding filament mechanism' used in muscle (see below), may also cause cellular contraction and produce mechanical force sufficient for cell movement (Fukui, 1993). The force produced by the polymerisation of actin is sufficient to deform the cell membrane, leading to protrusions such as pseudopodia, which are supported by filamentous networks and osmotic pressure. Following protrusion, events such as membrane adhesion to the substrate and myosin mediated retraction of the filament network propel the cell forward (Condeelis, 1993).

Cell polarity has been studied in yeast where it is important in mating systems. In *Sacchromyces cerevisiae* two mating types, a and α, secrete hormones (a-factor and α-factor) which bind to receptors on the cell surface of opposite mating types and trigger a series of events leading to the formation of diploid cells and mating (Read *et al.*, 1992). One of these events is the formation of cell polarity and the formation of characteristically shaped cells with a pointed bud. Bud formation has been shown to depend on proteins which regulate cytokinesis (Flescher *et al.*, 1993). Further analysis showed that actin is required for buds to form (Read *et al.*, 1992).

MUSCLE TYPE ACTINS

Muscle type actins are one of the major components of muscle fibres and are integral to the biological mechanism which creates muscular force.

Muscle fibres are large cells formed by the fusion of individual cells. The majority of the cytoplasm of muscle fibres are made of myofibrils. Myofibrils are long cylindrical structures composed of concatemers of contractile units called sarcomeres (Pollard and Weihing, 1974). Each sarcomere is composed of parallel and overlapping bundles of thin and thick filaments and have a Z disc at each end of the sarcomere. Thin filaments are composed of actin, are attached to a Z disc at either end of the sarcomere and extend into the centre of the sarcomere. The thick filaments are composed of bundles of myosin II and overlap with thin filaments emanating from either end of the sarcomere (Figure 1.3).

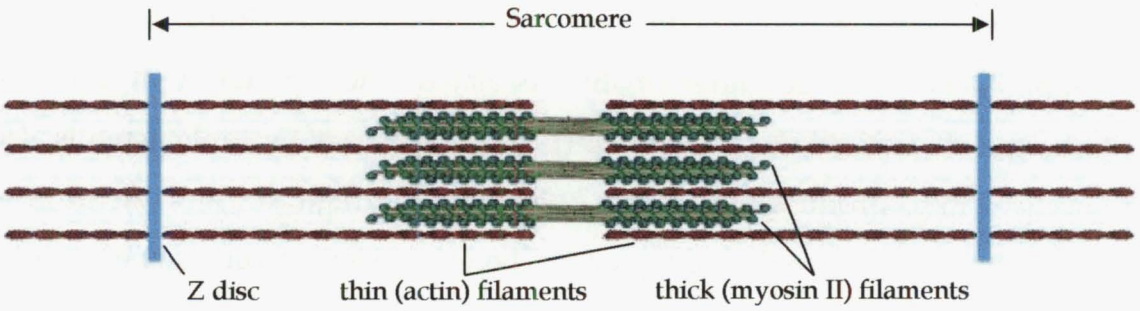


Figure 1.3. Sarcomere structure. Sarcomeres are contractile units which are arranged in concatemers within the myofibrils which compose muscle fibres. Sarcomeres have a Z disc at each end and contain parallel and overlapping thin and thick filaments. Thin filaments extend into the sarcomere from a Z disc and are composed of actin filaments. Thick filaments are situated in the middle of the sarcomere, overlap with thin filaments and are composed of myosin II bundles. (From Alberts *et al.*, 1994).

Muscle contraction is an ATP dependent mechanism in which the myosin II of thick filaments interact with and slide along the actin of thin filaments (Rayment *et al.*, 1993). Myosin II is a dimeric protein with a long tail and two globular heads which can hydrolyse ATP. Myosin II heads protrude from the thick filament and associate tightly with the adjacent thin filament. When an ATP molecule binds the myosin II head the binding affinity of myosin II to actin weakens. As the myosin II head interacts with the bound ATP

a conformational change occurs which causes the myosin II head to be displaced along the thin filament and ATP to be hydrolysed. Following hydrolysis the myosin II head associates weakly with a new area of thin filament, causing the inorganic phosphate liberated by ATP hydrolysis to be lost and the myosin II head to associate tightly with the thin filament. The energy created by the loss of inorganic phosphate allows the myosin II head to regain its original conformation, resulting in the movement of the thick filament in relation to the thin filament. This cycle is repeated leading to the shortening of the sarcomere and the contraction of muscle (Rayment *et al.*, 1993).

ACTIN GENE FAMILIES

Actin is encoded by a multigene family. Although the proteins produced by gene family members may often display extreme conservation, the maintenance of multiple genes is believed to have a functional basis (Rubenstein, 1990). Multiple genes can be differentially regulated and specialise in function (Fyrberg *et al.*, 1983; Mounier *et al.*, 1991). The existence of multiple genes is also a mechanism to produce high levels of a protein in high demand. Actin gene families will be reviewed in Chapter 2. The expression of actin genes will be reviewed in Chapter 3. The evolution of actin gene families will be reviewed in Chapter 4.

AIMS OF THIS STUDY

Molluscs are interesting research animals as they are a diverse group of organisms which represent one of the main clades of Bilaterians. The New Zealand black-footed abalone, *H. iris*, is important due to the economic and customary value associated with its meat and shell. The majority of *H. iris* research focuses on biology, ecology and husbandry. The literature investigating *H. iris* genetics is depauperate. As aquaculture of *H. iris* progresses, better understanding of abalone genetics is required to assist in technological innovation and productivity.

Actin is of interest as it is essential to the structure of eukaryotic cells and metazoan muscle. Actin is integral to a range of processes at the cellular and organismal level, such as maintenance of cell structure and animal locomotion via muscular contraction. The actin gene family is of interest for three reasons. Firstly, multiple genes elevate production of a protein in high demand. Secondly, multiple genes employ differential expression to regulate actin's numerous functions. Thirdly, gene families of highly conserved genes can be used to study the process of molecular evolution and diversification among taxa.

The aims of this research are to investigate the actin gene family in the New Zealand abalone. The objectives are:

1. To identify and characterise actin gene family members (Chapter 2)
2. Analyse the expression of abalone actin genes during embryonic development and in adult tissues (Chapter 3)
3. Investigate the evolutionary relationship of the abalone actin gene family (Chapter 4)

Chapter 2:

THE ACTIN GENE FAMILY OF *HALIOTIS IRIS*

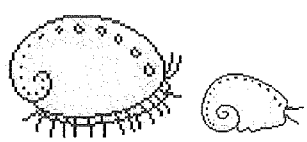


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

This chapter will describe the isolation and structural analysis of putative actin genes from the New Zealand black-footed abalone, *Haliotis iris*. *H. iris* putative actin genes will be compared to actins from invertebrate taxa with well characterised actin families. These invertebrates are from the Diptera (*Drosophila melanogaster* and *Bactrocera dorsalis*), Lepidoptera (*Bombyx mori*) and Echinodermata (*Pisaster ochraceus, Strongylocentrotus purpuratus* and *S. franciscanus*). *H. iris* actins will also be compared to actins which have been isolated from other mollusc species (*H. discus hannai, H. rufescens, Biomphalaria glabrata* (M line), *B. alexandrina, B. obstructa, B. tenagophila, B. pfeifferi, Helisoma trivolvis, Patella vulgata, Aplysia californica, Placopecten magellanicus, Crassostrea gigas, Mytilus galloprovincialis and Dreissena polymorpha*). A summary of the invertebrate actin genes considered in this chapter are listed in Table 2.1. A schematic of invertebrate genes is given in the Results section, Figure 2.18.

CHARACTERISED INVERTEBRATE ACTIN GENE FAMILIES

GENE NUMBER

Studies in the fruit fly *Drosophila*, the oriental fruit fly *Bactrocera dorsalis* (also known as *Dacus dorsalis*), the Mediterranean fruit fly *Ceratitis capitata* and the house fly *Musca domesticus* suggest that dipteran actin gene families contain six members (Fyrberg *et al.*, 1980; Haymer *et al.*, 1990; He and Haymer, 1991; Hadden and Sodja, 1994). The only study to identify all the genes of a dipteran actin gene family has been in *Drosophila* (Fyrberg, 1984).

In situ hybridisation of a *Drosophila* actin probe (a clone isolated from a genomic library using chicken and *Dictyostelium* actin cDNA as probes) to *Drosophila* polytene chromosomes confirmed that there are six *Drosophila* actin genes and that they are dispersed throughout the genome (Fyrberg *et al.*, 1980; Fyrberg *et al.*, 1981). There is no evidence to support a functional cause for the

distribution of *Drosophila* actin genes (Fyrberg, 1984). The six actin genes have been named after their chromosomal positions, they are *act5C, act 42A, act88F, act57B, act87E* and *act79B* (Zulauf *et al.*, 1981).

Four of the six putative *Bactrocera* actin genes, designated *BdA1, BdA2, BdA3* and *BdA5*, have been isolated and sequenced (He and Haymer, 1994).

Three actin genes were isolated from a *Bombyx mori* (domestic silkworm) genomic DNA library by screening with a *Drosophila* actin cDNA clone (Mounier and Prudhomme, 1986). A *Bombyx* actin clone, of the gene *A3*, was used as a Southern hybridisation probe to confirm the actin gene number in *Bombyx*. Hybridisation results indicated that a multigene family of four members, *A1, A2, A3* and *A4,* exists in *Bombyx* (Mounier and Prudhomme, 1991).

Screening of sea star *Pisaster ochraceus* genomic and tube feet cDNA libraries and analysis of developmental expression patterns suggests that *Pisaster* actins can be divided into three groups and that there are five non-allelic genes (Kovesdi *et al.*, 1984; Kovesdi and Smith, 1985). Southern hybridisation analysis of restricted genomic DNA with an actin coding region probe showed that there is extensive polymorphism for *Pisaster* actins (Kovesdi *et al.*, 1984). Southern hybridisation analysis of genomic DNA with a tube foot actin UTR probe suggests there are at least three actin alleles in the tube feet (Kovesdi *et al.*, 1984).

The structure of the actin gene family has been well characterised in the sea urchin *Strongylocentrotus purpuratus*. Early work suggested that there were 5-20 genes representing five actin subtypes (Durica *et al.*, 1980). Further characterisation by hybridisation analysis with 3' terminal probes derived from five previously described actin genes (Scheller *et al.*, 1981; Cooper and Crain, 1982) was used to confirm the gene number (Lee *et al.*, 1984). Results indicated the presence of four actin gene subtypes: *CyI, CyII, CyIII* and *M* (Shott *et al.*, 1984). Southern blots also showed that eight genes were detectable and that these genes showed allelic variation between the three individuals examined (Lee *et al.*, 1984). The *M* and *CyI* subtypes both contained a single gene (*M* and

CyI), whereas the *CyII* and *CyIII* subtypes each contained three genes (*CyIIa, CyIIb, CyIIc* and *CyIIIa, CyIIIb, CyIIIc,* respectively).

In the sea urchin *S. franciscanus* the actin gene copy number is estimated to be between 15 and 20 (Johnson *et al.*, 1983). Two genes have been sequenced, *SfA 15A* and *SfA 15B* (Foran *et al.*, 1985).

GENE STRUCTURE

The six *Drosophila* actin genes all encode a predicted protein of 376 aa. The intron positions in *Drosophila* genes vary; *act57A* has an intron within codon 14, *act79B* and *act88F* have an intron within codon 309 (Fyrberg *et al.*, 1981; Fyrberg, 1984). The intron sizes or sequences have not been determined. Introns have not been detected in the other *Drosophila* actins.

Bactrocera actin genes encode predicted proteins of 376 aa (He and Haymer, 1994). *BdA1* and *BdA5* contain an intron within codon 309 (91 and 74 nt, respectively). *BdA2* contains an intron within codon 14 (174 nt). *BdA3* does not contain an intron.

Bombyx actin genes encode predicted proteins of 376 aa. *A3* and *A4* both contain a 92 nt intron within codon 117 (Mange *et al.*, 1996), *A1* and *A2* do not contain introns (Mounier and Prudhomme, 1986).

Two *Pisaster* actin genes have been characterised, *Cy* and *M* (Kowbel and Smith, 1989). Both genes encode predicted proteins of 376 aa. The intron patterns and sizes differ between the genes. Both genes contain introns between codons 42/43 (166 nt and 697 nt in *Cy* and *M*, respectively) and 122/123 (136 nt and 485 nt) and within codon 205 (268 nt and 273 nt), however *M* contains two further introns within codons 150 (350 nt) and 269 (588 nt). The existence of an intron in the 5' UTR of the *Pisaster* genes is suggested by putative splice sites. A further *Pisaster* actin, named FAT (for Fat Actin Transcript), was detected during analysis of the *Pisaster* spermatogenic cycle by northern hybridisation with an actin probe (Boom and Smith, 1989). This gene encoded a larger transcript (3.5 kb) than the *Cy* and *M* transcripts (2.3 and 2.1 kb, respectively).

Preliminary sequencing showed that FAT showed 80% sequence similarity at the nucleotide level and over 90% identity at the amino acid level to *Cy*.

Five *Strongylocentrotus purpuratus* actins have been sequenced, *M* (Crain *et al.*, 1987), *CyI* (Cooper and Crain, 1982), *CyIIb* (Durica *et al.*, 1988), *CyIIIa* (Akhurst *et al.*, 1987) and *CyIIIb* (Durica *et al.*, 1988). The predicted proteins of these genes are all 376 aa. The intron positions in *M* are between codons 42/43 and 122/123 and within codons 205 and 309. All the cytoplasmic genes sequenced share introns of variable length between codons 122/123 and within codon 205. Unlike the organisation of actin genes in *Drosophila*, the actin genes of *Strongylocentrotus purpuratus* show physical linkage. Two linkage groups are detectable, *CyI/CyIIa/CyIIb* and *CyIIIa/CyIIIb* (Scheller *et al.*, 1981; Lee *et al.*, 1984; Minor *et al.*, 1987).

Two actin genes from *Strongylocentrotus franciscanus* have been sequenced, *SfA 15A* and *SfA 15B* (Foran *et al.*, 1985). Like *Strongylocentrotus purpuratus*, these genes form a linkage group, being within 5.7 kb of each other on the same chromosome and in the same transcriptional orientation. Both genes contain two introns of variable length at the same positions. The first intron is between codons 122/123 and is 236 nt in *SfA 15A* and 224 nt in *SfA 15B*, the second intron is within codon 205 and is 172 nt in *SfA 15A* and 191 nt in *SfA 15B*.

ACTIN GENE TYPE

Expression analysis has been used to determine whether genes encode cytoplasmic or muscle actins (for details see Chapter 3). In *Drosophila* there are two cytoplasmic actin genes (*act5C* and *act42A*) and four muscle actin genes (*act57B*, *act79B*, *act87E* and *act88F*) (Fyrberg *et al.*, 1983; Fyrberg, 1984). The four isolated *Bactrocera* actin genes encode muscle actins (He and Haymer, 1984). *Bombyx* contains two cytoplasmic actin genes (*A3* and *A4*) and two muscle actin genes (*A1* and *A2*) (Mounier and Prudhomme, 1986; Mounier and Prudhomme, 1991; Mange *et al.*, 1996).

MOLLUSCAN ACTIN GENES

GASTROPODS

Partial sequences of three actin genes have also been isolated from the New Zealand black-footed abalone, *H. iris* (Flint, 2000). In this study actin gene regions were amplified by polymerase chain reaction (PCR) using primers derived from the common carp, *Cyprinus carpio*. Three primer combinations were used to amplify actin. When the PCR products from the three primer sets were cloned and sequenced they were found to represent three actin gene subtypes. The first primer set amplified a 381 nt fragment from trochophore cDNA. The second primer set amplified a 669 nt fragment from trochophore cDNA and genomic DNA. The third primer set amplified a 954 nt fragment from trochophore cDNA. The 381 nt sequence was 17 and 15% different to the 669 and 954 nt sequences, respectively, which were 9% different to each other. The 381 nt product will be referred to as *H.irisA1*, the 669 nt product as *H.irisA2* and the 954 nt product as *H.irisA3*.

The complete coding sequence of an actin from *Haliotis discus hannai* is available on the GenBank database (accession number AY380809, deposited by Ma *et al.*, 2003, unpublished). This gene predicted a protein of 376 aa.

A partial sequence of an *Haliotis rufescens* (red abalone) actin gene has been isolated and characterised in a study which aimed to isolate abalone actin gene promoters for gene constructs for transfer into abalone (Gomez-Chiarri *et al.*, 1999). The isolated gene spanned approximately 1,600 nt, and contained 917 nt of the 5' regulatory region, all of exon 1, an intron between codons 41/42 and part of exon 2. The available nucleotide sequence predicts 172 aa, 45.9% of the full actin protein. The predicted protein sequence is missing the second amino acid, suggesting a full length protein of 375 aa.

The *Patella vulgata* actin gene family has been investigated by Van Loon *et al.* (1993). Actin clones were isolated from a cDNA library prepared from 16 h-old larvae by screening with a hamster actin cDNA probe. Three of the nine clones that contained an actin related insert were sequenced. These

sequences were named *pPA1*, *pPA2* and *pPA3*. Each sequence encoded a predicted protein of 375 aa. Comparison of the three actin sequences showed high intraspecific variability, there was 4.9-13.7% sequence difference at the nucleotide level and 3.2-4% sequence difference at the amino acid level.

Because of the high similarity in the coding region, the 3' untranslated regions (UTRs) of the *Patella* actin clones were used to distinguish between actin subtypes. Analysis of the 3' UTRs differentiated the clones into four subtypes, named *pPA1*, *pPA2*, *pPA3* and *pA17* after the clone from which they were isolated. The members of each subtype were grouped based on high similarity between the 3' UTRs. The similarity was not 100%, but the difference was usually less than 5%. The maximum dissimilarity between actins of one subtype was found in the pPA3 group, two of its members differed by 15.3% (Van Loon *et al.*, 1993).

The number of *Patella* actin genes in three of the subtypes was determined by screening Southern blots of sperm DNA from one individual with the 3' UTR of *pPA2*, *pPA3* and *pPA17*. The 3' UTR of *pPA1* could not be used as a probe because it was only 39 nt in length. Results suggest there is one *pPA2* gene, two to four *pPA3* genes and one to three *pPA17* genes. Hybridisation patterns, obtained with a *pPA3* coding region probe, and the existence of library clones which hybridise to a coding region probe but not to 3' UTR probes, suggest more actin genes are present within the *Patella* genome (Van Loon *et al.*, 1993). Unfortunately, expression of the *Patella* actin subtypes was not investigated, preventing the assignment of cytoplasmic or muscle actin functions to the subtypes.

In the opistobranch *Aplysia californica* (Californian sea hare), Southern hybridisation analysis with *Eco*R I, *Hin*d III, *Bam*H I and *Bgl* II restricted DNA infers the presence of at least three to five genes (DesGroseillers *et al.*, 1994).

Two actin genes have been isolated from *Aplysia*. Both genes were isolated from an abdominal ganglion cDNA library. The first gene was expressed in the muscular sheath that surrounds ganglia and was considered to be a muscle specific actin (DesGroseillers *et al.*, 1990). The muscle gene was

used as a probe to isolate the second gene from the cDNA library. The second gene was expressed in neurons, but also in ovotestis and kidney, suggesting it may represent a cytoplasmic actin with ubiquitous expression (DesGroseillers *et al.*, 1994).

The *Aplysia* muscle actin sequence was composed of a 56 nt 5' UTR, a 1,131 nt coding region and a 1,182 nt 3' UTR (DesGroseillers *et al.*, 1990). The *Aplysia* cytoplasmic actin sequence was composed of a 66 nt 5' UTR, a 1,131 nt coding region and a 396 nt 3' UTR (DesGroseillers *et al.*, 1994). The predicted proteins of both genes are 376 aa long and were 95.2% identical to each other (DesGroseillers *et al.*, 1994).

An actin gene was isolated from the freshwater pulmonate snail *Biomphalaria glabrata* (Lardans *et al.*, 1997). This gene was isolated from a cDNA library by PCR using primers derived from the *Aplysia* muscle and *H. rufescens* gene sequences. This sequence, *Bg Act*, contained a 1,391 nt sequence, composed of a 70 nt 5' UTR, a 1,131 nt coding region and a 193 nt 3' UTR. The coding sequence predicted a protein of 376 aa.

PCR primers derived from the *Bg Act* sequence were used to amplify actin genes from other pulmonates (Adema, 2002). PCR using genomic DNA obtained full gene sequences from six *Biomphalaria* species, *B. glabrata* (M line), *B. alexandrina, B. tenagophila, B. obstructa, B. pfeifferi* and *H. trivolvis*. These genes were composed of two exons separated by a single intron between codon 42/43. The coding region was 1,131 nt long and predicted a protein of 376 aa. The intron length varied, the intron lengths were 671 nt (*H. trivolvis*), 734 nt (*B. pfeifferi*), 765 nt (*B. tenagophila* and *B. obstructa*), 769 nt (*B. alexandrina*) and 794 nt (*B. glabrata* M line). Comparison to molluscan actins suggests that these genes represent cytoplasmic type actins.

Southern hybridisation analysis of genomic DNA of *B. glabrata* (M line) and *H. trivolvis* digested with *Hind* III or *Hae* III suggested there are at least five members in pulmonate actin families (Adema, 2002).

BIVALVES

One actin has been isolated from a *Placopecten magellanicus* (sea scallop) adductor muscle cDNA library (Patwary *et al.*, 1996). This gene was composed of a 64 nt 5' UTR, a 1,131 nt coding region and a 544 nt 3' UTR. The gene encodes a predicted protein of 376 aa.

Investigation of actin gene number by Southern hybridisation of isolated actin cDNA probe to *Placopecten* genomic DNA indicated the presence of 12-15 genes. Comparison of adductor actin cDNA hybridisation to genomic DNA from 10 individual scallops revealed that there are at least three polymorphic loci, one of which contains up to five alleles (Patwary *et al.*, 1996).

The screening of a *Crassostrea gigas* (Pacific oyster) cDNA library isolated a 1,439 nt sequence, which encoded an actin of 376 aa, named *GIA*. Further screening of a genomic library with the *C. gigas GIA* clone isolated a clone containing untranslated sequences of *GIA*. There is a single intron of 643 nt between codons 42/43 (Cadoret *et al.*, 1999). A second *C. gigas* gene has been isolated from a mantle cDNA library. *C. gigas actin 2* encodes a predicted protein of 376 aa (Miyamoto *et al.*, 2002). The function of *C. gigas actin 2* was not confirmed by expression, but as it was isolated from a mantle cDNA library it may represent a muscle actin.

CEPHALOPODS

The number of actin genes in cephalopods was investigated by Southern hybridisation. Genomic DNA from the squid *Ommastrephes bartramii* and the octopod *Vampyroteuthis infernalis* were restricted with *Ava* I, *Eco*R I, *Hin*d III and *Pst* I and probed with an *O. bartramii* actin sequence. Banding confirmed the presence of three actin genes in both species (Carlini *et al.*, 2000).

PCR with degenerate vertebrate actin primers amplified two to three genes from 44 cephalopod taxa (representing the Nautiloidea subclass and four orders of the Coleoidea subclass; Sepioidea, Teuthoidea, Octopoda and Vampyromorpha), 82 sequences in total (Carlini *et al.*, 2000). Parsimony analysis divided the actin genes into three groups, called *Actin I*, *Actin II* and *Actin III*,

supporting the gene number suggested by Southern hybridisation. Comparison of *Actin I, II* and *III* sequences to actin sequences available in GenBank placed *Actin I* sequences with mollusc muscle actin genes, *Actin II* sequences with mollusc cytoplasmic actin genes and *Actin III* sequences basal to all metazoan actin genes. Expression analysis of the actin genes was not performed.

Table 2.1. Summary of invertebrate actin genes. For each actin gene the commonly used name, accession number, reference and sequence details are shown.

Taxonomic Group	Species	Gene Name	Accession Number	Reference	Type	Sequence	Size (aa)	Intron Site (codon)	Intron Size (nt)
Diptera (Flies)	*Drosophila melanogaster*	*act5C*	NM078497.2	Adams *et al.*, 2000	Cy	mRNA	376	14	?
		act42A	NM078901.2		Cy	mRNA	376	-	-
		act79B	NM079643.1		M	mRNA	376	309	?
		act87B	NM079076.3		M	mRNA	376	-	-
		act88F	NM057743.3		M	mRNA	376	309	?
		act57B	NM079486.2		M	mRNA	376	-	-
	Bactrocera dorsalis	*BdA1*	L12253.1	He & Haymer, 1994	M	DNA	376	309	91
		BdA2	L12254.1	He & Haymer, 1994	M	DNA	376	14	174
		BdA3	L12255.1	He & Haymer, 1994	M	DNA	376	-	-
		BdA5	L12256.1	He & Haymer, 1994	M	DNA	376	309	74
Lepidoptera (Moths)	*Bombyx mori*	*A1*	X05185.1	Mounier *et al.*, 1987	M	DNA	376	-	-
		A2	X06363.1	Mounier *et al.*, 1987	M	DNA	376	-	-
		A3	U49854.1	Mange *et al.*, 1997	Cy	DNA	376	117	92
		A4	U49644.1	Mange *et al.*, 1996	Cy	DNA	376	117	92
Asteroidea (Starfish)	*Pisaster ochraceus*	*Cy*	M26501	Kowbel & Smith, 1989	Cy	DNA	376	42/43	166
								122/123	136
								205	268
		M	M26500	Kowbel & Smith, 1989	M	DNA	376	42/43	697
								122/123	485
								151	350
								205	273
								269	588
Echinoidea (Urchins)	*Strongylocentrotus franciscanus*	*SfA 15A*	X03075.1	Foran *et al.*, 1985	-	DNA	376	122/123	236
								205	172
		SfA 15B	X03076.1	Foran *et al.*, 1985	-	DNA	376	122/123	224
								205	191

	Strongylocentrotus purpuratus	*CyI*	J01202.1	Cooper & Crain, 1982	Cy	DNA	376	122/123	233
								205	181
		*CyIIa**	AF034254.1	Arnone *et al.*, 1998	Cy	DNA	15*	-	-
		CyIIb	M35323	Durica *et al.*, 1988	Cy	DNA	376	122/123	224
		CyIIc	-			inferred	-	-	
		CyIIIa	M30511	Akhurst *et al.*, 1987	Cy	DNA	376	122/123	>362
								205	371
		CyIIIb	M35324	Durica *et al.*, 1988	Cy	DNA	376	122/123	897
		CyIIIc	-		Cy	-	-	-	-
		*M**	X05739.1	Crain *et al.*, 1987	M	DNA	376	42/43	>158
			X05740.1	Crain *et al.*, 1987				122/123	>104
			X05741.1	Crain *et al.*, 1987				205	>134
			X05742.1	Crain *et al.*, 1987				-	-
			X05743.1	Crain *et al.*, 1987				309	>418
Gastropoda (Snails)	*Haliotis rufescens*	*actin**	AF032125	Gomez-Chiarri *et al.*, 1999	-	DNA	375?	41/42	128
	Haliotis discus hannai	*actin*	AY380809	Ma *et al.*, unpublished	-	mRNA	376	-	-
	Patella vulgata	*pPA1*	-	Van Loon *et al.*, 1993	-	mRNA	375	-	-
		pPA2	-	Van Loon *et al.*, 1993	-	mRNA	375	-	-
		pPA3	-	Van Loon *et al.*, 1993	-	mRNA	375	-	-
	Aplysia californica	*Cy*	U01352	DesGroseillers *et al.*, 1994	Cy	mRNA	376	-	-
		M	X52868	DesGroseillers *et al.*, 1990	M	mRNA	376	-	-
	Biomphalaria glabrata	*Bg Act*	Z72387.1	Lardans *et al.*, 1997	Cy	mRNA	376	-	-
	Biomphalaria glabrataM	*BgMACT*	AF329436.1	Adema *et al.*,2002	Cy	DNA	376	42/43	794
	Biomphalaria alexandrina	*BalACT*	AF329437.1	Adema *et al.*,2002	Cy	DNA	376	42/43	769
	Biomphalaria tenagophila	*BteACT*	AF329440	Adema *et al.*,2002	Cy	DNA	376	42/43	765
	Biomphalaria obstructa	*BobACT*	AF329439	Adema *et al.*,2002	Cy	DNA	376	42/43	765
	Biomphalaria pfeifferi	*BpfACT*	AF329438	Adema *et al.*,2002	Cy	DNA	376	42/43	734
	Helisoma trivolvis	*HtrACT*	AF329441	Adema *et al.*,2002	Cy	DNA	376	42/43	671
Bivalvia (Bivalves)	*Crassostrea gigas*	*GIA*	AF026063	Cadoret *et al.*, 1999	-	mRNA/DNA	376	42/43	643
		actin 2	AB071191	Miyamoto *et al.*, 2002		mRNA	376	-	-
	Placopecten magellanicus	*actin*	U55046	Patwary *et al.*, 1996	M	mRNA	376	-	-
	Mytilus galloprovincialis	*actin*	AF157491.1	Mitta & Cadoret, unpubd.	-	mRNA	376	-	-
	Dreissena polymorpha	*actin*	AF082863.1	Lamers *et al.*, unpubd.	Cy?	mRNA	376	-	-

* Partial sequence, ? Inferred function

AIMS

The actin gene families of many invertebrates have been well characterised. The number of genes, the gene structure and the gene types have been determined for dipterans (*D. melanogaster* and *B. dorsalis*), a lepidopteran (*B. mori*) and echinoderms (*P. ochraceus, S. purpuratus* and *S. franciscanus*). However, the actin gene family within a mollusc species has not been well studied. Actin sequences have been isolated from several mollusc species, but the number of actin genes, the gene structure and the gene type has not been determined within a species. The aims of this research are to investigate the actin gene family in a mollusc, *H. iris*, in order to better understand the organisation of an actin gene family in molluscs.

The specific aims of this chapter are to:

1. Identify, sequence and characterise actin gene sequences from *H. iris*
2. Estimate the actin gene number in *H. iris*
3. Investigate the structure of the *H. iris* gene family

METHODS

The Hazardous Substances and New Organisms (HASNO) Act of 1998 requires that modification of genetic material from indigenous species requires the approval of local Māori. Approval to use standard DNA recombination technology on *H. iris* DNA was not given, so alternative approaches to gene discovery and sequencing were employed. *H. iris* actin genes were isolated by polymerase chain reaction (PCR) using primers based on partial actin gene sequences of *H. iris* (Flint, 2000) and *H. rufescens* (Gomez-Chiarri *et al.*, 1999). Subtype-specific PCR, using primers that bind regions of difference in the three *H. iris* genes previously detected, was used to obtain the sequence of each of the actin genes. The 5' regulatory and coding regions of *H.irisA1* were obtained by PCR with *H. rufescens* regulatory region primers. The 5' coding regions of *H.irisA2* and *H.irisA3* were obtained by PCR with an ATG start site based primer.

All solutions were made using distilled deionised water (ddH_2O), which was purified using a Barnstead Nanopure Ultrapurification system (supplied by Medic). Preparation of reagents, buffers and gels are given in Appendix A; suppliers are listed in Appendix B.

GENOMIC DNA EXTRACTION

Genomic DNA was purified from *H. iris* sperm. Sexually mature males were collected from Kaikoura (South Island, New Zealand). Sperm were aspirated by syringe from the gonad, diluted with 30 volumes of extraction buffer (0.1 M Tris-HCl pH 8.0, 0.1 M NaCl, 20 mM Na_2EDTA) and stored at -20°C until required.

Sperm were thawed and further diluted with extraction buffer, to a total dilution of $^1/_{100}$ volumes. Protein was digested with proteinase K and lipids solubilised with sodium dodecyl sulfate (SDS) at concentrations of 100 µg/ml

and 1%, respectively, overnight with shaking at 50°C. Digested protein was separated from the nucleic acid by organic solvent extraction. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the sample, followed by mixing for 20 min. The organic and aqueous phases were separated by centrifugation at 20°C for 15 min at 12,000 × *g*. The aqueous layer was collected with a wide bore pipette. The organic fraction was back extracted by mixing with a volume of TE8 for 20 min. The extraction mix was centrifuged as above and the aqueous layer was collected and added to the first aqueous fraction. The pooled aqueous fractions were extracted once more by mixing with a volume of phenol:chloroform:isoamyl alcohol (25:24:1) for 20 min, followed by centrifugation as above. The aqueous layer was then extracted twice by mixing with a volume of chloroform:isoamyl alcohol (24:1) for 20 min. At this point the interphase was clear and the aqueous layer was collected.

The DNA was precipitated by adding $^1/_{10}$ volume of 3 M NaOAc pH 5.2 and two volumes of ice cold 100% EtOH. The sample was mixed and stored at -20°C overnight to aid precipitation. The DNA was scooped out of solution using a Pasteur pipette tip shaped into a hook and transferred to a 1.5 ml centrifuge tube. The DNA was rinsed twice with 1 ml of 70% EtOH to remove salt residue. The DNA was air dried to remove traces of alcohol and dissolved in TE8 to a final concentration of approximately 0.1 µg/µl.

The concentration, yield and purity of DNA were determined by spectrophotometry at absorbances of 260 and 280 nm (using a Ultrospec II, LKB Biochrom Ltd). Approximately, 0.1 ml of sperm gave 420 µg of genomic DNA, which was free of protein contamination ($A_{260}/A_{280} = 1.8$).

POLYMERASE CHAIN REACTION (PCR)

PRIMER DESIGN

PCR primers already available in the lab had been designed using the *Cyprinus* (common carp) actin sequence (NCBI accession number M24113.1, Liu *et al.*, 1990). Primer design was based on the partial coding regions of the three

putative actin genes of *H. iris* (Flint, 2000) and the *H. rufescens* actin sequence (Gomez-Chiarri *et al.*, 1999). As new regions of the actin genes became available during this research, further primers were designed to facilitate sequencing. The primers used during this research are listed in Table 2.2.

18S rRNA Primers

Primers for the amplification of the highly conserved 18S rRNA were designed for use as a positive control for PCR. The 18S rRNA primers were designed from a partial *H. iris* 18S rRNA sequence (accession number AF492441). The primers used were 18SF (5'-aaacggctaccacatccag-3') and 18SR (5'-tgtgcctgctttgaacactc-3'). The Tm of both primers was 68.0°C.

Table 2.2. PCR primers used to amplify putative *H. iris* actin genes. For each primer the direction, sequence, Tm and the species or actin subtype from which the primer was designed are given. The primer code refers to the coding system used in Figure 2.2. The code indicates whether a primer is generic (g) i.e. primer binds all *H. iris* actin genes, or specific (s) i.e. binds a particular actin subtype. The primer direction is indicated in the primer name, F is for forward primers, R is for reverse primers. The reference column indicates when the primer was designed.

Primer		Sequence	Tm	Species	Reference
Code	Name	(5'→3')	(°C)[a]		
g1	A266F	tcacagtttagagggcaagt	66	*H. rufescens*	This research
g2	A623F	agtcaactcaaggagcgc	67	*H.irisA1*	This research
g3	A911F	tacaaccatggatgatgatg	64	*H. rufescens*	This research
g4	ATGF(T3)[b]	atggatgatgatgttgctgc	66	*H. iris*	This research
g5	A989F(T3)	agagctgtcttcccctccat	70	Haliotids	This research
g6	A4F	atcagggtgtcatggttggt	68	*C. carpio*	Flint (2000)
g7	A1F	tgatggactctggtgatgg	62.3	*C. carpio*	In laboratory
s1	A1686F1	catcaaggagaagctctgc	68	*H.irisA1*	This research
s2	A1686F2	atcaaagaaaaattagctt	55	*H.irisA2*	This research
s3	A1686F3	catcaaggagaaactggcct	68	*H.irisA3*	This research
g8	A5R	gtcaccgacgtagctgtcct	72	*C. carpio*	Flint (2000)
g9	A2R*	gttgaaggtggtctcatg	60.5	*C. carpio*	In laboratory
g10	A1503R*	acaccatcaccagagtc	64	*H. iris*	This research
s4	A1717R1	caagggcgatgtagcagagc	72	*H.irisA1*	This research
s4	A1717R1(T7)[c]	caagggcgatgtagcagagc	72	*H.irisA1*	This research
s5	A1719R2	gtctaaagcaatgtaagcta	62	*H.irisA2*	This research
s6	A1721R3	aagtcaagagccayataggc	67	*H.irisA3*	This research
s7	A2IntronR(T7)	tttcatgctgattggtcgaa	61.5	*H.irisA2*	This research
s8	A3IntronR(T7)	cactctcgtcgtagttcgat	68	*H.irisA3*	This research
g11	A4R*	aatccacatctgctggaagg	68	*C. carpio*	Flint (2000)
g12	A3R(T7)	tagaagcatttgcggtggac	68	*C. carpio*	Flint (2000)

* Primers used for manual sequencing (not amplification of sequencing templates)

[a] Tm calculated at 1 M Na^+

[b] (T3) sequencing tag, 5'-attaaccctcactaaaggga-3'

[c] (T7) sequencing tag, 5'-taatacgactcactataggg-3'

PCR REACTIONS

PCR reactions were performed in volumes of 12.5 or 25 µl using 0.5 ml PCR tubes. PCR reactions contained a forward and reverse primer (at 0.5 µM each), dNTPs (0.2 mM of each), reaction buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), ddH_2O (to volume) and *Taq* polymerase (0.25 U, Roche Applied Sciences). Template DNA (20 ng) was added and the reactions were cycled using an Eppendorf Mastercycler® Gradient thermal cycler. PCR cycles included an initial denaturation at 94°C for 2 min, 30 cycles of denaturation at 94°C for 20 s, annealing at 50-70°C (depending on primer Tm) for 20 s and extension at 72°C for 20-60 s (depending on the expected size of the product), followed by a final extension at 72°C for 5 min. PCRs were optimised by annealing temperature gradients.

ANALYSIS

PCR products were separated by electrophoresis through 1-2% agarose/ 0.5 × TBE buffer gels at 5 V/cm, using a HE33 Mini Horizontal Submarine electrophoresis tank (Hoefer, supplied by Amersham Biosciences) or a Wide Mini-Sub Cell GT (BioRad) and a Gene Power Supply 200/400 (Pharmacia, supplied by Amersham Biosciences). Gels were stained with ethidium bromide (0.5 µg/ml in ddH_2O) for 20 min. DNA was visualised by UV transillumination at 300 nm and photographed (UltraLum gel documentation system, supplied by SciTech).

SEQUENCING

TEMPLATE PREPARATION

Sequencing templates were amplified by PCR at optimised conditions in 3 × 25 µl reactions then pooled. Templates were purified prior to sequencing by precipitation. If non-specific products were present in the PCR the product of interest was excised from an agarose gel and purified by glass milk extraction. When multiple products of similar size were produced, products were resolved

by non-denaturing gel electrophoresis and excised, then extracted by crush n' soak purification.

Precipitation Purification

PCR products were precipitated at room temperature by adding an equal volume of 4 M NH_4OAc pH 5.2 and two volumes of 100% isopropanol. The solution was mixed and precipitated at room temperature for 20 min. The DNA was pelleted by centrifugation at room temperature for 10 min at 12,000 × *g*. The pellet was washed with 100 µl of 70% EtOH. The DNA pellet was air dried and dissolved in 15-40 µl of ddH_2O depending on the size of the pellet. The DNA concentration was estimated by agarose gel electrophoresis of 2 µl of sample, a DNA molecular weight marker with bands of known concentration was used as a reference.

Glass Milk Purification

The PCR product of interest was excised from an agarose gel with a scalpel and purified with glass milk (Ausubel *et al.*, 1997). The agarose slice was dissolved in 2.5 volumes of 6 M NaI at 55°C for 5 min. Glass milk was added (5 µl to extract up to 5 µg of DNA) and the solution was incubated at room temperature for 5 min, with occasional mixing, to allow the glass milk to bind the DNA. The glass milk/DNA was pelleted by brief centrifugation at 12,000 × *g* and the supernatant was removed. The pellet was washed by dissolution (by vortexing) in 500 µl cold wash solution (10 mM Tris-HCl pH 7.4, 0.5 mM Na_2EDTA, 50 mM NaCl, 50% EtOH), followed by centrifugation as above. The wash was repeated twice more. Following washing all of the wash solution supernatant was removed and the pellet was briefly air dried. The DNA was eluted from the glass milk by incubation with 30 µl of TE8 at 45°C for 5 min. Following elution the glass milk was pelleted by centrifugation at 15,000 × *g* for 1 min and the DNA supernatant collected. The DNA concentration was estimated by agarose gel electrophoresis as described above.

Crush n' Soak Purification

Similar sized PCR products were resolved by polyacrylamide gel electrophoresis, excised with a scalpel and extracted by the crush n' soak method (Sambrook *et al.*, 1989). The gel was cut into small pieces and DNA was extracted with two volumes of elution buffer (0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$, 1 mM EDTA pH 8.0, 0.1% SDS) at 37°C overnight. Gel debris was pelleted by centrifugation at 4°C for 1 min at 12,000 × *g*. The supernatant was collected with a finely drawn out Pasteur pipette. Residual DNA was retrieved from the gel debris by vortexing the gel in 0.5 volume of elution buffer, centrifugation as above and pooling the supernatant with the first supernatant. Two volumes of cold 100% EtOH were added to the supernatant and the DNA was precipitated on ice for 30 min. DNA was collected by centrifugation at 4°C for 10 min at 12,000 × *g*. The DNA pellet was dissolved in 200 µl of TE7.6 and re-precipitated by the addition of $^1/_8$ volume 3 M NaOAc pH 5.2 and two volumes cold 100% EtOH. DNA was pelleted by centrifugation at 4°C for 10 min at 12,000 × *g* and rinsed with 1 ml of 70% EtOH. The pellet was air dried and dissolved in 10 µl TE7.6. Dilutions of $^1/_{10}$, $^1/_{100}$ and $^1/_{500}$ were used as templates for PCR.

SEQUENCING REACTIONS

PCR products were sequenced via radiolabelled or fluorescently labelled primer cycle sequencing, using the Thermo Sequenase Cycle Sequencing Kit (USB, supplied by Amersham Biosciences). Radiolabelled primers were end labelled with γ^{33}P-dATP isotope (3000 Ci/mmol, NEN, supplied by Biolab Scientific). Fluorescently labelled T3 or T7 primers (IRD800 fluorophore, LiCor) were used to sequence templates amplified with PCR primers carrying T3 or T7 tags.

Radioactive Primer Labelling

Primers were radiolabelled by adding a γ^{33}P-dATP molecule to the 5' end of the primer with T4 polynucleotide kinase (PNK). Primer (20 pmoles) was mixed with 1.125 µl of γ^{33}P-dATP (10 µCi/µl), 1 × PNK buffer and 2.5 U of PNK in a

5 µl reaction. The reaction was incubated at 37°C for 30 min, then the enzyme was denatured by heating at 95°C for 5 min. Radiolabelled primers were stored at -20°C and were used for up to 3 weeks.

Reactions

A master mix was made for each sequencing reaction. The master mix was mixed by gently pipetting up and down, followed by brief centrifugation. Following aliquoting of the master mix to the ddNTP, the samples were overlaid with mineral oil.

For radioactive sequencing, the master mix contained approximately 100 fmoles of DNA template, 1 × reaction buffer, 1.5 µl of radiolabelled primer and 2 U of Thermo Sequenase enzyme. Two microlitres of the master mix was aliquoted to four tubes, each containing 2 µl of each ddNTP termination mix (ddATP, ddGTP, ddCTP or ddTTP). For fluorescent sequencing, the master mix contained approximately 200 fmoles of DNA template, 1 × reaction buffer, 0.5 µl of fluorescently labelled primer and 0.5 U of Thermo Sequenase enzyme. One microlitre of the master mix was aliquoted to 1 µl of each ddNTP termination mix.

Reactions were cycled in an Eppendorf Mastercycler® Gradient thermal cycler. The cycle conditions were: an initial step of 95°C for 5 min, followed by 29 cycles of 95°C for 30 s, primer annealing at 50-60°C for 30 s and extension at 72°C for 1 min, followed by nine cycles of 95°C for 30 s and 72°C for 1 min. For radioactive sequencing, the annealing step varied between 50 and 60°C depending on the Tm of the primer. For fluorescent sequencing, the annealing temperature was 50°C.

Following cycling, 2 µl (radioactive sequencing) or 1 µl (fluorescent sequencing) of stop buffer was aliquoted to each tube. The reactions were spun down briefly, then analysed by electrophoresis or stored at -20°C until electrophoresis (within less than a week from cycling). Reactions were denatured by heating at 94°C for 3 min, followed by chilling on ice, prior to electrophoresis.

ELECTROPHORESIS

Radioactive sequencing reactions were analysed by polyacrylamide gel electrophoresis using a Model S2 Sequencing Gel apparatus (Life Technologies) and an LKB 2197 Power Supply (supplied by Amersham Biosciences). The first 200-300 nt of reactions were analysed on 6% acrylamide (19:1)/1 × TBE gels, further reads (300-550 nt) were analysed on 5% acrylamide (19:1)/1 × TBE gels. Electrophoresis was performed at a constant power of 70 W (voltage ~1500 V, current ~38 mA) for 2 h. Following electrophoresis the gel was transferred to a filter paper solid support (Schleicher and Schuell) and dried at 80°C for 1 h on cycle 2 under vacuum, using a BioRad model 583 gel drier.

Fluorescent sequencing reactions were analysed by polyacrylamide gel electrophoresis on a LiCor 4000L automated sequencer. Samples were analysed through 3.5% acrylamide (19:1)/1 × TBE gels. Electrophoresis was performed at 2000 V, 25 mA and 45 W at 45°C, with the motor speed set at 2 and the channel at 3.

AUTORADIOGRAPHY

Radioactive sequencing gels were exposed using BioMax MR film (Kodak, supplied by Radiographic Supplies). Gels were exposed for 16 to 72 h depending on intensity of the signal detected by a Geiger counter. Film was processed by incubating in developer (Agfa G150, $^1/_6$ dilution) for 2 min, washing in stop solution (water with a splash of glacial acetic acid) for 30 s and fixing in fix solution (containing Ilford Hypam Fixer, $^1/_5$ dilution, and Ilford Hypam Hardener, $^1/_{41}$ dilution) for 4 min. The film was washed in running tap water for 15 min and then rinsed in ddH_2O before being air dried.

ANALYSIS

Overlapping sequences were aligned by eye. Nucleotide sequences (nucleotide and amino acid) were aligned using ClustalX 1.8 (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) (Thompson *et al.*, 1997). Protein sequences were aligned with ClustalW 1.7 (ftp://ftp-igbmc.u-strasbg.fr/pub/clustalW/).

Alignments were formatted using BioEdit 5.0.9 (www.mbio.ncsu.edu/BioEdit/bioedit.html) (Hall, 1999). Restriction enzyme cut sites were determined using DNA Strider™ 1.2 (www.personal.psu.edu/faculty/d/s/dsg11/labmanual/Plasmids/Strider_files/DNA_sequence_files.htm). Predicted protein sequences were determined using the ExPASy Translate tool (http://us.expasy.org/tools/dna.html), MEGA 2.1 (www.megasoftware.net/) (Kumar *et al.*, 2001) or BioEdit 5.0.9. Sequence identity was analysed using BLAST 2.2.8 (www.ncbi.nlm.nih.gov/BLAST/) (Altschul *et al.*, 1997).

SOUTHERN HYBRIDISATION

Southern hybridisation was used to confirm the number of actin genes present in *H. iris*. Confirmation of actin gene number was achieved by cutting genomic DNA with restriction enzymes that cut outside the coding regions of the identified actin genes, followed by probing with an actin probe that binds the highly conserved coding region of the actin genes sequenced. This approach was used to confirm the number of actin genes in *D. melanogaster* (Fyrberg *et al.*, 1980) and sea urchins (Lee *et al.*, 1984). The comparison of hybridisation banding between individuals may help identify actin alleles (Lee *et al.*, 1984). Hybridisation bands on the autoradiogram should correspond to individual genes within the *H. iris* genome, assuming that none of the actin genes have altered restriction sites.

ELECTROPHORESIS AND TRANSFER

Restriction Digestion

DNA Strider™ 1.2 was used to find the restriction digestion sites of the *H. iris* actin sequenced during this study (Appendix E). The enzymes *Cla* I and *Pst* I did not cut within the coding sequences available for the *H. iris* putative actin genes. *Eco*R I is predicted to cut *H.irisA1* within the regulatory region of the gene, *Hin*d III is predicted to cut *H.irisA1c* within intron B. These cut sites

would be upstream of the actin coding region probe, so hybridisation banding would still represent gene number.

H. iris DNA was digested with various restriction enzymes (Table 2.3). Restriction digests contained 5 U/µg of enzyme, 1 × enzyme specific restriction buffer and DNA at a concentration of 0.2 µg/µl. All the reaction components, except the restriction enzyme, were mixed together and heated at 65°C for 10 min to aid in the resuspension of the DNA. The reaction was chilled on ice prior to the addition of enzyme. Reactions were incubated overnight at the optimum temperature for the enzyme used. The extent of cutting was checked by running a 0.3 µg sample of the digestion reaction on a 0.8% agarose/0.5 × TBE gel. If a high molecular weight band (corresponding to that of an uncut control) was present more enzyme (20 U) was added and the digestion reaction was incubated further.

Table 2.3. Restriction enzymes used to digest genomic DNA for Southern hybridisation analysis. The enzymes used (and their recognition sites) are listed, along with supplier, buffer used and optimum temperature for activity.

Enzyme	Recognition Sequence	Supplier	Buffer	Temperature (°C)
Cla I	AT↓CGAT	USB	M	37
*Eco*R I	G↓AATTC	USB	H	37
*Hin*d III	A↓AGCTT	USB	M	37
Pst I	CTGCA↓G	Roche	H	37

Electrophoresis

Restriction digestion products were separated by electrophoresis through a 0.8% agarose/0.5 × TBE gel, using a Wide Sub-Cell GT electrophoresis tank (BioRad) and a Gene Power Supply 200/400 (Pharmacia, supplied by Amersham Biosciences). Five micrograms of restricted *H. iris* DNA was loaded per well. Electrophoresis was performed at 5 V/cm on 15 cm wide × 20 cm long gels. Gels were stained with ethidium bromide (0.5 µg/ml in ddH_2O) for 20 min and the DNA was visualised by UV transillumination at 300 nm and photographed (UltraLum gel documentation system, supplied by SciTech).

Southern Transfer

Gels were treated to improve the efficiency of DNA transfer. Treatments used were based on those specified for the use of Hybond-XL blotting membrane (Amersham Biosciences). The gel was incubated in 0.125 M HCl for 10 min, with shaking, to depurinate the DNA. Following depurination, the DNA was denatured by incubation in 1.5 M NaCl/0.5 M NaOH for 30 min with shaking. The gel was rinsed with ddH_2O between treatments.

DNA was transferred from the gel to Hybond-XL blotting membrane (Amersham Biosciences) overnight by capillary blotting as described by Southern (1975) and the Amersham Biosciences instructions. The transfer buffer used was 1.5 M NaCl/0.5 M NaOH.

Once transfer was complete, the membrane was dried between two pieces of filter paper (Schleicher and Schuell), then fixed by exposure to UV (254 nm, for 2 min).

HYBRIDISATION

Probe Labelling

Probes were labelled with α^{32}P-dCTP (3000 Ci/mmol) using a Random Primed DNA Labeling Kit (Roche Applied Sciences). The cloned 954 nt product (*H.irisA3*) was used as a DNA template (Flint, 2000). The template (50 ng) was made up to a 9 µl volume with ddH_2O, boiled for 10 min, then chilled on ice. Hexanucleotide mix (2 µl), dNTPs (dATP, dTTP and dGTP, 1 µl at 0.5 mM) and α^{32}P-dCTP (50 µCi) were added and mixed with the template by carefully pipetting up and down. Klenow enzyme (2 U) was added and the reaction was incubated at 37°C in a heating block for 2 h. Prior to use the probe was denatured by boiling for 10 min, then chilled on ice.

Hybridisation

Hybridisation was carried out in bottles using a Hybaid oven (supplied by SciTech). The membrane was pre-wetted in 2 × SSC before prehybridisation solution was added. Prehybridisation solution consisted of 6 × SSC, 5 × Denhardt's solution, 0.5% SDS and 0.1% blocking agent; 100 µl of solution

was used per square centimetre of membrane. Prehybridisation was performed at 65°C for 2 h.

Denatured probe was added to fresh prehybridisation solution, mixed and then exchanged with the prehybridisation fluid in the hybridisation bottle. Hybridisation was performed at 65°C for 16 h.

The next morning the hybridisation solution was removed and stored at -20°C. The membrane was briefly washed in 2 × SSC at room temperature. To remove non-specific binding, the membrane was washed at 65°C with 2 × SSC for 15 min and 2 × SSC/0.1% SDS (w/v) for 30 min. At this point, the membrane was removed from the bottle and the strength of the radioactive signal was determined using a Geiger counter. If the background signal was still strong the membrane was washed further with 1 × SSC/0.1% SDS, at 65°C, for 15 min, and if necessary with 0.1 × SSC/0.1% SDS, at 65°C, for 15 min. The membrane was then wrapped in plastic wrap and autoradiography was performed.

Autoradiography

The membranes were exposed in Kodak film cassettes, with intensifying screens fitted, using X-Omat AR film (Kodak, Radiographic Supplies). Exposure was carried out at -70°C, for a period ranging from 24 -48 h depending on the strength of the signal.

Films were developed as for sequencing films, as outlined in the Sequencing-Autoradiography section (page 42).

RESULTS

POLYMERASE CHAIN REACTION (PCR)

18S rRNA

An *H. iris* 18S rRNA PCR was designed as a control for PCR and RT-PCR work. A ribosomal gene was chosen as a control for PCR and RT-PCR, as ribosomal genes are constitutively expressed. Molluscan ribosomal gene sequences were retrieved from the GenBank database (www.ncbi.nlm.nih.gov). Searches for complete 18S rRNA gene sequences from vetigastropod molluscs isolated 12 entries. Of these two were sequences from haliotids, *H. discus hannai* (accession number AF082177, Yoon and Kim, 2000) and *H. tuberculata* (accession number AF120511, Giribet and Wheeler, 2002). A further search for haliotid 18S rRNA gene sequences lead to the retrieval of a partial *H. iris* 18S rRNA gene sequence (accession number AF492441, Reece and Stokes, 2003). The sizes of the haliotid gene sequences were *H. discus hannai* 1,858 nt, *H. tuberculata* 1,809 nt and *H. iris* 1,810 nt. The 12 vetigastropod sequences and the *H. iris* sequence were aligned to identify regions of high sequence conservation (Appendix D). PCR primers that bound within a conserved gene region were designed. Primers were located at nt 390-409 and 771-790 of the *H. iris* sequence, amplifying a predicted 401 nt product. The Tm of 18SF and 18SR was 68°C. PCR was performed at annealing temperatures between 59.5 and 68.9°C. A single product of the expected size was amplified between 59.5 to 64.8°C (Figure 2.1).

ACTIN

H. iris actin genes were amplified by PCR using *Cyprinus* (carp), *H. rufescens* (red abalone) and *H. iris* derived primers. Subtype-specific *H. iris* primers were designed based on the partial actin gene sequences isolated by Flint (2000). Following amplification of initial products, further primers were designed and used to amplify the remaining regions of the actin genes (actin PCRs are listed in Table 2.4, an overview of actin primer binding sites is given in Figure 2.2).

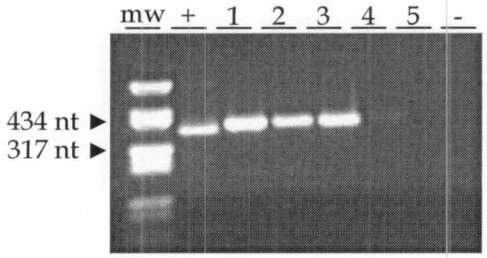


Figure 2.1. PCR amplification of *H. iris* 18S rRNA. PCR products were analysed on a 1.5% agarose/0.5 × TBE buffer gel. The molecular weight marker (mw) was pBluescript M13+ genomic DNA digested with *Hae* III. The positive control (+) was *H. iris* genomic DNA amplified with actin primers A1F and A2R at 50.1°C (386 nt). The negative control (-) was ddH_2O amplified with actin primers A1F and A2R at 50.1°C. 1-5 show 18SF/18SR PCR products amplified over an annealing temperature gradient of 59.5, 62.2, 64.8, 67.1 and 68.9°C. The *H. iris* 18S rRNA PCR product is 401 nt.

Table 2.4. Summary of actin PCRs. Primers, primer codes used in Figure 2.2, expected PCR product sizes and observed PCR product sizes are shown. Dashes (-) indicate that no product is expected.

Forward Primer		Reverse Primer		Expected	Observed Product (nt)		
Name	Code	Name	Code	Product (nt)	*H.irisA1*	*H.irisA2*	*H.irisA3*
A1686F1	s1	A3R	g12	496	496	-	-
A1686F2	s2			496	-	496	-
A1686F3	s3			496	-	-	496
A1F	g7	A1717R1	s4	210	210	-	-
		A1719R2	s5	212	-	212	-
		A1721R3	s6	214	-	-	214
A4F	g6	A1717R1	s4	546	546	-	-
		A1719R2	s5	548	-	548	-
		A1721R3	s6	550	-	-	550
A989F(T3)	g5	A1717R1	s4	728	757 812[a] 865[b] 884[c]	-	-
		A1719R2	s5	730	-	1,663	-
		A1721R3	s6	732	-	-	1,164
A266F	g1	A1717R1	s4	1,492	1,492 580* 800*	-	-
		A2IntronR(T7)	s7	834	-	0	-
		A2IntronR(T7)	s8	1,023	-	-	0
A266F	g1	A5R	g8	976	976	-	-
A911F	g3	A2IntronR(T7)	s7	187	-	187	-
ATGF(T3)	g4	A2IntronR(T7)	s8	369	-	-	369
A623F	g2	A1717R1	s4	1,228	1,228 1,850* 2,500*	-	-
		A1719R2	s5	2,134	-	0	-
		A1721R3	s6	1,635	-	-	0

* Non-specific product, identity not determined.

[a] *H.irisA1a*

[b] *H.irisA1b*

[c] *H.irisA1c*

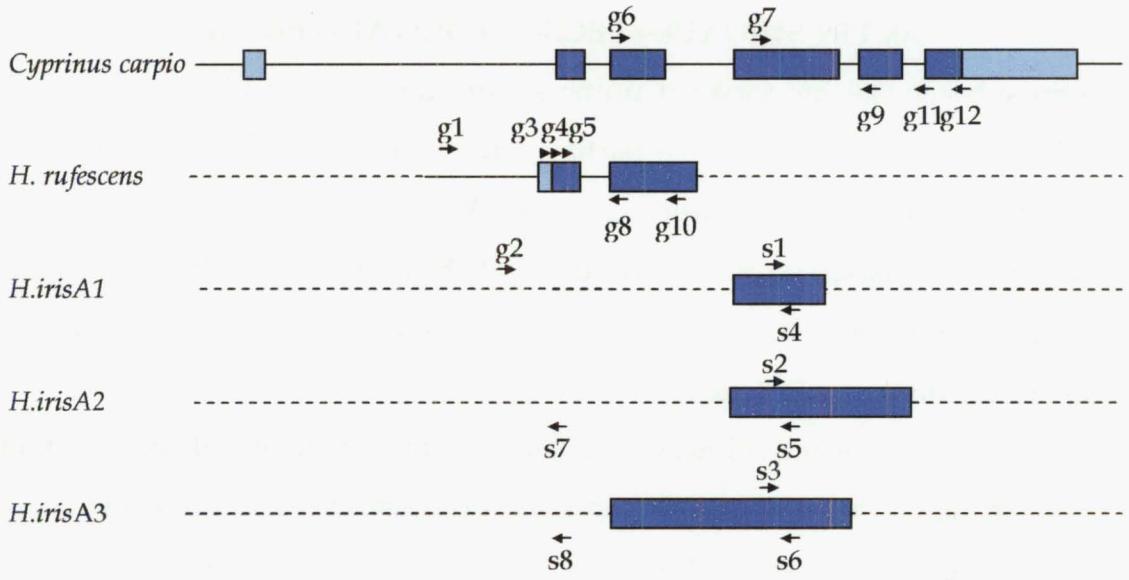


Figure 2.2. Overview of actin gene sequences used to design PCR primers and relative primer binding positions. Dark blue boxes represent exons, solid lines represent introns, light blue boxes represent untranslated regions and broken lines indicate undetermined sequence. Primers are shown as arrows (facing the direction of amplification) with a code number indicated (identities of codes as given in Table 2.2). Primer codes with a 'g' indicate a generic primer, codes with an 's' indicate a subtype-specific primer. Primers which bind within the undetermined sequences were designed during this research as sequence data became available. The sequences used were *Cyprinus carpio* (accession number M24113.1) and *H. rufescens* (accession number AF032125.1). The *H.irisA1*, *H.irisA2* and *H.irisA3* sequences were determined from sequencing of cloned actin PCR products amplified with *Cyprinus* primers (Flint, 2000).

ACTIN SUBTYPE-SPECIFIC FORWARD PRIMERS

Actin subtype-specific forward primers were designed to bind to a region of high diversity within the three partial putative actin gene products identified by Flint (2000). These were named A1686F1, A1686F2 and A1686F3 (nomenclature based on primer location at 1,686 nt within a haliotid actin guide sequence composed of *H. rufescens* and *H. iris* partial sequences and F1, F2 or F3 depending on the actin gene it amplifies).

Subtype-specific primers were used with the generic A3R primer, which binds at the end of the actin coding region of *Cyprinus*. The predicted sizes of the amplicons were 496 nt. The Tm of A1686F1, A1686F2 and A1686F3 were 68, 55 and 68°C, respectively; the Tm of A3R was 68°C. PCR was optimised using annealing temperature gradients (Figure 2.3). All three primer pairs amplified single products within the annealing temperatures used. A1686F1/A3R amplified products between 56.8-62.2°C. A1686F2/A3R amplified products between 50.1-52.3°C. A1686F3/A3R amplified products between 56.8-62.2°C.

ACTIN SUBTYPE-SPECIFIC REVERSE PRIMERS

Actin subtype-specific reverse primers were also designed to bind within the high diversity region of the three Flint (2000) actin sequences. Following the nomenclature system outlined above, these were named A1717R1, A1719R2 and A1721R3. The Tm for these primers were 72, 62 and 67°C, respectively. PCRs were performed with three generic forward primers, A1F, A4F and A989F(T3) (Tm 62, 68 and 70°C, respectively) and optimised using annealing temperature gradients.

The predicted size of PCR products amplified with A1F and the three subtype-specific reverse primers were 210, 212 and 214 nt, respectively. A single PCR product was obtained for A1F/A1717R1 between 54.4 and 64.8°C (Figure 2.4A). A single A1F/A1719R2 product was obtained between 50.9 and 56.8°C (Figure 2.4B). A single A1F/A1721R3 product was amplified between 54.4 and 62.2°C (Figure 2.3C).

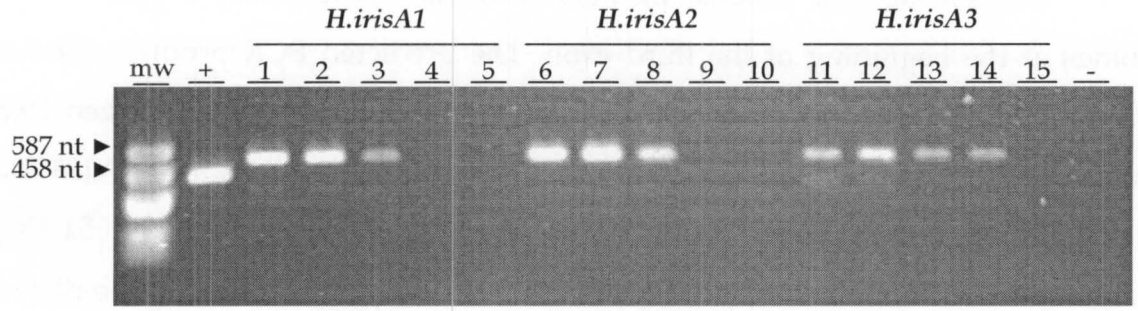


Figure 2.3. PCR amplification of *H. iris* actin genes with subtype-specific forward primers and generic primer A3R. PCR products were analysed on a 1.2% agarose/0.5 × TBE buffer gel. The molecular weight marker (mw) was pBluescript M13+ genomic DNA digested with *Hae* III. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 64.8°C (401 nt). The negative control (-) was ddH_2O amplified with 18SF/18SR at 64.8°C. 1-5 show A1686F1/A3R products amplified at 56.8, 59.5, 62.2, 64.8 and 67.1°C (496 nt). 6-10 show A1686F2/A3R products amplified at 50.1, 50.9, 52.3, 54.4 and 56.8°C (496 nt). 11-15 show A1686F3/A3R products amplified at 56.8, 59.5, 62.2, 64.8 and 67.1°C (496 nt).

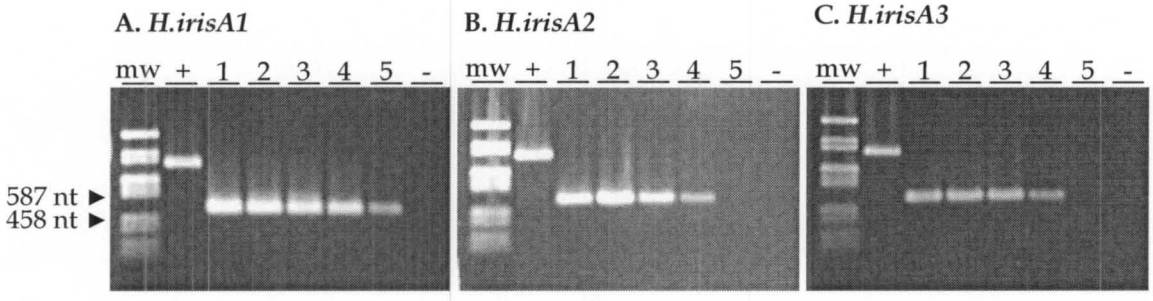


Figure 2.4. PCR amplification of *H. iris* actin genes with generic primer A1F and subtype-specific reverse primers. PCR products were analysed on 1.5% agarose/0.5 × TBE buffer gels. The molecular weight marker (mw) was pBluescript M13+ genomic DNA digested with *Hae* III. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 62.2°C (401 nt). **A. *H.irisA1*.** A1F/A1717R1R products (210 nt) amplified at 54.4, 56.8, 59.5, 62.2 and 64.8°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A1F/A1717R1 at 50.9°C. **B. *H.irisA2*.** A1F/A1719R2 products (212 nt) amplified at 50.9, 52.3, 54.4, 56.8 and 59.5°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A1F/A1719R2 at 50.9°C. **C. *H.irisA3*.** A1F/A1721R3 products (214 nt) amplified at 54.4, 56.8, 59.5, 62.2 and 64.8°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A1F/A1721R3 at 50.9°C.

Subtype-specific reverse primers were used with A4F, a primer that binds at the beginning of the third exon. The predicted PCR product size for A4F/A1717R1 was 546 nt. Single PCR products were obtained between 59.5 and 64.8°C (Figure 2.5A). The predicted PCR product size for A4F/A1719R2 was 548 nt. The expected product was amplified between 50.9 and 54.4°C, although faint non-specific products were visible (Figure 2.5B). The predicted PCR product size for A4F/A1721R3 was 550 nt. Single PCR products were amplified between 59.5 and 64.8°C (Figure 2.5C).

Subtype-specific reverse primers were also used with A989F(T3), a primer that binds in exon 1, upstream of the putative intron. Based on the intron size of the known *H. rufescens* sequence the expected sizes of the PCR products were 719, 721 and 723 nt, respectively. A989F(T3)/A1717R1 amplified four products between 62.2 and 68.9°C, sized 757, 812, 865 and 884 nt (Figure 2.6A). These were resolved on, and isolated from, a 4% non-denaturing polyacrylamide gel (Figure 2.10). A989F(T3)/A1719R2 amplified a single product of 1,663 nt at 56.8°C (Figure 2.6B). A989F(T3)/A1721R3 amplified a single product of 1,164 nt between 64.8 and 67.1°C (Figure 2.6C).

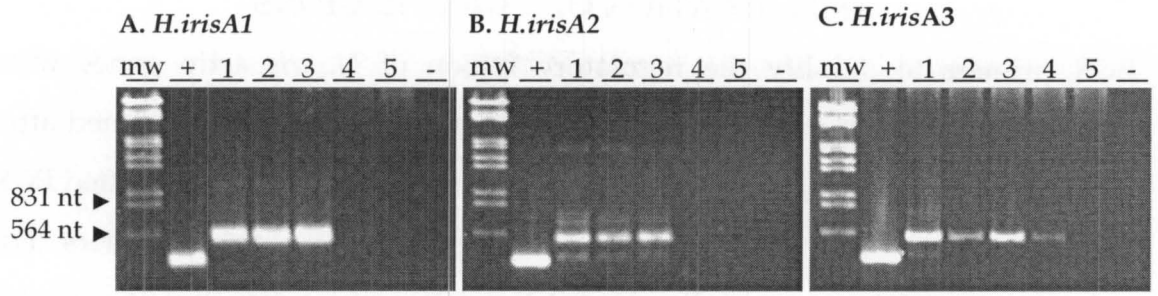


Figure 2.5. PCR amplification of *H. iris* actin genes with generic A4F and subtype-specific reverse primers. PCR products were analysed on 1.5% agarose/0.5 × TBE buffer gels. The molecular weight marker (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 62.2°C (401 nt). **A. *H.irisA1.*** A4F/A1717R1 products (546 nt) amplified at 59.5, 62.2, 64.8, 67.1 and 68.9°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A4F/A1717R1 at 50.9°C. **B. *H.irisA2.*** A4F/A1719R2 products (548 nt) amplified at 50.9, 52.3, 54.4, 56.8 and 59.5°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A4F/A1719R2 at 50.9°C. **C. *H.irisA3.*** A4F/A1721R3 products (550 nt) amplified at 56.8, 59.5, 62.2, 64.8 and 67.1°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A4F/A1721R3 at 50.9°C.

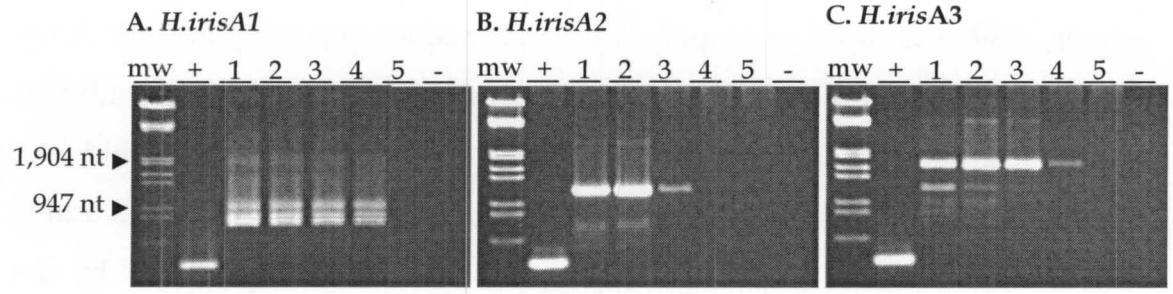


Figure 2.6. PCR amplification of *H. iris* actin genes with generic A989F(T3) and subtype-specific reverse primers. PCR products were analysed on 1.5% agarose/0.5 × TBE buffer gels. The molecular weight marker (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 62.2°C (401 nt). **A. *H.irisA1.*** A989F(T3)/A1717R1 products (757, 812, 865 and 884 nt) amplified at 62.2, 64.8, 67.1, 68.9 and 70.1°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A989F(T3)/A1717R1 at 50.9°C. **B. *H.irisA2.*** A989F(T3)/A1719R2 products (1,663 nt) amplified at 52.3, 54.4, 56.8, 59.5 and 62.2°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A989F(T3)/A1719R2 at 50.9°C. **C. *H.irisA3.*** A989F(T3)/A1721R3 products (1,164 nt) amplified at 59.5, 62.2, 64.8, 67.1 and 68.9°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A989F(T3)/A1721R3 at 50.9°C.

AMPLIFICATION OF 5' GENE REGIONS

PCR primers to amplify the regulatory region of *H. iris* actin genes were designed from the *H. rufescens* regulatory region sequence. A266F, named after its location in the *H. rufescens* sequence, was the only primer that amplified PCR products. A266F was used with actin subtype-specific reverse primers. For *H.irisA1*, A1717R1 was used. For *H.irisA2* and *H.irisA3*, intron-based subtype-specific primers (A2IntronR(T7) and A3IntronR(T7)) were used, as A1719R2 and A1721R3 would have amplified large products due to expansion in intron B. The Tm of A266F was 66°C; the Tm of A1717R1, A2IntronR(T7) and A3IntronR(T7) were 72, 61.5 and 68°C, respectively. For *H.irisA1*, *H.irisA2* and *H.irisA3* the predicted amplicons were 1,492, 834 and 1,023 nt in size, respectively. For *H.irisA1* a strong product, smaller than the expected size, was amplified between 46.8 and 50.3°C. Several smaller sized amplicons were also produced (approximately 580 and 800 nt) (Figure 2.7A). For *H.irisA2*, faint products were produced by A266F/A2IntronR(T7) at 46.8°C. These products were approximately 600 and 1,300 nt in size, not the predicted product of 834 nt (Figure 2.7B). For *H.irisA3*, A266F/A3IntronR(T7) did not produce amplicons at any annealing temperatures (Figure 2.7C).

As the *H.irisA1* A266F/A1717R1 product was smaller than expected and the PCR contained non-specific products, PCR using A266F and a generic primer, A5R, was used to amplify a *H.irisA1* sequencing template. As A266F did not amplify significant products with the *H.irisA2* and *H.irisA3* subtype-specific intron-based primers, the product of a A266F/A5R PCR should be a *H.irisA1* amplicon. As the A266F/A5R amplicon would include the intron region, amplicons from other actin subtypes would be distinguishable by size differences if they were present. A single product of the expected size, 976 nt, was amplified between 58.8 and 59.5°C (Figure 2.8A). This product was used as an *H.irisA1* template for sequencing.

To amplify the 5' coding regions of *H.irisA2* and *H.irisA3* for sequencing, PCR with primers that bound near or at the ATG start site were performed.

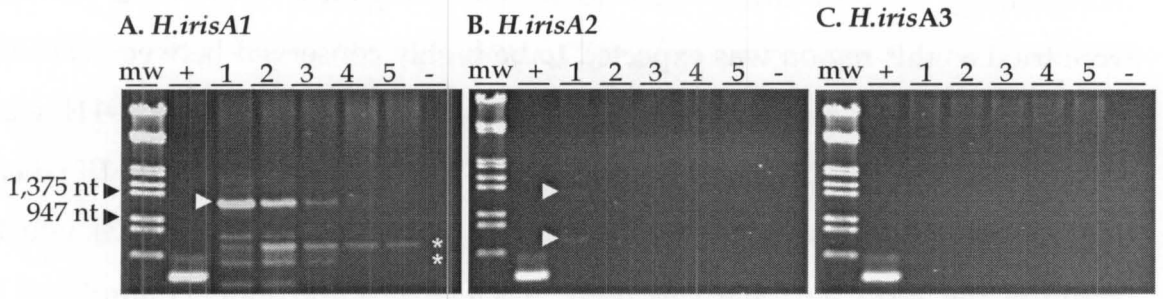


Figure 2.7. Amplification of *H. iris* actin gene regulatory regions with A266F and subtype-specific reverse primers. PCR products were analysed on 1% agarose/0.5 × TBE buffer gels. The molecular weight marker (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 60.8°C. **A. *H.irisA1.*** A266F/A1717R1 products amplified at 46.8, 50.3, 52.2, 55.4 and 58.1°C (lanes 1-5). A white arrowhead indicates the main product. Asterisks indicate non-specific products. The negative control (-) was ddH_2O amplified with A266F/A1717R1 at 46.8°C. **B. *H.irisA2.*** A266F/A2IntronR(T7) products amplified at 46.8, 50.3, 52.2, 55.4 and 58.1°C (lanes 1-5). White arrowheads indicate the amplified products. The negative control (-) was ddH_2O amplified with A266F/A2IntronR(T7) at 46.8°C. **C. *H.irisA3.*** A266F/A3IntronR(T7) products amplified at 46.8, 50.3, 52.2, 55.4 and 58.1°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A266F/A3IntronR(T7) at 46.8°C.

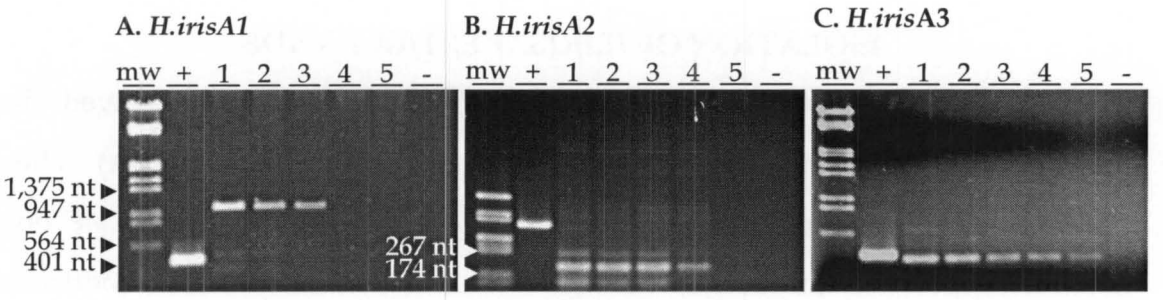


Figure 2.8. Amplification of 5′ gene region of *H. iris* genes. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 60.8°C (401 nt). **A. *H.irisA1.*** A266/A5R products (976 nt) amplified at 54.4, 56.8, 59.5, 62.2 and 64.8°C (lanes 1-5). The molecular weight marker (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. The negative control (-) was ddH_2O amplified with A266F/A5R PCR primers at 54.4°C. PCR products were analysed on a 1% agarose/0.5 × TBE gel. **B. *H.irisA2.*** A911/A2IntronR(T7) products (187 nt) amplified at 34.1, 42.0, 44.7, 47.1, and 50.0°C (lanes 1-5). The molecular weight marker (mw) was pBluescript M13+ genomic DNA digested with *Hae* III. The negative control was ddH_2O amplified with A911/A2IntronR(T7) at 34.1°C. PCR products were analysed on a 2% agarose/0.5 × TBE gel. **C. *H.irisA3.*** ATGF(T3)/A3IntronR(T7) products (369 nt) amplified at 48.4, 50.4, 52.5, 54.7 and 56.8°C (lanes 1-5). The molecular weight (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. The negative control (-) was ddH_2O amplified with ATG/A3IntronR(T7) at 64.8°C. PCR products were analysed on a 1.5% agarose/0.5 × TBE gel.

As regulatory region primers failed to amplify templates, ATG based primers were tried as this region was expected to be highly conserved between *Haliotis* actins. A911F binds just upstream of and across the ATG start site. A911F was used to amplify a 187 nt *H.irisA2* product at 47.1°C (Figure 2.8B). A911F failed to amplify a product with A3IntronR(T7), so a second primer, ATGF(T3), which bound at the ATG start site was used. ATGF(T3)/A3IntronR(T7) amplified a single *H.irisA3* product of 369 nt between 52.5-56.8°C (Figure 2.8C).

In an attempt to isolate the regulatory region of *H.irisA2* and *H.irisA3*, a further actin regulatory region primer was designed from the *H.irisA1* sequence (determined from sequencing the A266F/A5R amplicon). This primer was A623F, named after its location in the *H. rufescens* sequence. Amplification with A623F and subtype-specific primers were done between annealing temperatures of 46.8 and 58.1°C. For *H.irisA1* (A623F/A1717R1), a main product of the expected size (1,228 nt) and two non-specific products (approximately 1,850 and 2,500 nt in size) were produced at all annealing temperatures (Figure 2.9A). A623F failed to amplify *H.irisA2* or *H.irisA3* regulatory regions when used with A2IntronR(T7) or A3IntronR(T7) (Figure 2.8B and Figure 2.8C).

ISOLATION OF *H.IRISA1* EXTRA BANDS

Electrophoretic analysis of A989F(T3)/A1717R1 PCR products showed that multiple products of similar size had been amplified (Figure 2.6A). These products were analysed further by electrophoretic separation on a 4% non-denaturing polyacrylamide gel. Polyacrylamide gel electrophoresis confirmed that four products were present, sized 884, 865, 812 and 757 nt (Figure 2.10A). These products were excised from the polyacrylamide gel and extracted by the crush n' soak method. The purified DNA was used as a template for PCR (Figure 2.10B). Amplification of DNA eluted from each band amplified a single product corresponding in size to the four bands of the original A989F(T3)/A1717R1 PCR.

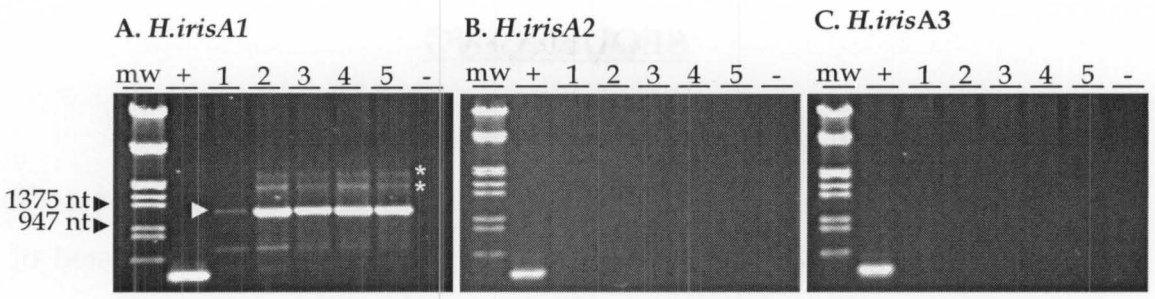


Figure 2.9. Amplification of *H. iris* actin gene regulatory regions with *H. iris* A623F and subtype-specific reverse primers. PCR products were analysed on 1% agarose/0.5 × TBE buffer gels. The molecular weight marker (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. The positive control (+) was *H. iris* genomic DNA amplified with 18SF/18SR at 60.8°C. **A. *H.irisA1*.** A623F/A1717R1 (1,228 nt) products amplified at 46.8, 50.3, 52.2, 55.4 and 58.1°C (lanes 1-5). The expected product is indicated with a white arrowhead. Non-specific products are indicated by white asterisks. The negative control (-) was ddH_2O amplified with A623F/A1717R1 at 46.8°C. **B. *H.irisA2*.** A623F/A2IntronR(T7) products amplified at 46.8, 50.3, 52.2, 55.4 and 58.1°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A623F/A2IntronR(T7) at 46.8°C. **C. *H.irisA3*.** A623F/A3IntronR(T7) products amplified at 46.8, 50.3, 52.2, 55.4 and 58.1°C (lanes 1-5). The negative control (-) was ddH_2O amplified with A266F/A3IntronR(T7) at 46.8°C.

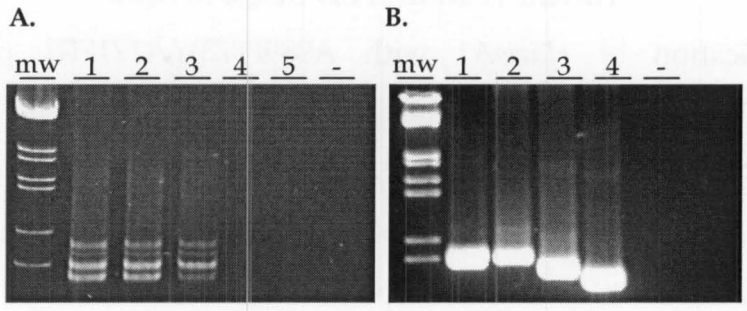


Figure 2.10. Isolation of *H.irisA1* A989F(T3)/A1717R1 PCR products. The molecular weight marker (mw) was λ genomic DNA digested with *Eco*R I and *Hin*d III. **A. A989F(T3)/A1717R1 PCR products** were resolved by electrophoresis on a 4% non-denaturing polyacrylamide/ 1 × TBE gel. 1-5 show products (884, 865, 812 and 757 nt) amplified by A989F(T3)/A1717R1 at 62.2, 64.8, 67.1, 68.9 and 70.1°C. The negative control (-) was ddH_2O amplified with A989F(T3)/A1717R1 at 62.2°C. **B. Isolated and purified *H.irisA1* PCR products**. PCR products were analysed on a 1.5% agarose/0.5 × TBE gel. 1-4 show each of the reamplified *H.irisA1* PCR products, in order of decreasing size. The negative control (-) was ddH_2O amplified with A989F(T3)/A1717R1 at 62.2°C.

SEQUENCING

H.IRISA1, H.IRISA2 AND H.IRISA3 SEQUENCES

The sequence of *H.irisA1*, including the 5' untranslated regions and coding regions, was determined. This sequence was 1,960 nt long and consisted of a 5' untranslated region of 658 nt, an intron of 174 nt and a coding sequence of 1,128 nt (Figure 2.11 and Figure 2.12). The untranslated region contained a TATA box between nt -100 and -95. This regulatory motif was 29 nt upstream of an untranslated exon (exon 1) of 65 nt (at -65 to -1 nt). The untranslated region also included CAAT regulatory boxes at nucleotide positions -234 to -229, -226 to -221 and -126 to -121. Exon 2 (123 nt) and exon 3 (1,005 nt) were separated by intron B (174 nt) and encoded a predicted protein of 375 aa.

For *H.irisA2* and *H.irisA3*, the sequences from the start codon to the stop codon were determined. In both genes, exon 2 was 123 nt and exon 3 was 1,005 nt, encoding predicted proteins of 375 aa. The introns of *H.irisA2* and *H.irisA3* differed in size, causing the PCR product sizes to vary. The *H.irisA2* sequence was 2,026 nt and included a 1,078 nt intron (intron B) (Figure 2.11 and Figure 2.13). The *H.irisA3* sequence was 1,709 nt and included a 581 nt intron (intron B) (Figure 2.11 and Figure 2.14).

H.IRISA1 RELATED SEQUENCES

PCR amplification of *H.irisA1* with A989F(T3)/A1717R1 produced four amplicons. These were excised, purified and sequenced. The smallest sized product (757 nt) was identified as *H.irisA1*. The three other products were named *H.irisA1a*, *H.irisA1b* and *H.irisA1c* (on basis of descending size) as they had been amplified with an *H.irisA1* specific primer.

H.irisA1a, *H.irisA1b* and *H.irisA1c* were composed of partial sequence data for exons 2 (42 nt) and 3 (541 nt) and the entire intron B region (Figure 2.11). The sequences obtained for the three *H.irisA1*-type PCR products predicted proteins of 194 aa. *H.irisA1a* is shown in Figure 2.15. The *H.irisA1a* sequence was 884 nt long and contained an intron B sequence of 301 nt.

H.irisA1b is shown in Figure 2.16. The *H.irisA1b* sequence was 865 nt long and contained an intron B sequence of 282 nt. *H.irisA1c* is shown in Figure 2.17. The *H.irisA1c* sequence was 812 nt long and contained an intron B sequence of 229 nt.

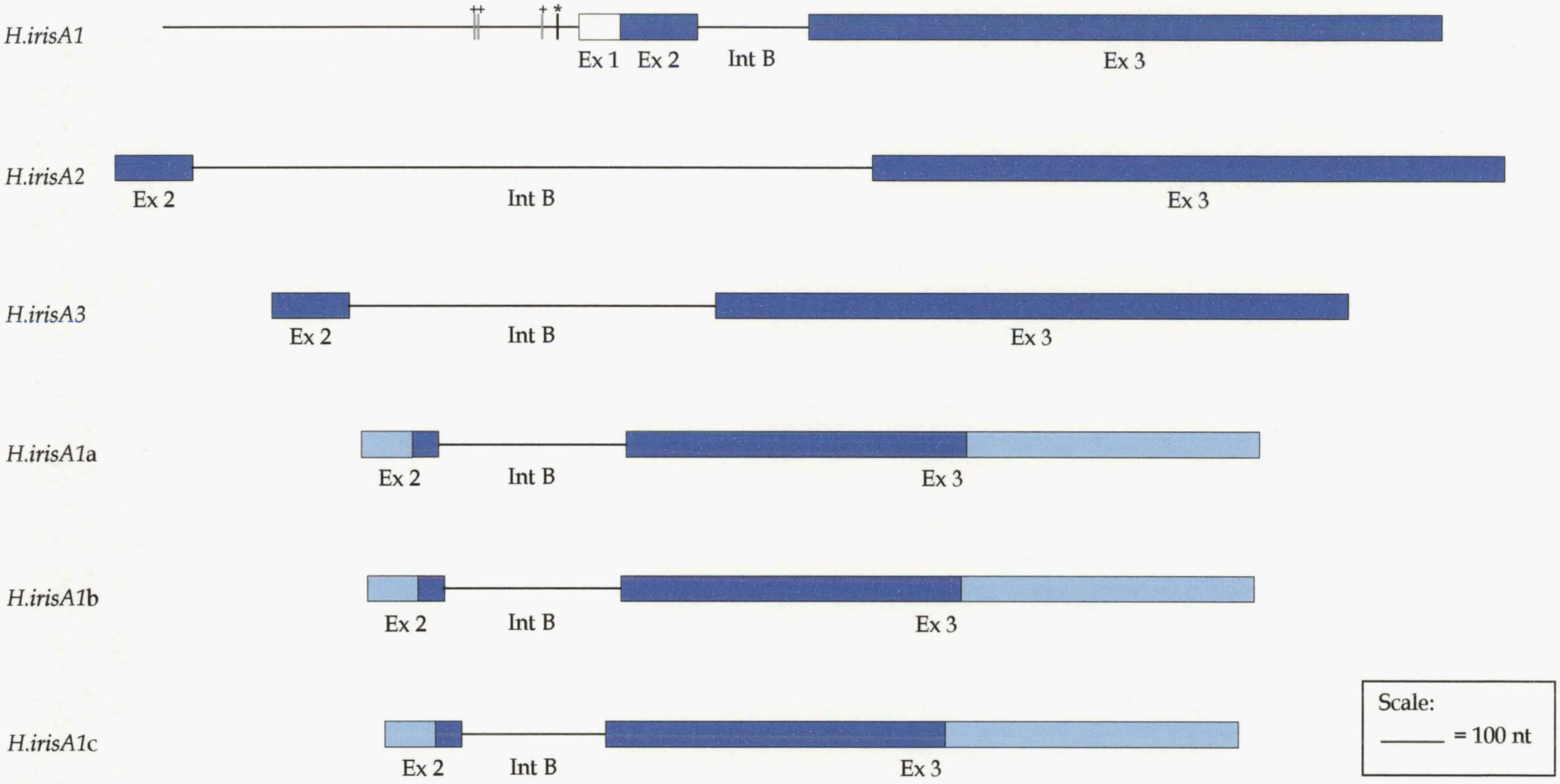


Figure 2.11. Structure of putative *H. iris* actin genes. Dark blue bars represent sequenced exons, white bars indicate untranslated exons, pale blue bars represent predicted exons (unsequenced regions). Lines represent non-coding regions which have been sequenced. The asterisk (*) indicates a TATA box, the pluses (+) indicate CAAT boxes. The genes are shown to scale.

Figure 2.12. *H.irisA1* nucleotide sequence and predicted amino acid sequence. Lower case letters represent untranslated regions, upper case letters represent coding nucleotides. The TATA box, CAAT boxes, initiation and stop codons are underlined. Primer binding sites are shown in bold (for details see Table 2.2). The predicted amino acid sequence is shown below the nucleotide sequence. The nucleotide and amino acid positions are given to the right.

```
aacaacatagcaactttgcatgcatggtgagaacaaacattggcagggaattccatttgt  [-599]
atattcattagatattttactttactgtatcgaatccgtattacatactgcatttctttt  [-539]
ctgaattatggttgtaacattacaacattactaaggaggcgtaagcgttgtatgctgttt  [-479]
agagtctatgcatttccaaataattctgaaatatgcatttaacttatttttattacgga  [-419]
acaagcatcatttgcatgcatacatactagtcaactcaaggagcgcttcccattctccgg  [-359]
aatcatatcatgtggcgttcactgcttgtgggagcgtttgggggacgggtctggtacat  [-299]
cctattaggggaaaagagttcagtgagctggatgcagttgacccctgactgtgacgcatt  [-239]
cttccacatgtacacatgaaaggctcttatgcaatataatacattaggagggtatagtta  [-179]
tcattacctacaattacattggttacgtttcaaytattcttaaacacaatctgattggtc  [-119]
ttctactgatgggggcagtataaaagacgcatggatgctracctcgatagtgcATTCGCT  [-59]
TTCAGTCTTCATTTGACTCGCATCTTGTTTCCTGTTTCYCACACAACTACCTACAGTCAT  [2]
                                                          M    [1]
GGATGATGATGTTGCTGCATTGGTCTGTGACAACGGCTCCGGCATGTGCAAAGCCGGTTT  [62]
  D  D  D  V  A  A  L  V  C  D  N  G  S  G  M  C  K  A  G  F  [21]
CGCCGGTGACGACGCTCCCAGAGCTGTCTTCCCCTCCATCGTCGGCCGTCCCAGACATCA  [122]
  A  G  D  D  A  P  R  A  V  F  P  S  I  V  G  R  P  R  H  Q  [41]
Ggtaacaccatttattgaacatagaagttcaagatttgaagtattagtttttcgggtgt  [182]
tttggaagacattataatccgcttgagtagtgtaacctggtgtcgttaaaattgaatctg  [242]
tcggtagatcgcagtcgaatatacgtagataatgacaaatgatttcgtgtttcagGGCGT  [302]
  G  V                                                        [43]
GATGGTCGGTATGGGTCAGAAAGACAGCTACGTCGGTGACGAGGCTCAGTCCAAGAGAGG  [362]
  M  V  G  M  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K  R  G  [63]
TATCCTCACCCTCAAGTACCCCATCGAGCACGGTATCGTCACCAACTGGGACGACATGGA  [422]
  I  L  T  L  K  Y  P  I  E  H  G  I  V  T  N  W  D  D  M  E  [83]
GAAGATCTGGCATCACACCTTCTACAACGAGCTCCGCGTTGCACCAGAGGAACACCCCGT  [482]
  K  I  W  H  H  T  F  Y  N  E  L  R  V  A  P  E  E  H  P  V  [103]
CCTCTTGACAGAGGCTCCCCTCAACCCTAAGGCCAACCGTGAAAAGATGACCCAGATCAT  [542]
  L  L  T  E  A  P  L  N  P  K  A  N  R  E  K  M  T  Q  I  M  [123]
GTTCGAGACCTTCAACTCTCCAGCCATGTATGTGGCCATCCAGGCTGTTCTGTCTCTCTA  [602]
  F  E  T  F  N  S  P  A  M  Y  V  A  I  Q  A  V  L  S  L  Y  [143]
CGCTTCTGGTCGTACCACGGGTATTGTTTTGGATTCTGGTGATGGTGTCACCCACACTGT  [662]
  A  S  G  R  T  T  G  I  V  L  D  S  G  D  G  V  T  H  T  V  [163]
CCCCATCTATGAAGGTTACGCCCTTCCCCACGCCATCATGAGGTTGGATCTTGCCGGCCG  [722]
  P  I  Y  E  G  Y  A  L  P  H  A  I  M  R  L  D  L  A  G  R  [183]
TGACCTGACTGATTACCTCATGAAGATCCTCACTGAGCGTGGCTACTCCTTCACCACCAC  [782]
  D  L  T  D  Y  L  M  K  I  L  T  E  R  G  Y  S  F  T  T  T  [203]
CGCTGAGAGAGAAATCGTCAGAGACATCAAGGAGAAGCTCTGCTACGTCGCTCTCGACTT  [842]
  A  E  R  E  I  V  R  D  I  K  E  K  L  C  Y  V  A  L  D  F  [223]
CGAGCAAGAGATGTCTACAGCTGCTTCCTCTTCCTCCCTGGAGAAGAGCTACGAGTTGCC  [902]
  E  Q  E  M  S  T  A  A  S  S  S  S  L  E  K  S  Y  E  L  P  [243]
CGATGGTCAGGTCATCACCATTGGTAACGAGAGGTTCCGTTGCCCCGAATCCCTCTTCCA  [962]
  D  G  Q  V  I  T  I  G  N  E  R  F  R  C  P  E  S  L  F  Q  [263]
ACCATCCTTCTTGGGAATGGAATCTGCTGGTATACACGAGACCACATACAACTCTATCAT  [1022]
  P  S  F  L  G  M  E  S  A  G  I  H  E  T  T  Y  N  S  I  M  [283]
GAAGTGCGATGTTGATATCCGTAAGGACTTGTACGCCAACACTGTTCTCTCAGGAGGTAC  [1082]
  K  C  D  V  D  I  R  K  D  L  Y  A  N  T  V  L  S  G  G  T  [303]
CACCATGTTCCCAGGTATCGCCGACAGAATGCAGAAGGAGATCACAGCCCTTGCCCCCAG  [1142]
  T  M  F  P  G  I  A  D  R  M  Q  K  E  I  T  A  L  A  P  S  [323]
CACGATGAAGATCAAAATCATCGCTCCCCCAGAGAGGAAATACTCCGTCTGGATCGGTGG  [1202]
  T  M  K  I  K  I  I  A  P  P  E  R  K  Y  S  V  W  I  G  G  [343]
CTCCATCCTTGCCTCTCTGTCCACCTTCCAACAGATGTGGATCAGCAAACAAGAGTACGA  [1262]
  S  I  L  A  S  L  S  T  F  Q  Q  M  W  I  S  K  Q  E  Y  D  [363]
CGAATCTGGTCCATCCATTGTCCACCGCAAATGCTTCTAA                      [1302]
  E  S  G  P  S  I  V  H  R  K  C  F  *                       [375]
```

```
ATGGATGATGATGTTGCTGCGTTGGTTATCGACAACGGCTCCGGGATGTGTAAGGCCGGG   [60]
 M  D  D  D  V  A  A  L  V  I  D  N  G  S  G  M  C  K  A  G     [20]
TTCGTCGGAGACGACCCACCCAGAGCTGTCTTCTCCTCAATCGTCGGCCGCCCAAGACAT   [120]
 F  V  G  D  D  P  P  R  A  V  F  S  S  I  V  G  R  P  R  H     [40]
CAAgtaagttcttcagcttgtaagaaggttataataatatattcgaccaatcagcatgaa   [180]
 Q                                                              [41]
aggtaaatactatgtctgccattacatgaaagtgattgaatggcttaatatgtccccaaa   [240]
tatattaatacacctccgtttattatcacatggttagaataggtcttcagcaaccaatgt   [300]
tgttgtagaacgtgactatgcttgtcgtaagaggcgaataacgggatcgggtggtgaggc   [360]
tcactgacttgattgatgcatgtcatcgtatcccaaatgcgtagattgatgcgcatgata   [420]
ttaatcactatattgtctggtccggactcgattatttagagaccgtcgtcatatagctga   [480]
aatattgctgagtgcggcgttaaacaacaaacacacaaatattgctatatcctgtcgtgc   [540]
agatgtggttgtacagtaacgctaagtttctgaaaatttataattttgataaaaacgagc   [600]
aactatgaaactattgaaaaaaagccccaatgagaagagagattttgaacctgaactcaa   [660]
cgacagtgcacagtcaagaaagcgtttaatttccaataaccattaagcattaaaattgta   [720]
tagttatagaacaggaatattacaacactgaccatataaactcatattgcatagcatatt   [780]
cttagtttagaagtctttgtagggaatagtaatgatcatacaaccaggggcttctgtgaa   [840]
ctgcaataccatgaaactcgattgaaaataattatataagatatatgtacttttttgta   [900]
aaaaaggaattagttgttaaatttagtgatgtaatcatagggctaccccacatgttaaag   [960]
ttttttaaaatgtgtataaaatcgtgtgcgaaatgcgtgttgataatttctttgtgcac   [1020]
ggaacgttttgcgactgaaagcaagcctgtgtaaattattgacattgatagtaattcgga   [1080]
attgaaatcaactttcctatctgtattacatataaatcatatattaaagttgctcttata   [1140]
gttgtgaaatgaaaatatgctttgtaattatgtgtaatgttaaaactttctttgcttcca   [1200]
gGGTGTGATGGTCGGTATGGGACAGAAGGACAGCTACGTCGGTGACGAAGCTCAATCAAA   [1260]
  G  V  M  V  G  M  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K   [61]
GCGAGGTATTCTCACTCTCAAGTACCCCATTGAGCACGGTATCGTCACCAACTGGGACGA   [1320]
  R  G  I  L  T  L  K  Y  P  I  E  H  G  I  V  T  N  W  D  D   [81]
CATGGAGAAGATCTGGCATCACACATTCTACAACGAGCTCCGTGTTGCTCCAGAGGAGCA   [1380]
  M  E  K  I  W  H  H  T  F  Y  N  E  L  R  V  A  P  E  E  H   [101]
CCCTGTCCTTCTGACTGAGGCTCCACTCAACCCCAAGGCCAACCGTGAAAAGATGACCCA   [1440]
  P  V  L  L  T  E  A  P  L  N  P  K  A  N  R  E  K  M  T  Q   [121]
GATCATGTTTGAGACCTTCAACTCTCCAGCTATGTATGTGGCCATTCAGGCTGTTCTGTC   [1500]
  I  M  F  E  T  F  N  S  P  A  M  Y  V  A  I  Q  A  V  L  S   [141]
TCTGTACGCCTCTGGTCGTACGACGGGTATTGTTCTTGACTCTGGTGATGGTGTGTCTCA   [1560]
  L  Y  A  S  G  R  T  T  G  I  V  L  D  S  G  D  G  V  S  H   [161]
CACTGTTCCTATCTACGAGGGTTATGCCCTTCCCCACGCCATCATGAGACTGGACCTTGC   [1620]
  T  V  P  I  Y  E  G  Y  A  L  P  H  A  I  M  R  L  D  L  A   [181]
TGGGCGTGACCTGACAGACTACCTCATGAAGATCCTGACCGAGCGTGGTTACTCCTTCAC   [1680]
  G  R  D  L  T  D  Y  L  M  K  I  L  T  E  R  G  Y  S  F  T   [201]
CACCACTGCCGAGCGAGAAATCGTCAGGGACATCAAAGAAAAATTAGCTTACATTGCTTT   [1740]
  T  T  A  E  R  E  I  V  R  D  I  K  E  K  L  A  Y  I  A  L   [221]
AGACTTTGAACAGGAAATGCAGACTGCTGCATCTTCCTCTTCATTGGAAAAGAGTTACGA   [1800]
  D  F  E  Q  E  M  Q  T  A  A  S  S  S  S  L  E  K  S  Y  E   [241]
GCTTCCCGACGGTCAGGTCATCACTATTGGTAACGAGAGGTTCCGTTGTCCAGAGGCTCT   [1860]
  L  P  D  G  Q  V  I  T  I  G  N  E  R  F  R  C  P  E  A  L   [261]
CTTCCAGCCATCTTTCTTGGGTATGGAATCTGCTGGTATCCACGAAACAACATACAACTC   [1920]
  F  Q  P  S  F  L  G  M  E  S  A  G  I  H  E  T  T  Y  N  S   [281]
CATCATGAAGTGTGATGTTGATATTCGTAAAGACTTGTATGCCAACACTGTTCTCTCGGG   [1980]
  I  M  K  C  D  V  D  I  R  K  D  L  Y  A  N  T  V  L  S  G   [301]
AGGTACTACCATGTTCCCTGGTATCGCCGACAGAATGCAGAAGGAGATCACAGCCCTTGC   [2040]
  G  T  T  M  F  P  G  I  A  D  R  M  Q  K  E  I  T  A  L  A   [321]
TCCAGCAACAATGAAGATCAAGATCATCGCTCCTCCAGAGAGGAAATACTCCGTCTGGAT   [2100]
  P  A  T  M  K  I  K  I  I  A  P  P  E  R  K  Y  S  V  W  I   [341]
CGGTGGCTCCATCCTTGCCTCTCTGTCCACCTTCCAGCAGATGTGGATCAGCAAGCAGGA   [2160]
  G  G  S  I  L  A  S  L  S  T  F  Q  Q  M  W  I  S  K  Q  E   [361]
GTACGACGAGTCCGGCCCATCCATTGTCCACCGCAAATGCTTCTAA                 [2206]
  Y  D  E  S  G  P  S  I  V  H  R  K  C  F  *                  [375]
```

Figure 2.13. *H.irisA2* nucleotide sequence and predicted amino acid sequence. Lower case letters represent untranslated regions, upper case letters represent coding nucleotides. Primer binding sites are shown in bold (for details see Table 2.2). The initiation and stop codons are underlined. The predicted amino acid sequence is shown below the nucleotide sequence. The nucleotide and amino acid positions are given to the right.

ATGGATGATGATGTTGCTGCGTTGGTTATCGACAACGGCTCCGGCATGTGTAAGGCCGGG [60]
M D D D V A A L V I D N G S G M C K A G [20]
TTCGCCGGTGACGACGCCCCG**AGAGCTGTCTTCCCCTCCAT**CGTCGGTCGTCCCCGTCAT [120]
F A G D D A P R A V F P S I V G R P R H [40]
CAGgtaattcacttctgtcatcaatactactcactaatcaatggttcatcttaactggtc [180]
Q [41]
atgatatgtctatcatttagctattgttcaatatgataaattgctgtatatttttgtcac [240]
gtatcatttatgaaaacaatttaaatcgacttatggatatgcatgtgtgtttgtatatta [300]
tttcaaaatatgatctaataggggatatttgatgagtatatatcgtaat**atcgaactacg** [360]
aagagagtgtgaaaaatagtttcaagacaaacatgttctgtgcataaacgtttaaggttg [420]
cattaatttattggacatgtattttatgtatcttcataccaatagtatagaaataagcaa [480]
tgaagacgctttagtcgccctgccatagaatgattttgttaatcggtgacagaaatttca [540]
aatgatatctaccaattgctgtacaaacatggttacgataatgatgcggtattttgtgac [600]
gtcagtagtccgatggtgaaaatatgcatatgggtggtaatagttaatattttaaatgtc [660]
acagctaattattcatatatcactacatttttttttctt**tccagGGTGTGATGGTCGGT**A [720]
G V M V G [46]
TGGGACAGA**AGGACAGCTACGTCGGTGAC**GAGGCTCAGTCCAAGAGAGGTATCCTCACCC [780]
M G Q K D S Y V G D E A Q S K R G I L T [66]
TCAAGTACCCCATTGAGCACGGTATCGTCACCAACTGGGACGACATGGAGAAGATCTGGC [840]
L K Y P I E H G I V T N W D D M E K I W [86]
ATCACACCTTCTACAACGAGCTCCGTGTGGCTCCAGAGGAACATCCCGTCCTCCTGACCG [900]
H H T F Y N E L R V A P E E H P V L L T [106]
AGGCTCCCCTTAACCCCAAGGCCAACCGTGAAAAGATGACCCAGATCATGTTTGAGACCT [960]
E A P L N P K A N R E K M T Q I M F E T [126]
TCAACTCTCCAGCTATGTATGTGGCCATCCAGGCTGTTCTGTCTCTGTACGCCTCTGGTC [1020]
F N S P A M Y V A I Q A V L S L Y A S G [146]
GTACCACGGGTATTG**TTCTTG*ACTCTGGTGATGGT*GT**CACCCACACTGTCCCCATCTATG [1080]
R T T G I V L D S G D G V T H T V P I Y [166]
AAGGTTACGCCCTTCCCCACGCTATTCTTAGGTTGGACCTGGCTGGTCGTGATCTCACAG [1140]
E G Y A L P H A I L R L D L A G R D L T [186]
ACTACATGATGAAGATCCTGACTGAGCGTGGTTACTCCTTCACCACCACTGCCGAACGAG [1200]
D Y M M K I L T E R G Y S F T T T A E R [206]
AAATCGTCAGGGA**CATCAAGGAGAAA*CTGGCCT*ATGTGGCTCTTG**ACTTCGAACAGGAAA [1260]
E I V R D I K E K L A Y V A L D F E Q E [226]
TGCAGACTGCTGCCTCTTCCTCTTCATTGGAGAAGAGCTACGAGCTTCCCGACGGTCAGG [1320]
M Q T A A S S S S L E K S Y E L P D G Q [246]
TCATCACCATTGGTAACGAGAGGTTTCGTTGTCCAGAGGCTCTTTTCCAGCCTTCCTTCT [1380]
V I T I G N E R F R C P E A L F Q P S F [266]
TGGGTATGGAATCTGCTGGTATC**CACGAAACAACATACAAC**TCCATCATGAAATGTGATG [1440]
L G M E S A G I H E T T Y N S I M K C D [286]
TTGATATTCGTAAAGACTTATATGCCAACACTGTTCTCTCAGGAGGTACCACCATGTTCC [1500]
V D I R K D L Y A N T V L S G G T T M F [306]
CTGGTATCGCCGACAGAATGCAGAAGGAGATCACAGCCCTTGCTCCAGCAACAATGAAGA [1560]
P G I A D R M Q K E I T A L A P A T M K [326]
TCAAGATCATCGCACCCCCAGAGAGGAAATACTCGGTCTGGATAGGCGGCTCCATCCTTG [1620]
I K I I A P P E R K Y S V W I G G S I L [346]
CCTCTCTGTCCA**CCTTCCAACAGATGTGGAT**CAGCAAGCAGGAGTACGATGAGTCCGGCC [1680]
A S L S T F Q Q M W I S K Q E Y D E S G [366]
CATCCATT**GTCCACCGCAAATGCTTCTAA** [1709]
P S I V H R K C F * [375]

Figure 2.14. *H.irisA3* nucleotide sequence and predicted amino acid sequence. Lower case letters represent untranslated regions, upper case letters represent coding nucleotides. Primer binding sites are shown in bold (for details see Table 2.2). The initiation and stop codons are underlined. The predicted amino acid sequence is shown below the nucleotide sequence. The nucleotide and amino acid positions are given to the right.

```
AGAGCTGTCTTCCCCTCCATCGTCGGCCGTCCCAGACATCAGgtaatacagttgttttgg   [60]
 R  A  V  F  P  S  I  V  G  R  P  R  H  Q                      [14]
ttatgtattgaaacgtaaatcttgtaacttgatttcacatgtagaagtatctgttggatt   [120]
attcatctacatatgtatatatgacagaatatataacagaatgtctaaatgaataatatt   [180]
attactaaaaactgaggagcaccagaaactatacccctcagttgattttatcctatgttt   [240]
gatggaattaagtgtcagaagatcagagtaaattcttttcaaactgttttatgccttttt   [300]
taaacatacttacgtgtaatgacgtaatgaatatatgttttagGGAGTGATGGTCGGTAT   [360]
  G  V  M  V  G  M                                            [20]
GGGTCAGAAAGACAGCTACGTCGGTGACGAGGCTCAGTCCAAGAGAGGTATCCTCACCCT   [420]
  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K  R  G  I  L  T  L   [40]
TAAGTACCCCATCGAGCACGGTATCGTCACCAACTGGGACGACATGGAGAAGATCTGGCA   [480]
  K  Y  P  I  E  H  G  I  V  T  N  W  D  D  M  E  K  I  W  H   [60]
TCACACCTTCTACAACGAGCTCCGAGTGGCTCCAGAGGAGCACCCCGTCCTCCTGACAGA   [540]
  H  T  F  Y  N  E  L  R  V  A  P  E  E  H  P  V  L  L  T  E   [80]
GGCTCCCCTCAACCCCAAAGCCAACCGTGAAAAGATGACCCAGATCATGTTCGAGACCTT   [600]
  A  P  L  N  P  K  A  N  R  E  K  M  T  Q  I  M  F  E  T  F   [100]
CAACTCTCCAGCTATGTATGTGGCAATCCAGGCTGTTCTGTCTCTGTACGCTTCTGGTCG   [660]
  N  S  P  A  M  Y  V  A  I  Q  A  V  L  S  L  Y  A  S  G  R   [120]
TACCACGGGTATTGTTTTGGACTCTGGTGATGGTGTCACCCACACTGTACCCATCTATGA   [720]
  T  T  G  I  V  L  D  S  G  D  G  V  T  H  T  V  P  I  Y  E   [140]
AGGTTACGCCCTTCCCCACGCCATCATGAGGTTGGATCTTGCTGGCCGTGACCTGACTGA   [780]
  G  Y  A  L  P  H  A  I  M  R  L  D  L  A  G  R  D  L  T  D   [160]
TTACCCCATGAAGATCCTCACTGAGCGTGGCTACTCCTTCACCACCACCGCTGAGAGAGA   [840]
  Y  P  M  K  I  L  T  E  R  G  Y  S  F  T  T  T  A  E  R  E   [180]
AATCGTCAGAGACATCAAGGAGAAGCTGCTGTACATCGCCCTTG                   [884]
  I  V  R  D  I  K  E  K  L  L  Y  I  A  L                     [194]
```

Figure 2.15. *H.irisA1a* partial nucleotide sequence and predicted amino acid sequence. Lower case letters represent untranslated regions, upper case letters represent coding nucleotides. Primer binding sites are shown in bold (for details see Table 2.2). The predicted amino acid sequence is shown below the nucleotide sequence. The nucleotide and amino acid positions are given to the right.

```
AGAGCTGTCTTCCCCTCCATCGTCGGTCGTCCTAGACATCAGgtaagaaatggtgatgtg   [60]
 R  A  V  F  P  S  I  V  G  R  P  R  H  Q                      [14]
acatcagataacacttgcatttaaatatcaacatcagcatggtgttgaatcatatatcag   [120]
acgtattgtttctgtgagtggaattttggagttggtggcatgagtgagaaaagttcttga   [180]
ataaaaatggtctttcttacaattaaggagacgttccgacgaagattctgtcatctttat   [240]
caagggtatatataacccccaccattgtatttcatctttcactgaatcattttactaata   [300]
taaccatgttttgcttctttcaagGGTGTGATGGTCGGTATGGGTCAGAAAGACAGCTAC   [360]
 G  V  M  V  G  M  G  Q  K  D  S  Y                            [26]
GTCGGTGATGAGGCTCAGTCCAAAAGAGGTATTCTCACCCTCAAGTACCCCATCGAGCAC   [420]
 V  G  D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P  I  E  H    [46]
GGTATCGTCACCAACTGGGACGACATGGAGAAGATCTGGCATCACACCTTCTACAACGAG   [480]
 G  I  V  T  N  W  D  D  M  E  K  I  W  H  H  T  F  Y  N  E    [66]
CTCCGAGTTGCACCAGAGGAGCATCCCGTCCTCCTGACAGAGGCTCCCCTCAACCCCAAG   [540]
 L  R  V  A  P  E  E  H  P  V  L  L  T  E  A  P  L  N  P  K    [86]
GCCAACCGTGAAAAGATGACCCAGATCATGTTCGAGACCTTCAACTGTCCAGCTACGTAT   [600]
 A  N  R  E  K  M  T  Q  I  M  F  E  T  F  N  C  P  A  T  Y    [106]
GTGGCCATCCAGGCTGTTCTGTCTCTGTACGCTTCTGGTCGTACCACGGGTATTGTTTTG   [660]
 V  A  I  Q  A  V  L  S  L  Y  A  S  G  R  T  T  G  I  V  L    [126]
GACTCTGGTGATGGTGTCACACACACTGTCCCCATCTATGAAGGTTACGCCCTTCCCCAC   [720]
 D  S  G  D  G  V  T  H  T  V  P  I  Y  E  G  Y  A  L  P  H    [146]
GCCATCATGAGATTGGATCTTGCCGGCCGTGACCTGACTGATTACCTCATGAAGATCCTC   [780]
 A  I  M  R  L  D  L  A  G  R  D  L  T  D  Y  L  M  K  I  L    [166]
ACTGAGCGCGGTTACTCCTTCACCACCACTGCTGAGAGAGAAATCGTCAGAGACATCAAG   [840]
 T  E  R  G  Y  S  F  T  T  T  A  E  R  E  I  V  R  D  I  K    [186]
GAGAAGCTGCTGTACATCGCCCTTG                                      [865]
 E  K  L  L  Y  I  A  L                                        [194]
```

Figure 2.16. *H.irisA1b* partial nucleotide sequence and predicted amino acid sequence. Legend as for Figure 2.15.

```
AGAGCTGTCTTCCCCTCCATCGTCGGCCGTCTTAGACACCAGgtaaaataacttttttta   [60]
 R  A  V  F  P  S  I  V  G  R  L  R  H  Q                      [14]
aacaaaaaccaacggctgagttactgtctgcaaatagtgtaatttcacgttctatcagaa   [120]
cattattatgggtgtaactggttagaaaatggagttgattaagcttgagcaatggataat   [180]
aaaagggcccctcttgtgaaaaaaatagatattccttttatgtaacatatttagtagatc   [240]
atagtttcttaagattatgttttctttatagGGCGTGATGGTCGGTATGGGTCAGAAAGA   [300]
  G  V  M  V  G  M  G  Q  K  D                                 [24]
CAGCTATGTCGGTGACGAGGCTCAGTCCAAGAGAGGTATCCTCACCCTCAAGTACCCCAT   [360]
  S  Y  V  G  D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P  I  [44]
CGAACACGGTATCGTCACCAACTGGGACGACATGGAGAAGATCTGGCATCACACCTTCTA   [420]
  E  H  G  I  V  T  N  W  D  D  M  E  K  I  W  H  H  T  F  Y  [64]
CAACGAGCTCCGAGTTGCACCAGAGGAGCATCCCGTCCTCCTGACAGAGGCTCCTCTCGA   [480]
  N  E  L  R  V  A  P  E  E  H  P  V  L  L  T  E  A  P  L  D  [84]
CCCCAAGGCCAACCGAGAAAAGATGACACAGATCATGTTCGAAACCTTCAACTCTCCAGC   [540]
  P  K  A  N  R  E  K  M  T  Q  I  M  F  E  T  F  N  S  P  A  [104]
TATGTATGTGGCCATCCAGGCTGTTCTGTCTCTGTACGCTTCTGGTCGTACCACGGGTAT   [600]
  M  Y  V  A  I  Q  A  V  L  S  L  Y  A  S  G  R  T  T  G  I  [124]
TGTTTTGGATTCTGGTGATGGTGTCACCCACACTGTCCCCATCTATGAAGGTTACGCCCT   [660]
  V  L  D  S  G  D  G  V  T  H  T  V  P  I  Y  E  G  Y  A  L  [144]
TCCCCACGCCATCATGAGGTTGGATCTTGCCGGCCGTGACCTGACTGATTACCTCATGAA   [720]
  P  H  A  I  M  R  L  D  L  A  G  R  D  L  T  D  Y  L  M  K  [164]
GATCCTCACTGAGCGTGGCTACTCCTTCACCACCACCGCTGAGAGAGAAATCGTCAGAGA   [780]
  I  L  T  E  R  G  Y  S  F  T  T  T  A  E  R  E  I  V  R  D  [184]
CATCAAGGAGAAGCTGCTGTACATCGCCCTTG                               [812]
  I  K  E  K  L  L  Y  I  A  L                                 [194]
```

Figure 2.17. *H.irisA1c* partial nucleotide sequence and predicted amino acid sequence. Legend as for Figure 2.15.

RESTRICTION ENZYME SITES

The predicted restriction enzyme sites of *H.irisA1*, *H.irisA2*, *H.irisA3*, *H.irisA1a*, *H.irisA1b* and *H.irisA1c* were found using DNA Strider™ 1.2. The presence of cut sites for 192 restriction enzymes (those available from NEB) were tested. Of these enzymes 46 did not cut within the known *H. iris* actin sequences. The full predicted restriction enzyme site list is given Appendix E, a selection of predicted restriction sites are shown in Table 2.5.

Table 2.5. Selected restriction enzyme sites for *H. iris* actin genes.

Restriction Enzyme	Recognition Site	*H.iris A1*	*H.iris* A1a	*H.iris* A1b	*H.iris* A1c	*H.iris A2*	*H.iris A3*
Alu I	ag↓ct	7	5	5	6	9	8
Apa I	gggcc↓c	0	0	0	1	0	0
*Bam*H I	g↓gatcc	0	0	0	0	0	0
Cla I	at↓cgat	0	0	0	0	0	0
*Eco*R I	g↓aattc	1	0	0	0	0	0
Hae III	gg↓cc	4	2	3	5	5	5
*Hin*d III	a↓agctt	0	0	0	1	0	0
Not I	gc↓ggccgc	0	0	0	0	0	0
Pst I	ctgca↓g	0	0	0	0	0	0
Pvu II	cag↓ctg	1	0	0	0	0	0
Rsa I	gt↓ac	7	4	3	3	7	6
Xba I	t↓ctaga	0	0	0	0	0	0

SEQUENCE ANALYSIS

GENE IDENTIFICATION

The identities of the putative *H. iris* actin sequences were investigated by comparison to actin gene sequences deposited in the GenBank database using BLAST 2.2.8 (Altschul *et al.*, 1997). BLAST results showed that *H.irisA1*, *H.irisA2* and *H.irisA3* were very similar to actin sequences from a range of taxa, including molluscs, echinoderms and vertebrates (Table 2.6). *H. iris* putative actin genes showed greater than 85% sequence similarity to actin genes from molluscs, with up to 95% sequence similarity to actin sequences from other haliotids. *H. iris* actin genes showed sequence similarities of 82-91% and 83-91% to echinoderms and vertebrates, respectively. A diagrammatic comparison of gene structure of invertebrate actin genes is shown in Figure 2.18.

The sequences of *H.irisA1*, *H.irisA2* and *H.irisA3* were aligned to the corresponding partial actin sequences obtained by Flint (2000) using ClustalX 1.8. The percentage sequence difference between each pair of sequences was calculated. Bases which were ambiguous in the partial sequences were noted and were not included in the calculation of percentage sequence difference. Where non-synonymous changes occurred alternative amino acid were indicated.

Alignment of *H.irisA1* to the 381 nt actin sequence showed that there were 29 nt substitutions between the sequences, a difference of 7.6% (Figure 2.19). In comparison, alignment of the *H.irisA2* to the 669 nt sequence showed one nucleotide substitution from 669 nt, a sequence difference of 0.1% (Figure 2.20). The alignment of *H.irisA3* to the 954 nt sequence showed 10 nucleotide substitutions from 954 nt, a sequence difference of 1.0% (Figure 2.21).

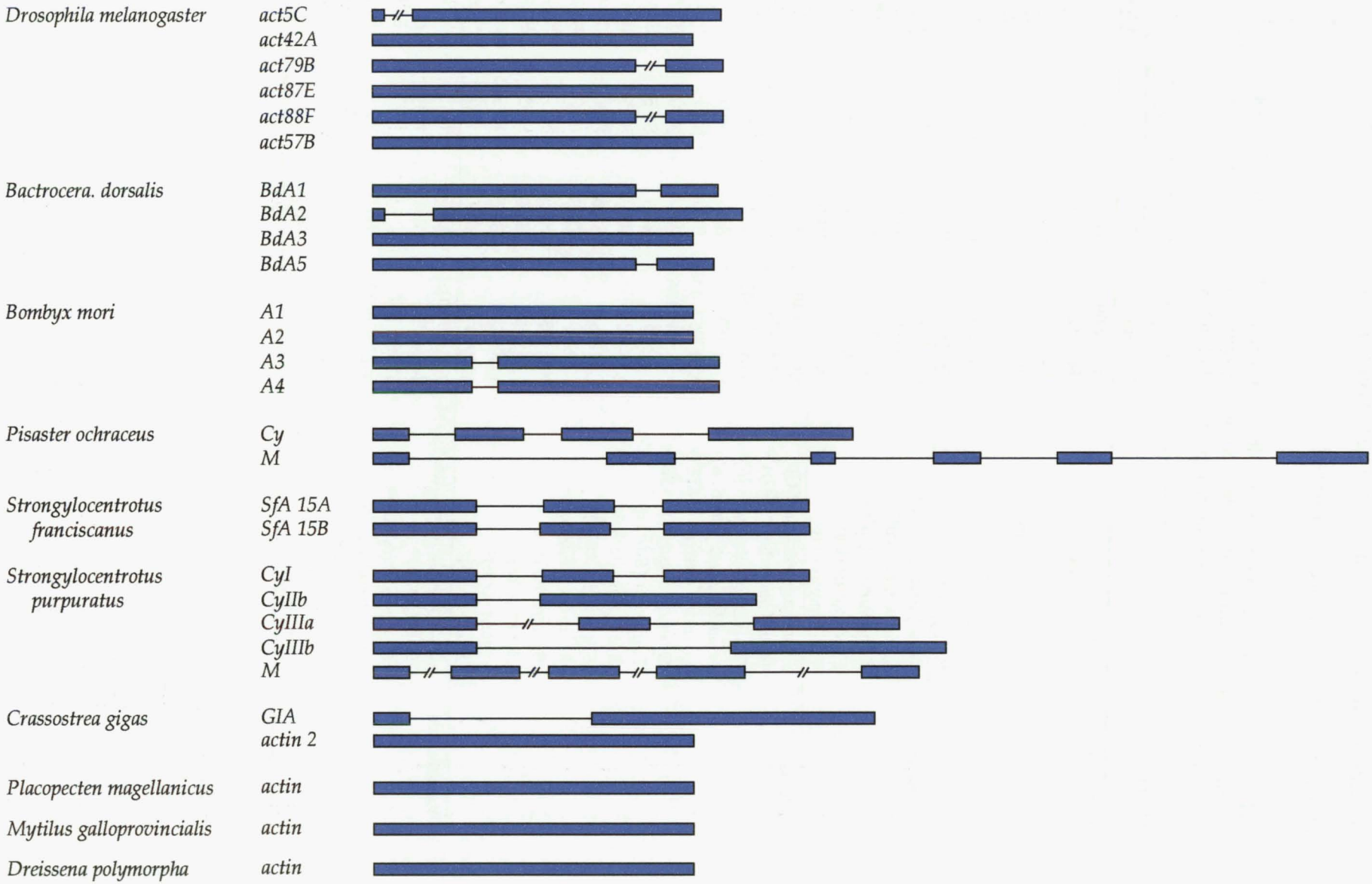
The non-synonymous nucleotide changes from the 381 nt sequence to *H.irisA1* predicted the following amino acid changes: a methionine to a valine at aa 219, a glycine to a valine at aa 231, a serine to an alanine at aa 231 and a phenylalanine to a tyrosine at aa 279. From the 954 nt sequence to *H.irisA3* the predicted amino acid changes were: valine to an isoleucine at aa 64 and a

proline to an alanine at aa 310. There were no non-synonymous substitutions from the 669 nt sequence to *H.irisA2*.

Table 2.6. Percentage sequence similarity of *H.irisA1*, *H.irisA2* and *H.irisA3* to selected mollusc, echinoderm and vertebrate actin nucleotide sequences. For each species the gene name, accession number and sequence type (DNA or mRNA) are given.

Phyla	Species	Accession Number	Sequence Similarity (%)		
			H.irisA1	*H.irisA2*	*H.irisA3*
Molluscs	*H. discus hannai*[m]	AY380809	89.0	91.3	94.0
	H. rufescens[d*]	AF032125	95.0	88.5	90.5
	A. californica (*Cy*)[m]	U01352	87.5	87.0	85.5
	A. californica (*M*)[m]	X52868	88.0	85.5	86.0
	C. gigas (*GIA*)[m]	AF026063	85.5	82.5	85.0
	C. gigas (2)[m]	AB071191	85.5	86.0	86.0
	P. magellanicus[m]	U55046	86.0	87.6	85.5
Echinoderms	*S. purpuratus* (*cyIIb*)[d]	M35323	90.8	86.3	86.8
	S. purpuratus (*cyIIIa*)[m]	M30511	82.0	82.0	82.5
	S. purpuratus (*cyIIIb*)[d]	M35324	86.5	86.0	86.8
	P. ochraceus (*Cy*)[d]	M26501	89.0	89.7	85.8
	P. ochraceus (*M*)[d]	M26500	91.0	90.0	88.7
Insects	*D. melanogaster* (*act5C*)[m]	NM_078497.2	91.5	83.5	83.0
	D. melanogaster (*act42A*)[m]	NM_078901.2	80.0	81.0	79.0
	D. melanogaster (*act88F*)[m]	NM_079643.1	83.0	91.3	82.0
	D. melanogaster (*act57B*)[m]	NM_079076.3	86.0	89.0	83.0
	D. melanogaster (*act87E*)[m]	NM_057743.3	91.0	89.3	91.0
	D. melanogaster (*act79B*)[m]	NM_079486.2	82.0	81.0	87.7
	B. dorsalis (*BdA1*)[d]	L12253.1	82.5	86.5	83.0
	B. dorsalis (*BdA2*)[d]	L12254.1	92.5	81.5	80.0
	B. dorsalis (*BdA3*)[d]	L12255.1	89.0	82.7	89.4
	B. dorsalis (*BdA5*)[d]	L12256.1	87.8	87.3	86.25
	B. mori (*A1*)[d]	X05185.1	86.5	88.7	81.5
	B. mori (*A2*)[d]	X06363.1	81.0	90.5	87.7
	B. mori (*A3*)[d]	U49854.1	87.5	82.7	80.0
	B. mori (*A4*)[d]	U49644.1	86.0	90.7	92.6
Vertebrates	*C. carpio* (skeletal muscle α)[m]	AY395870	83.5	83.5	91.0
	C. carpio (β)[d]	M24113	85.2	85.0	85.2
	O. mossambicus (α)[m]	AB037866	84.0	83.3	83.5
	O. mossambicus (β)[m]	AB0378651	86.5	86.0	85.5
	O. mykiss (α)[m*]	AF503211	87.0	83.0	86.0
	O. mykiss (β)[m]	AF157514	86.5	83.5	84.0
	P. flavescens[m]	AY332493	89.0	85.5	85.0

[m] mRNA
[d] DNA
* Partial sequence


Drosophila melanogaster
act5C
act42A
act79B
act87E
act88F
act57B
Bactrocera. dorsalis
BdA1
BdA2
BdA3
BdA5
Bombyx mori
A1
A2
A3
A4
Pisaster ochraceus
Cy
M
Strongylocentrotus franciscanus
SfA 15A
SfA 15B
Strongylocentrotus purpuratus
CyI
CyIIb
CyIIIa
CyIIIb
M
Crassostrea gigas
GIA
actin 2
Placopecten magellanicus
actin
Mytilus galloprovincialis
actin
Dreissena polymorpha
actin

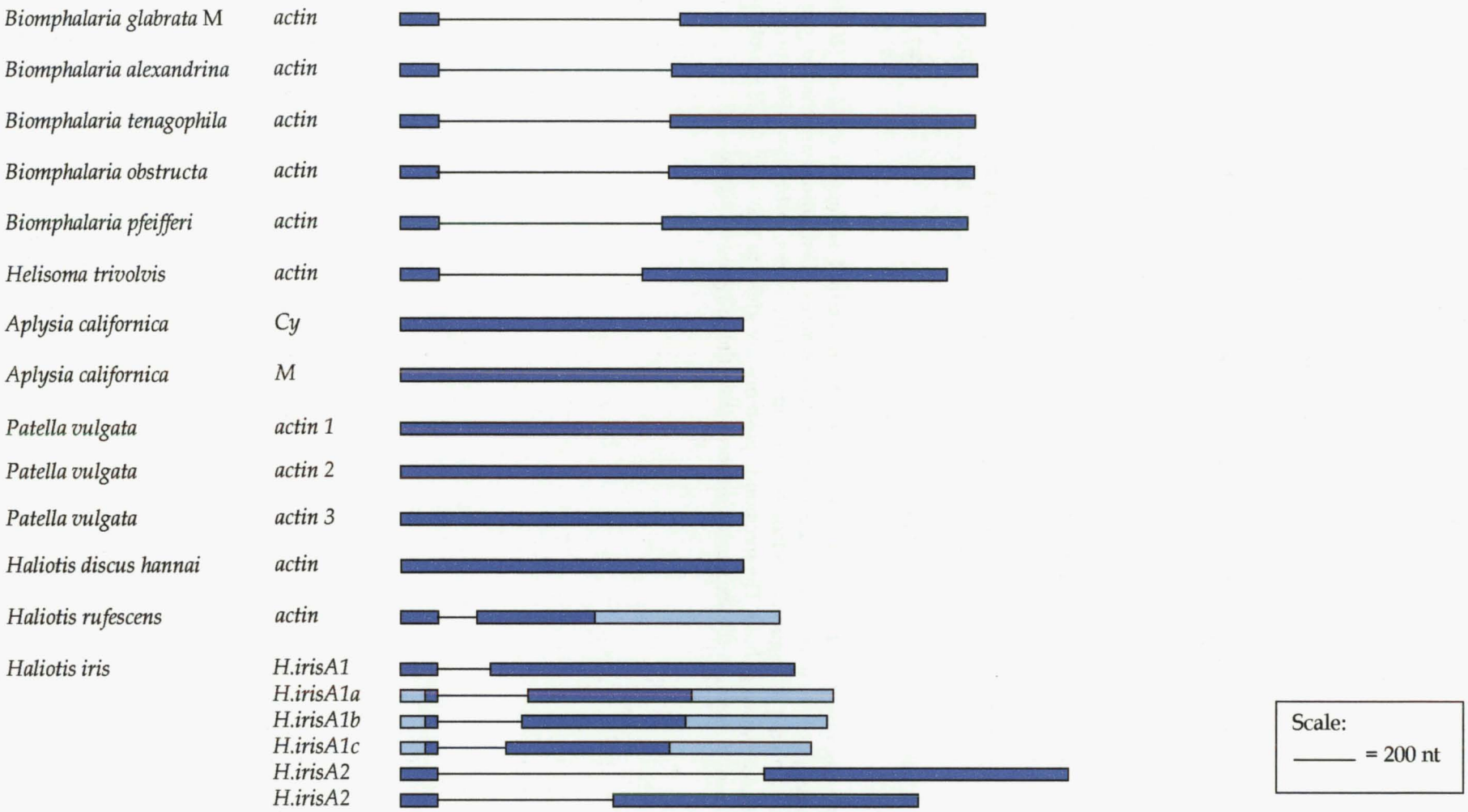


Figure 2.18. Gene structure of invertebrate and *H. iris* actin genes. Dark blue bars represent exons, pale blue bars represent predicted exons (unsequenced regions), lines represent sequenced introns, broken lines represent unsequenced introns. The genes are shown to scale.

```
381 nt        GACTCTGGTGATGGTGTCACCCACACTGTCCCCATCTATGAAGGTTATGC  [50]
H.irisA1      ..T.....................................C..        [683]

381 nt        CCTTCCCCACGCCATCATGAGGTTGGATCTTGCCGGCCGTGACCTGACTG  [100]
H.irisA1      ..................................................  [733]

381 nt        ATTACCTCATGAAGATCCTCACTGAGCGCGGCTACTCCTTCACCACCACC  [150]
H.irisA1      ........................T.........................  [783]

381 nt        GCTGAGAGAGAAATCGTCAGAGACATCAAGGAGAAGCTCTGCTACATCGC  [200]
H.irisA1      .............................................G....  [833]

381 nt        CCTTGATTTCGAGCAGGAAATGGCTACCGCCTCATCTTCTTCATCCCTGG  [250]
H.irisA1      T..C..C........A..G...T....A..TG.T..C.....C.......  [883]

381 nt        AGAAGAGCTACGAGTTGCCCGACGGTCAAGTCATCACCATCGGAAACGAG  [300]
H.irisA1      ....................T.....G..........T..T......  [933]

381 nt        AGGTTCCGWTGCCCAGAGTCTCTCTTCCAGCCATCCTTCTTGGGTATGGA  [350]
H.irisA1      ........T.....C..A..C........A..............A.....  [983]

381 nt        ATCTGCTGGTATCCATGARACCACCTTCAAC                     [381]
H.irisA1      ............A..C..G.....A.A....                     [1014]
```

Figure 2.19. Alignment of *H.irisA1* to the *H. iris* 381 nt actin 1 sequence of Flint (2000). Identical nucleotides are indicated by dots (.). Non-synonymous substitutions are shaded grey. Differences involving ambiguous nucleotides are indicated by boxes. Subtype-specific primer binding sites are shown in bold. The nucleotide positions are given to the right, for *H.irisA1* the nucleotide numbers refer to the sequence position within the full gene sequence.

```
669 nt      GACTCTGGTGATGGTGTGTCTCACACTGTTCCTATCTACGAGGGTTATGC  [50]
H.irisA2    ..................................................  [1587]

669 nt      CCTTCCCCACGCCATCATGAGACTGGACCTTGCTGGGCGTGACCTGACAG  [100]
H.irisA2    ..................................................  [1637]

669 nt      ACTACCTCATGAAGATCCTGACCGAGCGTGGTTACTCCTTCACCACCACT  [150]
H.irisA2    ..................................................  [1687]

669 nt      GCCGARCGAGAAATCGTCAGGGACATCAAAGAAAAATTAGCTTACATTGC  [200]
H.irisA2    .....G............................................  [1737]

669 nt      TTTAGACTTTGAACAGGAAATGCAGACTGCTGCATCTTCCTCTTCATTGG  [250]
H.irisA2    ..................................................  [1787]

669 nt      AAAAGAGTTACGAGCTTCCCGACGGTCAGGTCATCACCATTGGTAACGAG  [300]
H.irisA2    ..................................T...............  [1837]

669 nt      AGGTTCCGTTGTCCAGAGGCTCTCTTCCAGCCAYCTTTCTTGGGTATGGA  [350]
H.irisA2    ..............................T...................  [1887]

669 nt      ATCTGCTGGTATCCACGAAACAACATACAACTCCATCATGAARTGTGATG  [400]
H.irisA2    ..........................................G.......  [1937]

669 nt      TTGATATYCGTAAAGACTTGTATGCCAACACTGTTCTCTCRGGAGGYASY  [450]
H.irisA2    .......T..................................G.....T.CT  [1987]

669 nt      ACCATGTTCCCTGGTATCGCCGACAGAATGCAGAAGGAGATCACAGCCCT  [500]
H.irisA2    ..................................................  [2037]

669 nt      TGCTCCAGCAACAATGAAGATCAAGATCATCGCTCCTCCAGAGAGGAAAT  [550]
H.irisA2    ..................................................  [2087]

669 nt      ACTCCGTCTGGATCGGTGGCTCYATCCTTGCCTCTCTGTCCACCTTCCAG  [600]
H.irisA2    ....................C.............................  [2137]

669 nt      CAGATGTGGATCAGCAAGCAGGAGTACGACGAGTCCGGCCCATCCATTGT  [650]
H.irisA2    ..................................................  [2187]

669 nt      CCACCGCAAATGCTTCTAA                                 [669]
H.irisA2    ..................                                  [2206]
```

Figure 2.20. Alignment of *H.irisA2* to the *H. iris* 669 nt actin 2 sequence of Flint (2000). Identical nucleotides are indicated by dots (.). Non-synonymous substitutions are shaded grey. Differences involving ambiguous nucleotides are indicated by boxes. Subtype-specific primer binding sites are shown in bold. The nucleotide positions are given to the right, for *H.irisA2* the nucleotide numbers refer to the sequence position within the full gene sequence.

954 nt	TATCAGGGTGTCATGGTTGGTATGGGACAGAAGGACAGCTACGTCGGTGA	[50]
H.irisA3	.tc.....G.....C.....	[748]
954 nt	CGAGGCTCAGTCCAAGAGAGGTGTCCTCACCCTCAAGTACCCCATTGAGC	[100]
H.irisA3	.....A.....	[798]
954 nt	ACGGTATCGTCACCAACTGGGACGACATGGAGAAAATCTGGCATCACACC	[150]
H.irisA3	.....G.....	[848]
954 nt	TTCTACAACGAGCTCCGTGTGGCTCCAGAGGAACATCCCGTCCTCCTGAC	[200]
H.irisA3	.....	[898]
954 nt	CGAGGCTCCCCTTAACCCCAAGGCCAACCGTGAAAAGATGACCCAGATCA	[250]
H.irisA3	.....	[948]
954 nt	TGTTTGAGACCTTCAACTCTCCAGCTATGTATGTGGCCATCCAGGCTGTT	[300]
H.irisA3	.....	[998]
954 nt	CTGTCTCTGTACGCCTCTGGTCGTACCACGGGTATTGTTCTTGACTCTGG	[350]
H.irisA3	.....	[1048]
954 nt	TGATGGTGTCACCCACACTGTCCCCATCTATGAAGGTTACGCCCTTCCCC	[400]
H.irisA3	.....	[1098]
954 nt	ACGCTATTCTTAGGTTGGACCTGGCTGGTCGTGATCTCACAGACTACATS	[450]
H.irisA3	.....G	[1148]
954 nt	ATGAAGATCCTCACTGAGCGTGGTTACTCCTTCACCACCACTGCCGAACG	[500]
H.irisA3	.....G.....	[1198]
954 nt	AGAAATCGTCAGGGA**CATCAAGGAGAAACTGGCCTATR**TGGCTCTTGACT	[550]
H.irisA3	.....G.....	[1248]
954 nt	TCGAACAGGAAATGCAGACTGCTGCCTCTTCCTCTTCATTGGAGAAGAGC	[600]
H.irisA3	.....	[1298]
954 nt	TACGAGCTTCCCGACGGTCAGGTCATCACCATTGGTAACGAGAGGTTYCG	[650]
H.irisA3	.....T..	[1348]
954 nt	TTGTCCAGAGGCTCTTTTCCAGCCTTCCTTCTTGGGTATGGAATCTGCTG	[700]
H.irisA3	.....	[1398]
954 nt	GTATCCACGAAACAACATACAACTCCATCATGAAATGTGATGTTGATATT	[750]
H.irisA3	.....	[1448]
954 nt	CGTAAAGACTTATATGCCAACACTGTTCTCTCAGGAGGTACCACCATGTT	[800]
H.irisA3	.....	[1498]
954 nt	CCCTGGTATCCCCGACAGAATGCAGAAGGAGATCACAGCCCTTGCTCCAG	[850]
H.irisA3	.....G.....	[1548]
954 nt	CAACAATGAAGATCAAGATCATCGCACCCCCAGAGAGGAAATACTCAGTC	[900]
H.irisA3	.....G...	[1598]
954 nt	TGGATAGGCGGCTCCATCCTTGCCTCTCTGTCCACCTTCCAGCAGATGTG	[950]
H.irisA3	.....A.....	[1648]

954 nt	GATT	[954]
H.irisA3	...C	[1652]

Figure 2.21. Alignment of *H.irisA3* to the *H. iris* 954 nt actin 3 sequence of Flint (2000). Identical nucleotides are indicated by dots (.). Non-synonymous substitutions are shaded grey. Differences involving ambiguous nucleotides are indicated by boxes. Subtype-specific primer binding sites are shown in bold. The nucleotide positions are given to the right, for *H.irisA3* the nucleotide numbers refer to the sequence position within the full gene sequence.

H. IRIS ACTIN GENES

The cDNA of the three *H. iris* actin genes were aligned using ClustalX 1.8 (Figure 2.22). The percentage sequence similarities between the three genes' coding regions were calculated. The 5′ and 3′ flanking sequences that were generated by primers were excluded from percentage sequence similarity estimations (the first 20 nt of *H.irisA1* and *H.irisA2* were removed. Not all of these nucleotides were generated by primers, but they were all removed to balance the alignment).

Comparison of *H.irisA1* to *H.irisA2* showed 131 nt differences from 1,088 nt (indicating 12.0% difference/88.0% similarity between the sequences), of these changes 109 were synonymous substitutions (83.2%). *H.irisA1* and *H.irisA3* showed 107 nt differences from 1,088 nt (9.8% difference/90.2% similarity), of these changes 89 were synonymous substitutions (83.2%). *H.irisA2* and *H.irisA3* showed 77 differences from 1,088 nt (7.1% difference/92.9% similarity), of these changes 65 were synonymous substitutions (84.4%).

```
                                                  C
H.irisA1    ATGGATGATGATGTTGCTGCATTGGTCTGTGACAACGGCTCCGGCATGTG    [50]
H.irisA2    ...................G.....TATC.............G.....    [50]
H.irisA3    ...................G.....TATC...................    [50]
                                                  I

                      A                   A                   P
H.irisA1    CAAAGCCGGTTTCGCCGGTGACGACGCTCCCAGAGCTGTCTTCCCCTCCA    [100]
H.irisA2    T..G.....G...T...A......C.A..............T....A.    [100]
H.irisA3    T..G.....G.....................C..G..........T.....    [100]
                      V                   P                   S

H.irisA1    TCGTCGGCCGTCCCAGACATCAGGGCGTGATGGTCGGTATGGGTCAGAAA    [150]
H.irisA2    ..........C..A........A..T.................A.....G    [150]
H.irisA3    .......T......C.T........T.................A.....G    [150]

H.irisA1    GACAGCTACGTCGGTGACGAGGCTCAGTCCAAGAGAGGTATCCTCACCCT    [200]
H.irisA2    ...................A.....A..A...C.......T.....T..    [200]
H.irisA3    ..................................................    [200]

H.irisA1    CAAGTACCCCATCGAGCACGGTATCGTCACCAACTGGGACGACATGGAGA    [250]
H.irisA2    ............T.....................................    [250]
H.irisA3    ............T.....................................    [250]

H.irisA1    AGATCTGGCATCACACCTTCTACAACGAGCTCCGCGTTGCACCAGAGGAA    [300]
H.irisA2    ................A.................T.....T........G    [300]
H.irisA3    ................................T..G..T.........    [300]

H.irisA1    CACCCCGTCCTCTTGACAGAGGCTCCCCTCAACCCTAAGGCCAACCGTGA    [350]
H.irisA2    .....T.....TC....T........A........C.............    [350]
H.irisA3    ..T.........C....C...........T.....C.............    [350]

H.irisA1    AAAGATGACCCAGATCATGTTCGAGACCTTCAACTCTCCAGCCATGTATG    [400]
H.irisA2    ....................T...................T.......    [400]
H.irisA3    ....................T...................T.......    [400]

H.irisA1    TGGCCATCCAGGCTGTTCTGTCTCTCTACGCTTCTGGTCGTACCACGGGT    [450]
H.irisA2    .......T................G.....C...........G......    [450]
H.irisA3    ........................G.....C.................    [450]

                                      T
H.irisA1    ATTGTTTTGGATTCTGGTGATGGTGTCACCCACACTGTCCCCATCTATGA    [500]
H.irisA2    ......C.T..C.............GT.T........T..T.....C..    [500]
H.irisA3    ......C.T..C.............................S.....    [500]

                                    M
H.irisA1    AGGTTACGCCCTTCCCCACGCCATCATGAGGTTGGATCTTGCCGGCCGTG    [550]
H.irisA2    G.....T.....................AC....C.....T..G....    [550]
H.irisA3    ....................T..TC.T........C..G..T..T....    [550]
                                    L

                        L
H.irisA1    ACCTGACTGATTACCTCATGAAGATCCTCACTGAGCGTGGCTACTCCTTC    [600]
H.irisA2    .......A..C.....................G..C........T.....    [600]
H.irisA3    .T..C..A..C...A.G...........G...........T.....     [600
                        M

                                                          C
H.irisA1    ACCACCACCGCTGAGAGAGAAATCGTCAGAGACATCAAGGAGAAGCTCTG    [650]
H.irisA2    ........T..C...C.............G........A..A..AT.AGC    [650]
H.irisA3    ........T..C..AC............G.............A..GGC    [650]
                                                          A
```

```
                    V                                                   S
H.irisA1    CTACGTCGCTCTCGACTTCGAGCAAGAGATGTCTACAGCTGCTTCCTCTT    [700]
H.irisA2    T...A.T...T.A.....T..A..G..A...CAG..T.....A..T..C.    [700]
H.irisA3    ...T..G.....T........A..G..A...CAG..T.....C..T..C.    [700]
                    I                                                   Q

H.irisA1    CCTCCCTGGAGAAGAGCTACGAGTTGCCCGATGGTCAGGTCATCACCATT    [750]
H.irisA2    .T..AT....A.....T......C.T.....C.............T...    [750]
H.irisA3    .T..AT................C.T.....C.................    [750]

                                                S
H.irisA1    GGTAACGAGAGGTTCCGTTGCCCCGAATCCCTCTTCCAACCATCCTTCTT    [800]
H.irisA2    ...................T..A..GG.T........G.....T.....    [800]
H.irisA3    .............T.....T..A..GG.T..T.....G..T........    [800]
                                                A

H.irisA1    GGGAATGGAATCTGCTGGTATACACGAGACCACATACAACTCTATCATGA    [850]
H.irisA2    ...T................C.....A..A..........C.......    [850]
H.irisA3    ...T................C.....A..A..........C.......    [850]

H.irisA1    AGTGCGATGTTGATATCCGTAAGGACTTGTACGCCAACACTGTTCTCTCA    [900]
H.irisA2    ....T..........T.....A........T................G    [900]
H.irisA3    .A..T..........T.....A.....A..T.................    [900]

H.irisA1    GGAGGTACCACCATGTTCCCAGGTATCGCCGACAGAATGCAGAAGGAGAT    [950]
H.irisA2    ........T..........T............................    [950]
H.irisA3    ...................T............................    [950]
                                S
H.irisA1    CACAGCCCTTGCCCCCAGCACGATGAAGATCAAAATCATCGCTCCCCCAG    [1000]
H.irisA2    ............T..AGCA..A..........G..........T....    [1000]
H.irisA3    ............T..AGCA..A..........G........A.......    [1000]
                                A
H.irisA1    AGAGGAAATACTCCGTCTGGATCGGTGGCTCCATCCTTGCCTCTCTGTCC    [1050]
H.irisA2    .................................................    [1050]
H.irisA3    ............G........A..C.........................    [1050]

H.irisA1    ACCTTCCAACAGATGTGGATCAGCAAACAAGAGTACGACGAATCTGGTCC    [1100]
H.irisA2    ........G................G..G..........G..C..C..    [1100]
H.irisA3    ........................G..G........T..G..C..C..    [1100]

H.irisA1    ATCCATTGTCCACCGCAAATGCTTCTAA                          [1128]
H.irisA2    ............................                          [1128]
H.irisA3    ............................                          [1128]
```

Figure 2.22. Alignment of *H.irisA1*, *H.irisA2* and *H.irisA3* cDNA sequences. *H.irisA2* and *H.irisA3* are shown aligned to the *H.irisA1* sequence. Identical nucleotides are indicated by dots (.). Initiation and stop codons are underlined. Non-synonymous substitutions are shaded grey, codons are boxed and alternative amino acids shown (the amino acid for *H.iris*A1 is shown above the alignment, the amino acid of the variant partial sequences are shown below the alignment). The first and last 20 nt were excluded from percentage sequence similarity calculations, as the sequence was determined from primers. The nucleotide positions are given to the right.

H. IRIS AMINO ACID SEQUENCES

The predicted amino acid sequences of the *H. iris* actins were aligned (Figure 2.23). The percentage sequence identities for the three proteins, with the primer derived sequences removed (seven amino acid were removed from the beginning and six amino acid were removed from the end of the putative actins to balance the alignment, as above), were calculated. ClustalW 1.7 identifies amino acid changes which form weakly conserved or strongly conserved groups (Appendix D), these are positively scoring amino acid changes in the Gonnet PAM250 matrix, which means they are changes which occur frequently amongst proteins from diverse taxa (Gonnet *et al.,* 1992). For the three actins the percentage of changes that involved amino acids within weakly conserved or strongly conserved groups were noted.

H.irisA1 and *H.irisA2* differed in 10 out of 362 aa (2.5% difference or 97.5% identity); of these changes six involved amino acids within weakly conserved groups (60%), three involved amino acids within strongly conserved groups (30%) and one was an unconserved change (10%). *H.irisA1* and *H.irisA3* differed in seven out of 362 aa (1.9% difference or 98.1% identity); of these changes two were within weakly conserved groups (28.6%), four were within strongly conserved groups (57.1%) and one was an unconserved change (14.3%). *H.irisA2* and *H.irisA3* differed in seven out of 362 aa (1.9% difference or 98.1% identity); three were within weakly conserved groups (42.9%), four were within strongly conserved groups (57.1%), there were no unconserved changes. The only non-conservative amino acid change between the *H. iris* actins was at aa 10. *H.irisA1* encoded a cysteine whereas *H.irisA2* and *H.irisA3* encoded an isoleucine.

```
H.irisA1    MDDDVAALVCDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQK  [50]
H.irisA2    .........I..........V...P.....S.................  [50]
H.irisA3    .........I.....................................  [50]
                              +   +     +

H.irisA1    DSYVGDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEE  [100]
H.irisA2    ..................................................  [100]
H.irisA3    ..................................................  [100]

H.irisA1    HPVLLTEAPLNPKANREKMTQIMFETFNSPAMYVAIQAVLSLYASGRTTG  [150]
H.irisA2    ..................................................  [150]
H.irisA3    ..................................................  [150]

H.irisA1    IVLDSGDGVTHTVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYSF  [200]
H.irisA2    .........S.....................................  [200]
H.irisA3    ........................L.............M..........  [200]
                    *                 *             *

H.irisA1    TTTAEREIVRDIKEKLCYVALDFEQEMSTAASSSSLEKSYELPDGQVITI  [250]
H.irisA2    ...............A.I........Q.....................  [250]
H.irisA3    ...............A..........Q.....................  [250]
                          + *         +

H.irisA1    GNERFRCPESLFQPSFLGMESAGIHETTYNSIMKCDVDIRKDLYANTVLS  [300]
H.irisA2    .........A.....................................  [300]
H.irisA3    .........A.....................................  [300]
                    *

H.irisA1    GGTTMFPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLS  [350]
H.irisA2    ......................A.........................  [350]
H.irisA3    ......................A.........................  [350]
                                 *

H.irisA1    TFQQMWISKQEYDESGPSIVHRKCF                        [375]
H.irisA2    .........................                        [375]
H.irisA3    .........................                        [375]
```

Figure 2.23. Alignment of predicted amino acid sequences of *H.irisA1*, *H.irisA2* and *H.irisA3* sequences. *H.irisA2* and *H.irisA3* are shown aligned to *H.irisA1*. Identical nucleotides are indicated by dots (.). Pluses (+) indicate weakly conserved amino acid substitutions, asterisks (*) indicate strongly conserved amino acid substitutions (as defined by the Gonnet PAM250 substitution matrix). The first seven and last six amino acids were not included in the percentage similarity calculation as they were derived from primer sequences. The amino acid numbers are given to the right.

COMPARISON TO MOLLUSC ACTIN GENES

H.irisA1, *H. irisA2* and *H.irisA3* were compared to mollusc actin cDNA sequences (Figure 2.24) and mollusc predicted protein sequences (Figure 2.25). The mollusc actins used for alignment were *H. discus hannai* (AY380809.1), *H. rufescens* (AF032123.1), *Biomphalaria glabrata* M line (AF329436), *B. alexandrina* (AF329437), *B. pfeifferi* (AF329438), *B. obstructa* (AF329439), *B. tenagophila* (AF329440), *Helisoma trivolvis* (AF329441), *Patella vulgata* subtype 1, subtype 2 and subtype 3, *Aplysia californica* cytoplasmic (U01352.1) and muscle (X52868.1), *Crassostrea gigas GIA* (AF026063.1) and *actin 2* (AB071191.1), *Mytilus galloprovincialis* (AF157491.1), *Dreissena polymorpha* (AF082863.1) and *Placopecten magellanicus* (U55046).

The molluscan cDNA sequences shared 594 nt from a total of 1,088 nt (primer sequences removed, as above), giving an overall percentage sequence similarity of 54.6% (or difference of 45.4%). Of the 494 nt substitutions between the mollusc actins, there were 422 synonymous substitutions (85.4% of substitutions) and 72 non-synonymous substitutions (14.6% of substitutions). An ambiguous *H. rufescens* nt (nt 305) was not considered as a change during estimation of sequence similarity as it was the only species with a variant nucleotide and may not encode a variant amino acid.

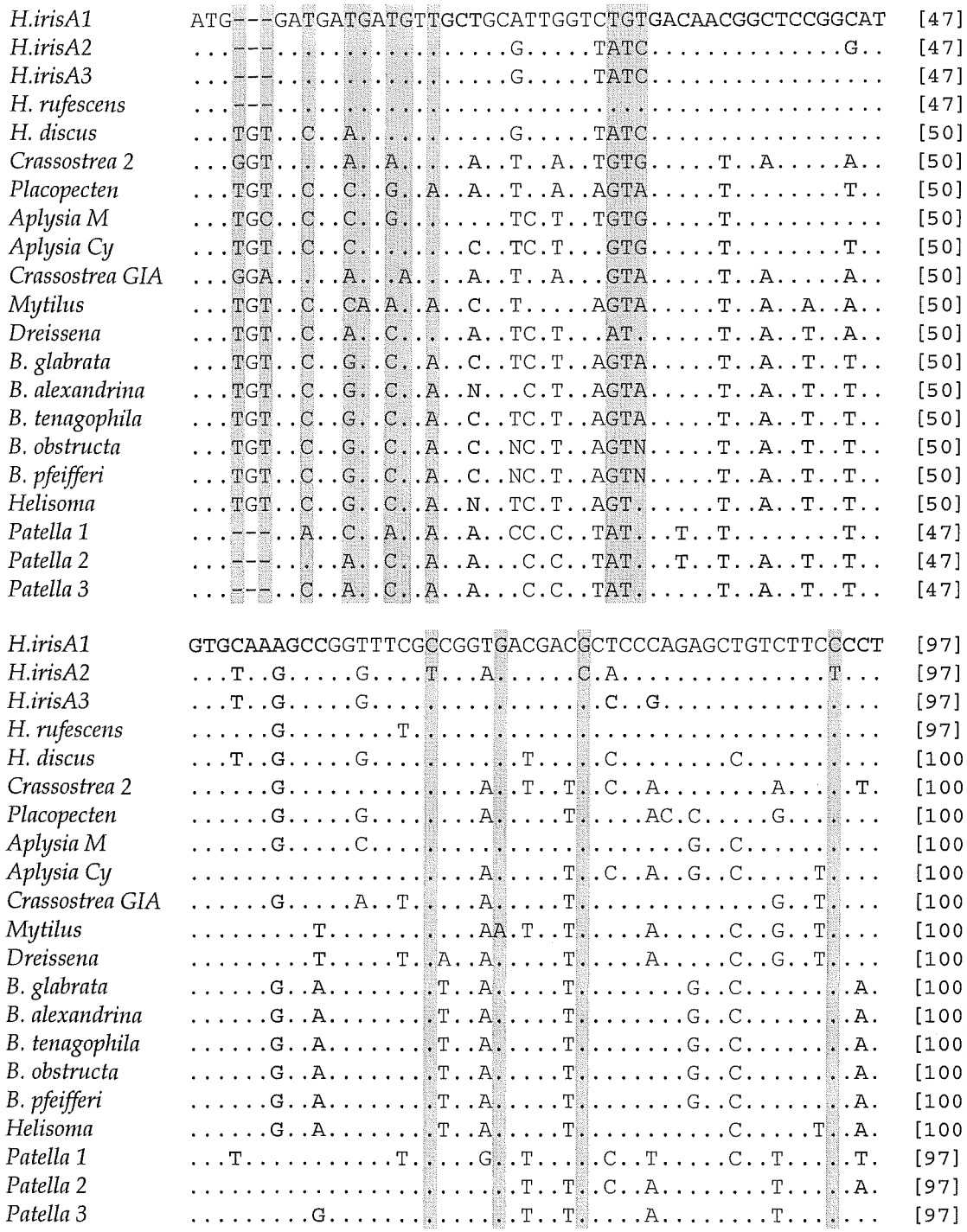

H.irisA1 ATG---GATGATGATGTTGCTGCATTGGTCTGTGACAACGGCTCCGGCAT [47]
H.irisA2 ...---.....G.....TATC.....G.. [47]
H.irisA3 ...---.....G.....TATC..... [47]
H. rufescens ...---..... [47]
H. discus ...TGT..C..A.....G.....TATC..... [50]
Crassostrea 2 ...GGT.....A..A.....A..T..A..TGTG.....T..A.....A.. [50]
Placopecten ...TGT..C..C..G..A..A..T..A..AGTA.....T.....T.. [50]
Aplysia M ...TGC..C..C..G.....TC.T..TGTG.....T..... [50]
Aplysia Cy ...TGT..C..C.....C..TC.T...GTG.....T.....T.. [50]
Crassostrea GIA ...GGA.....A..A....A..T..A...GTA.....T..A.....A.. [50]
Mytilus ...TGT..C..CA.A..A..C..T.....AGTA.....T..A..A..A.. [50]
Dreissena ...TGT..C..A..C.....A..TC.T...AT.....T..A..T..A.. [50]
B. glabrata ...TGT..C..G..C..A..C..TC.T..AGTA.....T..A..T..T.. [50]
B. alexandrina ...TGT..C..G..C..A..N...C.T..AGTA.....T..A..T..T.. [50]
B. tenagophila ...TGT..C..G..C..A..C..TC.T..AGTA.....T..A..T..T.. [50]
B. obstructa ...TGT..C..G..C..A..C..NC.T..AGTN.....T..A..T..T.. [50]
B. pfeifferi ...TGT..C..G..C..A..C..NC.T..AGTN.....T..A..T..T.. [50]
Helisoma ...TGT..C..G..C..A..N..TC.T..AGT.....T..A..T..T.. [50]
Patella 1 ...---..A..C..A..A..A..CC.C..TAT...T..T.....T.. [47]
Patella 2 ...---.....A..C..A..A...C.C..TAT...T..T..A..T..T.. [47]
Patella 3 ...---..C..A..C..A..A...C.C..TAT.....T..A..T..T.. [47]
H.irisA1 GTGCAAAGCCGGTTTCGCCGGTGACGACGCTCCCAGAGCTGTCTTCCCCT [97]
H.irisA2 ...T..G.....G....T...A.....C.A.....T... [97]
H.irisA3 ...T..G.....G.....C..G..... [97]
H. rufescens ......G.....T..... [97]
H. discus ...T..G.....G.....T.....C.....C..... [100]
Crassostrea 2 ......G.....A..T..T..C..A.....A.....T. [100]
Placopecten ......G.....G.....A.....T.....AC.C.....G..... [100]
Aplysia M ......G.....C.....G..C..... [100]
Aplysia Cy .....A.....T..C..A..G..C.....T.... [100]
Crassostrea GIA ......G.....A..T.....A.....T.....G..T.... [100]
Mytilus .....T.....AA.T..T.....A.....C..G..T.... [100]
Dreissena .....T.....T..A..A.....T.....A.....C..G..T.... [100]
B. glabrata ......G..A.....T..A.....T.....G..C.....A. [100]
B. alexandrina ......G..A.....T..A.....T.....G..C.....A. [100]
B. tenagophila ......G..A.....T..A.....T.....G..C.....A. [100]
B. obstructa ......G..A.....T..A.....T.....G..C.....A. [100]
B. pfeifferi ......G..A.....T..A.....T.....G..C.....A. [100]
Helisoma ......G..A.....T..A.....T.....C.....T..A. [100]
Patella 1 ...T.....T.....G..T.....C..T.....C..T.....T. [97]
Patella 2 .....T..T..C..A.....T.....A. [97]
Patella 3 .....G.....T..T.....A.....T..... [97]

Figure 2.24

```
H.irisA1          CCATCGTCGGCCGTCCCAGACATCAGGGCGTGATGGTCGGTATGGGTCAG [147]
H.irisA2          .A..........C..A........A..T.................A... [147]
H.irisA3          ..........T......C.T........T.................A... [147]
H. rufescens      ................T...........T........T............ [147]
H. discus         ..........T......C.T........T.................A... [150]
Crassostrea 2     ....T..T..AA.A..A...........A..C.....T............ [150]
Placopecten       ....T..T..AA.G...C.T..C.....T..C.....T............ [150]
Aplysia M         ..........A..C.....G........T..C.....T........C... [150]
Aplysia Cy        ..........AA.A...C.T..C.....T........T........C..A [150]
Crassostrea GIA   ....T.....A..C..............T........T........A... [150]
Mytilus           .......T..AA.A..A...........A..C.....T............ [150]
Dreissena         .G..T........A..A..G........T........T............ [150]
B. glabrata       .T..T..T...A.A..............T........T............ [150]
B. alexandrina    .T..T..T...A.A..............T........T............ [150]
B. tenagophila    .T..T..T...A.A..............T........T........C... [150]
B. obstructa      .T..T..T...A.A..............T........T........C... [150]
B. pfeifferi      .T..T..T...A.A..............T..T.....T............ [150]
Helisoma          .T..T..T...A.A..T...........T........T..C......... [150]
Patella 1         ....T........A..A..G........T..C.....T........A..A [147]
Patella 2         .A..T..T..A..A..A...........T..C.....T........A... [147]
Patella 3         ....T..T..A..A..A...........T..C.....T........A... [147]

H.irisA1          AAAGACAGCTACGTCGGTGACGAGGCTCAGTCCAAGAGAGGTATCCTCAC [197]
H.irisA2          ..G.....................A.....A..A...C.......T..... [197]
H.irisA3          ..G..................................................... [197]
H. rufescens      ..................................................... [197]
H. discus         ..G.............................A..................... [200]
Crassostrea 2     .................A..T..A..C...AG..................... [200]
Placopecten       .............A..A..T..A......AG..................... [200]
Aplysia M         ..G.................T.............................. [200]
Aplysia Cy        ..G........T.....A..T.............................. [200]
Crassostrea GIA   ..G........T..A..A........C...AG..................... [200]
Mytilus           ......TC......A..A..T..A..C...AG..................... [200]
Dreissena         ..G........T..T..A..T.....C...AG..................... [200]
B. glabrata       ..G........T..T.....T...........T..................... [200]
B. alexandrina    ..........T..T.....T...........T...........T..... [200]
B. tenagophila    ..........T..T.....T...........A..................... [200]
B. obstructa      ..........T..T.....T...........A..................... [200]
B. pfeifferi      ..........T..T.....T...........T...........T..... [200]
Helisoma          ..........T..T.....T.............................. [200]
Patella 1         ..........T..T..A..T..A.....A..T..................... [197]
Patella 2         ........T..T..T..A..T..A..C..A.....A.............. [197]
Patella 3         ........T..T..T..A.....A..C..A.....A.............. [197]
```

```
H.irisA1          CCTCAAGTACCCCATCGAGCACGGTATCGTCACCAACTGGGACGACATGG  [247]
H.irisA2          T............T.....................................  [247]
H.irisA3          ............T.....................................  [247]
H. rufescens      T........T.................C.....................  [247]
H. discus         ..............................................  [250]
Crassostrea 2     ...G.........T..A.............A........T.......  [250]
Placopecten       ............T.................A........T..T....  [250]
Aplysia M         ...G.....................C.....................T.......  [250]
Aplysia Cy        T............T..........T.....................T.......  [250]
Crassostrea GIA   ............T..A.....C.....................T.......  [250]
Mytilus           ...G..A.....A..T.....................A...........T....  [250]
Dreissena         ............T..A.....................................T..T....  [250]
B. glabrata       .T.G..A........T..A..T..C..T.....A........T.......  [250]
B. alexandrina    .T.G..A........T..A..T..C..T.....A........T.......  [250]
B. tenagophila    .T.G..A........T..A..T..C..T.....A........T.......  [250]
B. obstructa      .T.G..A........T..A..T..Y..T.....A........T.......  [250]
B. pfeifferi      .T.G..A........T..A..T..C..T.....A........T.......  [250]
Helisoma          ...G..A........T..A.....C..T.....A........T.......  [250]
Patella 1         .T.G..A........T..A.....................................T..T....  [247]
Patella 2         .T.G..A........T..A.....................................T..T....  [247]
Patella 3         .T.G..A..T.....T..A.....................................T..T....  [247]

H.irisA1          AGAAGATCTGGCATCACACCTTCTACAACGAGCTCCGCGTTGCACCAGAG  [297]
H.irisA2          ..........................A.....................T.....T.......  [297]
H.irisA3          ..........................................T..G..T.......  [297]
H. rufescens      ........C.....................A.....A..G..T.......  [297]
H. discus         ..............A.....................T..G..C.......  [300]
Crassostrea 2     ....A.........C.....................T..C.....C..A  [300]
Placopecten       ..........................................T..C..C..T...  [300]
Aplysia M         ..........................T.....G..T...C.C.......  [300]
Aplysia Cy        ..........................T.....G..T.....C..C...  [300]
Crassostrea GIA   ....A.........T..........T..A.....T..G..C.......  [300]
Mytilus           .A..A.....................A.....T.....C.....A  [300]
Dreissena         ..........................T........T..C..C.......  [300]
B. glabrata       ..................T........T.....TA.A.....C..T..A  [300]
B. alexandrina    ..................T........T.....TA.A.....C..T..A  [300]
B. tenagophila    ..................T........T.....TA.A.....C..T..A  [300]
B. obstructa      ..................T........T.....TA.A.....C..T..A  [300]
B. pfeifferi      ................T..T........T.....TA.A.....C..T..A  [300]
Helisoma          ..........................T.....GA.A.....C..T..A  [300]
Patella 1         .A..A.....................T..........T..A.....A.....C.....A  [297]
Patella 2         .A..................T........T..A..T..A.....T.....A  [297]
Patella 3         .A..................T........T..A..T..A.....T.....A  [297]
```

Figure 2.24

```
H.irisA1          GAACACCCCGTCCTCTTGACAGAGGCTCCCCTCAACCCTAAGGCCAACCG  [347]
H.irisA2          ..G....T.....TC....T........A........C...........  [347]
H.irisA3          .....T.........C....C...........T.....C...........  [347]
H. rufescens      ..G....YT.....YC.....................C...........  [347]
H. discus         ..G....T......C....T................C...........  [350]
Crassostrea 2     ..G.....................C....T.........................A.  [350]
Placopecten       ..G.....................C.....................C.........A.  [350]
Aplysia M         ..G.....................GC.C..T.....C...........C.........A.  [350]
Aplysia Cy        ..G.....................TC.C...................C.........A.  [350]
Crassostrea GIA   ........................C....C.....C..A........C.........A.  [350]
Mytilus           ..G....A.....TC....T........A.....T..C..A......A.  [350]
Dreissena         ..G....A.....A.....T..A..C..A........A.........A.  [350]
B. glabrata       ..G....A..A..TC.C........A..T..A........A......A.  [350]
B. alexandrina    ..G....A..A..TC.C........A..T..A........A......A.  [350]
B. tenagophila    ..G....A..A..TC.C........A..T..A........A......A.  [350]
B. obstructa      ..G....A..A..TC.C........A..T..A........A......A.  [350]
B. pfeifferi      ..G....A..A..TC.C........A..T..A........A......A.  [350]
Helisoma          ..G..T..A..A...C.C........A..A...........A......A.  [350]
Patella 1         .....T.........C.C.....A.....A.....T.....A......A.  [347]
Patella 2         .....T.........T..A..T..A..............C..A......A.  [347]
Patella 3         .....T.........T..A..T..A..............C..A......A.  [347]

H.irisA1          TGAAAAGATGACCCAGATCATGTTCGAGACCTTCAACTCTCCAGCCATGT  [397]
H.irisA2          ........................T.....................T....  [397]
H.irisA3          ........................T.....................T....  [397]
H. rufescens      ..........................................T....  [397]
H. discus         ..........................................T....  [400]
Crassostrea 2     A...........A.........................G....C.......  [400]
Placopecten       G.....................................G.C..C..T....  [400]
Aplysia M         A..G..................................G.C..C.......  [400]
Aplysia Cy        A..G.....................A.........A.A...........  [400]
Crassostrea GIA   A...........A.............................C.......  [400]
Mytilus           G.....................................TG.A.........  [400]
Dreissena         G........................T............A.C.....A....  [400]
B. glabrata       A........................T..A.........A.A.....T....  [400]
B. alexandrina    A........................T..A.........A.A.....T....  [400]
B. tenagophila    A..................A.....T..A.........A.A.....T....  [400]
B. obstructa      A..................T.....T..A.........A.A.....T....  [400]
B. pfeifferi      A........................T..A.........A.A.....T....  [400]
Helisoma          G........................T..A.........A.A.....T....  [400]
Patella 1         A..............A..T.....T..A...........A.....T....  [397]
Patella 2         A..............A........T..A........T..A.....T....  [397]
Patella 3         A..............A........T..A........T..A.....T....  [397]
```

```
H.irisA1          ATGTGGCCATCCAGGCTGTTCTGTCTCTCTACGCTTCTGGTCGTACCACG  [447]
H.irisA2          .........T.................G.....C..........G...  [447]
H.irisA3          ..........................G.....C..............  [447]
H. rufescens      G.........................G.....................  [447]
H. discus         ....A.....................G........C...........  [450]
Crassostrea 2     .C..C..........C..G.....C..G........C..........C  [450]
Placopecten       .C..C............C..C..C..G........C..........C  [450]
Aplysia M         .C..T..........C..G..C..C..G.....C..C..A........A  [450]
Aplysia Cy        .C..C............C.....C..G........C..........A  [450]
Crassostrea GIA   .C..C..........C..A.....C..G........C........A..C  [450]
Mytilus           .C..C..T........C..A..C..A..G..T.....C..........T  [450]
Dreissena         .C..T..T........A..AT....A..G..T..A..C..........T  [450]
B. glabrata       ....A........A.....G..C..AT.G.....A............A  [450]
B. alexandrina    ....A........A..A..G..C..AT.G.....A............A  [450]
B. tenagophila    ....A........A.....G..C..AT.G.....A............A  [450]
B. obstructa      ....A..........C..G..C..A..G.....A............A  [450]
B. pfeifferi      ....A........A..C..G..C..AT.G.....A............A  [450]
Helisoma          .C..T..G........C..A..C..A..G..T..A............A  [450]
Patella 1         ....T..T..T..A.....AT.....T.A.....C..........A..C  [447]
Patella 2         ....T..T.....A.....AT.A..CT.G..T..............C  [447]
Patella 3         ....T..T.....A.....AT.A..CT.G..................T  [447]

H.irisA1          GGTATTGTTTTGGATTCTGGTGATGGTGTCACCCACACTGTCCCCATCTA  [497]
H.irisA2          .........C.T..C..............GT.T........T..T.....  [497]
H.irisA3          .........C.T..C..............................  [497]
H. rufescens      .........C....C..............T..........T........  [497]
H. discus         .........C....C..............................  [500]
Crassostrea 2     .....C..CC....C..C..........................  [500]
Placopecten       .....C..CC.C..C..C..A.....................C..........  [500]
Aplysia M         .....C..GC.T..C.....C.....................C..........  [500]
Aplysia Cy        ........G.....C..............................  [500]
Crassostrea GIA   .....C..AC.C..C..C..A........GT.T.....A..........  [500]
Mytilus           .....C..AC.C..C.....A..........A.....C..A..A.....  [500]
Dreissena         ........GA....C..............GT.T........T..T..T..  [500]
B. glabrata       ........CA....C..............G..................  [500]
B. alexandrina    ........GA....C..............G..T..........T.....  [500]
B. tenagophila    ..A.....GC....C..A..........G..................  [500]
B. obstructa      ..A.....GC.A.................G..................  [500]
B. pfeifferi      ..A.....GC.A.....A..........A..T................  [500]
Helisoma          ........CA.A..C........C.....A..T..........T.....  [500]
Patella 1         ........A.....................TT.........A........  [497]
Patella 2         ........AC.....................................C..........  [497]
Patella 3         ........AC.A.....C...........TT.......A..T........  [497]
```

Figure 2.24

```
H.irisA1          TGAAGGTTACGCCCTTCCCCACGCCATCATGAGGTTGGATCTTGCCGGCC  [547]
H.irisA2          C..G.....T.....................AC....C.....T..G.  [547]
H.irisA3          ......................T..TC.T.......C..G..T..T.  [547]
H. rufescens      ----------------------------------------------  [497]
H. discus         C..G.........................AC.........T..A.  [550]
Crassostrea 2     C..........T..G........T...C.CC.TC....CT.G..T..T.  [550]
Placopecten       ...........T.............C.CC.TC.C..CT.G..T....  [550]
Aplysia M         C..G........T..G..........C.CC.TC....CT.G..T....  [550]
Aplysia Cy        C..G.....T.....................A........C..T....  [550]
Crassostrea GIA   C.............................A........C..T..A.  [550]
Mytilus           C..........T.............C.CT.TC.A..CT.G.....TA  [550]
Dreissena         ...........T.............TC..C.TC.T..C..G..T..TA  [550]
B. glabrata       .........T..T.....T..T..T..T........A...T.G..T..T.  [550]
B. alexandrina    .........T..T.....T..T..T..T........A...T.G..T..T.  [550]
B. tenagophila    .........T..T.....T..T..T..T........A...T.G..T..T.  [550]
B. obstructa      .........T..T.....T..T..T..T........A...T.G..T..T.  [550]
B. pfeifferi      .........T..T.....T..T.....T........A...T.G..T..T.  [550]
Helisoma          ...........T.....T..T.....T........A...T.G..T..T.  [550]
Patella 1         ...........TT.A..A.............A........C..T....  [547]
Patella 2         .........T.....................A......T.G..T..TA  [547]
Patella 3         ......C..T..........T..T...G.A..A......T.A..T....  [547]

H.irisA1          GTGACCTGACTGATTACCTCATGAAGATCCTCACTGAGCGTGGCTACTCC  [597]
H.irisA2          ..........A..C.................G..C........T......  [597]
H.irisA3          ....T..C..A..C...A.G...........G..........T......  [597]
H. rufescens      ----------------------------------------------  [497]
H. discus         .......C..A..C..............T..............T.....A  [600]
Crassostrea 2     ....T..C.....C.................G..C........T..T..A  [600]
Placopecten       .....T....C.....................C........T.....A  [600]
Aplysia M         .....T....C..C.....G.GT.......G..C.............T  [600]
Aplysia Cy        ....T..C..C..C.................G..C...A.G..T...AG.  [600]
Crassostrea GIA   ....T.....C..............A........A..A.....A.....T  [600]
Mytilus           .A..T..T.G....A..TGG.....A........C...A.A..T.....A  [600]
Dreissena         .A..T..T..A.....T.........A..........A........T..A  [600]
B. glabrata       .......C..A......T.G.....A..........A.G..A...AG.  [600]
B. alexandrina    .......C..A......T.G.....A..........A.A......AG.  [600]
B. tenagophila    .......C..A......T.G.....A..........A.A......AG.  [600]
B. obstructa      .......C..A......T.G.....A..........A.A......AG.  [600]
B. pfeifferi      .......C..A......T.G.....A..........A.G......AG.  [600]
Helisoma          .......T..A......T.G.....A........A...A.A......AG.  [600]
Patella 1         .A..TT....C..............A.....T..C...A.A..T.....A  [597]
Patella 2         .A..TT....C......T.G.....AG..T.G..C...A.A..T.....A  [597]
Patella 3         .A..TT....C......T.G.....A...T.G..C...A.A..T.....A  [597]
```

```
H.irisA1          TTCACCACCACCGCTGAGAGAGAAATCGTCAGAGACATCAAGGAGAAGCT  [647]
H.irisA2          ..........T..C...C...........G........A..A..AT.     [647]
H.irisA3          ..........T..C..AC...........G.............A..       [647]
H. rufescens      ------------------------------------------------    [497]
H. discus         ........................T.....G.............A..       [650]
Crassostrea 2     .....A........C.....................................  [650]
Placopecten       ............C.................G.............A..       [650]
Aplysia M         ............C........G.....GC.T.................      [650]
Aplysia Cy        ..........T..C.............C.T.................      [650]
Crassostrea GIA   ..........A..C.....................................A..  [650]
Mytilus           .....A........G.............T........T.....A..AT.    [650]
Dreissena         .....A.....A.....AC.T.....T...........T..A..A.....    [650]
B. glabrata       ........A..A......C.T.....T...C.T...........A.....    [650]
B. alexandrina    ..T.....A..A......C.T.....T...C.T...........A.....    [650]
B. tenagophila    ........A..A......C.T.....T...C.T...........A.....    [650]
B. obstructa      ........A..A......C.T.....T...C.T...........A.....    [650]
B. pfeifferi      .....T..A..A......C.T.....T...C.T...........A..A..    [650]
Helisoma          ........A..A......C.T.....T...C.............A..A..    [650]
Patella 1         ..........A......C........T...C.............A..AT.    [647]
Patella 2         .....T.....A..............T..TC....T........A..AT.    [647]
Patella 3         .....T.....A..............T...C....T.....A..A...T.    [647]

H.irisA1          CTGCTACGTCGCTCTCGACTTCGAGCAAGAGATGTCTACAGCTGCTTCCT  [697]
H.irisA2          AGCT...A.T...T.A.....T..A..G..A...CAG..T.....A..T.  [697]
H.irisA3          GGC...T..G.....T........A..G..A...CAG..T.....C..T.  [697]
H. rufescens      ------------------------------------------------    [497]
H. discus         .GA...T..A..A.............G......G....T.....C..T.    [700]
Crassostrea 2     ......T.....C.............G..A...G.C..C..C.......    [700]
Placopecten       ......T..T..C...........A.C......G.C..C..C..C..A.    [700]
Aplysia M         G..........C..G..........G......G.C..C..C..C....     [700]
Aplysia Cy        G.....T.A...C..T..........G......CAG..C.....C..A.    [700]
Crassostrea GIA   G........T..C..G........A.........A....T.........A.  [700]
Mytilus           G.....T..T.....T..T........G..A.....A..C..C.....T.    [700]
Dreissena         T.....T..T..A..T.....T.....G..A...CAA..T..A..CAGT.    [700]
B. glabrata       ......T..G.....A..T..T..A..G..A...CAA......T.CA...    [700]
B. alexandrina    ......T..G.....A..T..T..A..G..A...CAA......T.CA...    [700]
B. tenagophila    ......T..G.....A..T..T..A..G..A...CAA......T.CA...    [700]
B. obstructa      A..T..T..A.....A..T..T..A..G..A...CAA......A.CAG..    [700]
B. pfeifferi      A.....T..G.....A..T..T..A..G..A...CAC........CA...    [700]
Helisoma          ......T.........A..T..T..A..G..A...CAA......T..A...    [700]
Patella 1         A..T.....T..C...........A.....A...GGA........CAA..    [697]
Patella 2         A..T..T.CT.....T..T.....A.....A...G.A..G.....C..G.    [697]
Patella 3         G..T..T..T.....T..T.....A.....A.....A..G.....C..G.    [697]
```

Figure 2.24

```
H.irisA1          CTTCCTCCCTGGAGAAGAGCTACGAGTTGCCCGATGGTCAGGTCATCACC  [747]
H.irisA2          .C..T..AT....A.....T......C.T.....C............T  [747]
H.irisA3          .C..T..AT..............C.T.....C.............  [747]
H. rufescens      ----------------------------------------------  [497]
H. discus         .A.....................C.T.....C.............  [750]
Crassostrea 2     .C..T.....C............C.T.....C.............  [750]
Placopecten       .C..A.....C............C.T.....C.............  [750]
Aplysia M         .C..........A..........C.T.....C..A..........  [750]
Aplysia Cy        .C..T..................C.T.....C..C..A.......  [750]
Crassostrea GIA   .C..A..T..A..........T..AC.T.....C.............  [750]
Mytilus           .A..T.....A..A..........A..........A.....T.....  [750]
Dreissena         .AAG....T....A........T...C.C..T..C.............  [750]
B. glabrata       ....A..T..T........T..T...C.T..T..C.....A.......  [750]
B. alexandrina    ....A..T..T........T..T...C.T..T..C.....A.......  [750]
B. tenagophila    ....A..T..T........T..T...C.T..T..C.....A.......  [750]
B. obstructa      ....A..T..T........T..T...C.T..T..C.....A.......  [750]
B. pfeifferi      ....A..A..T........T..T...C.T..T..C.....A.......  [750]
Helisoma          ....A.....T..........T...C.T..T........A.......  [750]
Patella 1         .A..A..T..T.....A........A..A..T..C.....A.......  [747]
Patella 2         .A..A..TT....A.....T..T..AC.T..T........A........T  [747]
Patella 3         .A..A..TT....A.....T..T..AC.T..T........A........T  [747]

H.irisA1          ATTGGTAACGAGAGGTTCCGTTGCCCCGAATCCCTCTTCCAACCATCCTT  [797]
H.irisA2          ........................T..A..GG.T.........G.....T..  [797]
H.irisA3          ..................T.....T..A..GG.T..T.....G..T.....  [797]
H. rufescens      ----------------------------------------------  [497]
H. discus         ..C..A........A........T..A..GG..........G.......  [800]
Crassostrea 2     ..C...........A...........A..G..........G.......  [800]
Placopecten       ..C..A......C.T...A.G..T.....................G.......  [800]
Aplysia M         ..C..A......C.C...A.G........G..T.........G...AT.C.  [800]
Aplysia Cy        .....C......C.T...A.G..T.....GG..A.G.....G.......  [800]
Crassostrea GIA   .....C......C.A...A.G.....A..GG..A.G.....G.......  [800]
Mytilus           ...........A..A...A.G..T..A.....AT.A.............  [800]
Dreissena         .....A......C.A...........A..GG.AA.G.....G.....T..  [800]
B. glabrata       .....C..T...C.A...A.A..T..A...G.TACG.....G.....T..  [800]
B. alexandrina    .....C..T...C.A...A.A..T..A...G.TA.G.....G.....T..  [800]
B. tenagophila    .....C..T...C.A...A.A..T..A...G.TA.G.....G.....T..  [800]
B. obstructa      .....C..T...C.A...A.A..T..A...G.AT.G.....G.....T..  [800]
B. pfeifferi      .....C..T...C.A...A.A..T..A...G.TG.G.....G.....T..  [800]
Helisoma          .....C......C.T...A.A..T..T...G..A.G.A.........T..  [800]
Patella 1         ...........AC.A.....A.....A...G.T..............T..  [797]
Patella 2         ...........A..A.....A.....A...G.TG.T ..........T..  [797]
Patella 3         .....C..T..A..A.....A..T..A...A.G..T.........T..T..  [797]
```

```
H.irisA1          CTTGGGAATGGAATCTGCTGGTATACACGAGACCACATACAACTCTATCA  [847]
H.irisA2          ......T.....................C.....A..A.............C....  [847]
H.irisA3          ......T.....................C.....A..A.............C....  [847]
H. rufescens      ------------------------------------------------------  [497]
H. discus         ......T.....................C..T..A..A.............C....  [850]
Crassostrea 2     ......T.....................C..T..A....T.....T..C....  [850]
Placopecten       ......T.....................C.....C...............C....  [850]
Aplysia M         TC.T..G.....................C..........C.........C....  [850]
Aplysia Cy        .C....T.....................G.C..........C.T......C....  [850]
Crassostrea GIA   .C.T..........T.C..A..C..T..A..AT.........AG.....  [850]
Mytilus           ......T.....................C..T..A..............AG.....  [850]
Dreissena         .C....T.....................C.....C..T..A....C.....AG...T.  [850]
B. glabrata       .C.T..T......G..............C..T..A..A..C.........C....  [850]
B. alexandrina    .C.T..T......G..............C..T..A..A..C.........C....  [850]
B. tenagophila    .C.T..T......G..............C..T..A..A..C.........C....  [850]
B. obstructa      .C.T..T......G..............C..T..A..A..C.........C....  [850]
B. pfeifferi      .C.T..T.....................C..T..A..A..C.........C....  [850]
Helisoma          .C.T..T.....................C..T..A..A..C.........C....  [850]
Patella 1         ......T.....................C..T..A..............C....  [847]
Patella 2         ..A..T.........T.........T..T..A..A..T.....T.....A.  [847]
Patella 3         .A.T..C.........T.........T..T..A..A..T.....T.....A.  [847]

H.irisA1          TGAAGTGCGATGTTGATATCCGTAAGGACTTGTACGCCAACACTGTTCTC  [897]
H.irisA2          .......T...........T.....A........T.....................  [897]
H.irisA3          ....A..T...........T.....A.....A..T.....................  [897]
H. rufescens      ------------------------------------------------------  [497]
H. discus         ........................A.....................  [900]
Crassostrea 2     ..........C..C..C...........TC.............C..C..T  [900]
Placopecten       ..........C..C..C...........TC.............C..G  [900]
Aplysia M         ..........C..G..C...........TC.............C..C..G  [900]
Aplysia Cy        ..........C..G..C...........C.C.............C...  [900]
Crassostrea GIA   ....A..T.....C...........A...........T..T.T...C..G  [900]
Mytilus           .......T.....C.............................C..CT.G  [900]
Dreissena         .......T.....A...........A........T.....C..T  [900]
B. glabrata       ....A..T........C..T.....A........T..T.....AT.G  [900]
B. alexandrina    .......T........C..T.....A........T..T.....AT.G  [900]
B. tenagophila    ....A..T........C..T.....A........T..T.....AT.G  [900]
B. obstructa      ....A..T........C..T.....A........T..T.....AT.G  [900]
B. pfeifferi      ....A..T........C..T.....A........T..T.....AT.G  [900]
Helisoma          .......T........C..T.....A........T.....T.G  [900]
Patella 1         ....A..T.....C........A.....T.............C..CT.G  [897]
Patella 2         ....A.....................A........T.....T.....AT.A  [897]
Patella 3         ....A..T.................A........T.....T.....AT.A  [897]
```

Figure 2.24

```
H.irisA1          TCAGGAGGTACCACCATGTTCCCAGGTATCGCCGACAGAATGCAGAAGGA  [947]
H.irisA2          ..G.......T.............T.........................  [947]
H.irisA3          ..........................T.........................  [947]
H. rufescens      ------------------------------------------------  [497]
H. discus         ....................A...C..C.....T...C.T.........  [950]
Crassostrea 2     ..C.....................................T..T...C.T.........  [950]
Placopecten       ..C.....C.....................T.....TC.T.........  [950]
Aplysia M         ..C.....C.....................T.....T..T...C.C.........  [950]
Aplysia Cy        ........AT.....................T..C.........C.C.........  [950]
Crassostrea GIA   ..T.....................C..C..T..T...C.T.....A.....  [950]
Mytilus           ..T..T.....................................T.....................  [950]
Dreissena         ..T..T...T.T.....................C..T.....................  [950]
B. glabrata       ..T..T...T.....................C..T..A.............A..  [950]
B. alexandrina    ..T..T...T.....................C..T..A.............A..  [950]
B. tenagophila    ..T..T...T.....................C..T..A.............A..  [950]
B. obstructa      ..T..T...T.....................C..T..A.............A..  [950]
B. pfeifferi      ..T..T...T.....................C..T..A.............A..  [950]
Helisoma          ..T..T...T.....................C..T..A.............A..  [950]
Patella 1         ..T..T..C.........AT..T.....T.....T..........A..  [947]
Patella 2         ..T......T.T.......AT..T.....T................A..  [947]
Patella 3         ..T.....................A.....T.....T................A..  [947]

H.irisA1          GATCACAGCCCTTGCCCCCAGCACGATGAAGATCAAAATCATCGCTCCCC  [997]
H.irisA2          ..............T..AGCA..A...........G...........T.  [997]
H.irisA3          ..............T..AGCA..A...........G.......A....  [997]
H. rufescens      ------------------------------------------------  [497]
H. discus         ..........................A...........GG..........T.  [1000]
Crassostrea 2     .....C..T.....T..A.....C...........G.....T..C..A.  [1000]
Placopecten       A....C...T.G..T.........A...........G.....T.....A.  [1000]
Aplysia M         .....CT.T..G..T.........C...........G.............  [1000]
Aplysia Cy        ....GTC.......T..TCC...A...........G.....T..C..A.  [1000]
Crassostrea GIA   .G...C.....C..T...CCA..A.........T..GG....T.....A.  [1000]
Mytilus           A.....A.....T..A.....A.....................T..C..A.  [1000]
Dreissena         A..............T..ATC.....................G.....T.......  [1000]
B. glabrata       A.........T.G..T..ACCA..A...........G.....T.....A.  [1000]
B. alexandrina    A.........T.G..T..ACCA..A...........G.....T.....A.  [1000]
B. tenagophila    A.........T.G..T..ACCA..A...........G.....T.....A.  [1000]
B. obstructa      A.........T.G..T..ACCA..A...........G.....T.....A.  [1000]
B. pfeifferi      A.........T.G..T..ACCA..A...........G.....T.....A.  [1000]
Helisoma          A.........T.A..T..ACCA..A.........T..G.....T.....A.  [1000]
Patella 1         A.......T.T..G..T..AGCT..C.....A..T...G....T.......  [997]
Patella 2         A....T...T.G.....AGC...A.........A.....T..T..C..A.  [997]
Patella 3         A....CT..T.G......CCT..A.........A.....T..T..C..A.  [997]
```

```
H.irisA1          CAGAGAGGAAATACTCCGTCTGGATCGGTGGCTCCATCCTTGCCTCTCTG  [1047]
H.irisA2          ..................................................  [1047]
H.irisA3          ................G.......A..C.....................  [1047]
H. rufescens      --------------------------------------------------  [497]
H. discus         ............................A.........T.G.........  [1050]
Crassostrea 2     ..............................T......T.G..T......  [1050]
Placopecten       ...................................T.G..T......  [1050]
Aplysia M         .T...C.C..........A.....................G..T.....C  [1050]
Aplysia Cy        ................A.........................C...  [1050]
Crassostrea GIA   .T.............................T...........T.....C  [1050]
Mytilus           ..............................T......T.G..T..AT..  [1050]
Dreissena         ..............T........T.............G.........  [1050]
B. glabrata       ........................T..A..A..T..TT.A..A.....C  [1050]
B. alexandrina    ..............T........T..A..A..T..TT.A..A.....C  [1050]
B. tenagophila    ..............T........T..A..A..T..TT.A..A.....C  [1050]
B. obstructa      ..............T........T..A..A..T..TT.G..A.....C  [1050]
B. pfeifferi      ..............T........T..A..A..T..TT.A..A.....C  [1050]
Helisoma          ..............T........T..A..A..T...T.G..A..A..C  [1050]
Patella 1         ....AC.T..........A..........T......T.A.....A..C  [1047]
Patella 2         ............T.....A.....T.....T......T.G..T.....C  [1047]
Patella 3         ............T..T..A.....T.....T......T.G..T.....C  [1047]

H.irisA1          TCCACCTTCCAACAGATGTGGATCAGCAAACAAGAGTACGACGAATCTGG  [1097]
H.irisA2          ..........G................G..G..........G..C..  [1097]
H.irisA3          ............................G..G........T..G..C..  [1097]
H. rufescens      --------------------------------------------------  [497]
H. discus         ..A........G................G..G........T..G..C..  [1100]
Crassostrea 2     ............................G..G........T..G..C..  [1100]
Placopecten       ..............................G..A.....T..G..C..  [1100]
Aplysia M         ..........G...........TC...G..G........T..G..C..  [1100]
Aplysia Cy        ........................TC...G..........T..G..C..  [1100]
Crassostrea GIA   ..............................G.................  [1100]
Mytilus           ..............................G..A..T.............  [1100]
Dreissena         ..A........G................G..G.....T..T..GG....  [1100]
B. glabrata       ........................TC...........T.....  [1100]
B. alexandrina    ........................TC...........T.....  [1100]
B. tenagophila    ........................TC...........T..T..  [1100]
B. obstructa      ........................TC...........T..T..  [1100]
B. pfeifferi      ........................TC...........T..T..  [1100]
Helisoma          ..........G...........TC......G..A..T..T..  [1100]
Patella 1         ..A.....................TC.........A..T..T....C..  [1097]
Patella 2         ..G.....................TC.........A..T..T....C..  [1097]
Patella 3         ..G.....................TC.........A..T..T....C..  [1097]
```

Figure 2.24

H.irisA1	TCCATCCATTGTCCACCGCAAATGCTTCTAA	[1128]
H.irisA2	C.....C.....	[1128]
H.irisA3	C.....C.....	[1128]
H. rufescens	----------C----------	[497]
H. discus	C.....C.....A...A.G..G.....	[1131]
Crassostrea 2	...T.....C.....A.G.....	[1131]
Placopecten	C.....C.....A.G.....	[1131]
Aplysia M	C..C.....C..G.....G.....	[1131]
Aplysia Cy	C.....C.....A.G..G.....	[1131]
Crassostrea GIA	A.....C.....A.G.....	[1131]
Mytilus	C.....C.....A.G.....	[1131]
Dreissena	A.....A.....A.A..G.....	[1131]
B. glabrata	.....T..TA.G..G..T.....	[1131]
B. alexandrina	.....T..TA.G..G..T.....	[1131]
B. tenagophila	...T.....T..TA.G..G..T.....	[1131]
B. obstructa	...T.....T..TA.G..G..T.....	[1131]
B. pfeifferi	.....T..TA.G..G..T.....	[1131]
Helisoma	.....T..TA.G..G..T.....	[1131]
Patella 1	.....A.A.....	[1128]
Patella 2	...CG.A..C.....A.A.....T.....	[1128]
Patella 3	C..CG.A..C.....A.A.....	[1128]

Figure 2.24. Alignment of actin cDNA sequences of *H.irisA1, H.irisA2, H.irisA3* and other molluscs. Mollusc actins are shown aligned to *H. iris* actins. Identical nucleotides are indicated by dots (.). Non-synonymous substitutions are shaded grey. Dashes (-) indicate alignment gaps or lack of sequence data for *H. rufescens*. The nucleotide positions are given to the right. The first 20/23 nt (excluding/including the missing three nucleotides) and last 20 nt were not used in percentage similarity calculation as they were derived from primer sequences in *H. iris*. The box indicates an ambiguous nucleotide which was not considered a substitution.

COMPARISON TO MOLLUSC ACTIN PROTEINS

The *H. iris* predicted protein sequences were compared to the molluscan protein sequences (Figure 2.25). The amino acid position numbering used was based on the mollusc sequences (376 aa). The *H. iris*, *H. rufescens* and *P. vulgata* proteins were numbered under the mollusc system to aid comparison, they are missing the second amino acid found in other actins, so the amino acid positions should actually be one number less than those listed below. The mollusc and *H. iris* proteins shared 323 amino acids from a total of 362 amino acids, giving an overall percentage identity of 89.2% (10.8% difference).

Of the 39 amino acid changes, 23 were between amino acids within strongly conserved amino acid groups (59.0%), six were between amino acids within weakly conserved amino acid groups (15.4%) and 10 were unconserved changes (25.6%).

Comparison of variant sites showed that often one amino acid was favoured, but one or two variants existed, for other variant sites no favoured amino acid was evident. There were six variable sites where no amino acid was clearly favoured (aa 130, aa 229, aa 232, aa 262, aa 304 and aa 324) and 33 sites where one amino acid was favoured (aa 11, aa 23, aa 27, aa 33, aa 77, aa 87, aa 88, aa 98, aa 134, aa 154, aa 161, aa 171, aa 190, aa 191, aa 218, aa 220, aa 226, aa 233, aa 261, aa 263, aa 267, aa 272, aa 273, aa 275, aa 279, aa 280, aa 298, aa 307, aa 318, aa 319, aa 320 and aa 330). A summary of variable amino acid sites is given in Table 2.7, colours are used to highlight the variant amino acid(s). Where more than one variant amino acid is found, each variant is shaded a different colour, the darker the shade, the more common the variant.

```
H.irisA1          M-DDDVAALVCDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQ  [47]
H.irisA2          .-........I..........V...P.....S.................  [47]
H.irisA3          .-........I.....................................  [47]
H. rufescens      .-.....................................................  [47]
H. discus         .C.E.....I.....................................  [50]
Crassostrea 2     .G.EE.....V....................................  [50]
Placopecten       .C..E.....V....................................  [50]
Aplysia M         .C..E.....V....................................  [50]
Aplysia Cy        .C........V....................................  [50]
Crassostrea GIA   .G.E.I....V....................................  [50]
Mytilus           .C..K.....V............N........................  [50]
Dreissena         .C.E......I....................................  [50]
B. glabrata       .C.E......V....................................  [50]
B. alexandrina    .C.E......V....................................  [50]
B. tenagophila    .C.E......V....................................  [50]
B. obstructa      .C.E......V....................................  [50]
B. pfeifferi      .C.E......V....................................  [50]
Helisoma          .C.E......V....................................  [50]
Patella 1         .-E.E.....I....................................  [47]
Patella 2         .-.E......I....................................  [47]
Patella 3         .-.E......I....................................  [47]
                    +++*                 + + +      +

H.irisA1          KDSYVGDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPE  [97]
H.irisA2          .....................................................  [97]
H.irisA3          .....................................................  [97]
H. rufescens      ........................A.........R.............  [97]
H. discus         ...................................Q.............  [100]
Crassostrea 2     .....................................................  [100]
Placopecten       .....................................................  [100]
Aplysia M         ..............................................P..  [100]
Aplysia Cy        .....................................................  [100]
Crassostrea GIA   .....................................................  [100]
Mytilus           .....................................................  [100]
Dreissena         .....................................................  [100]
B. glabrata       .....................................................  [100]
B. alexandrina    .....................................................  [100]
B. tenagophila    .....................................................  [100]
B. obstructa      .....................................................  [100]
B. pfeifferi      .....................................................  [100]
Helisoma          .....................................................  [100]
Patella 1         .....................................................  [97]
Patella 2         .....................................................  [97]
Patella 3         .....................................................  [97]
                                        +           *          +
```

```
H.irisA1          EHPVLLTEAPLNPKANREKMTQIMFETFNSPAMYVAIQAVLSLYASGRTT [147]
H.irisA2          ..................................................  [147]
H.irisA3          ..................................................  [147]
H. rufescens      ..X.............................C................  [147]
H. discus         ..................................................  [150]
Crassostrea 2     ............................A.....................  [150]
Placopecten       ............................A.....................  [150]
Aplysia M         ............................A.....................  [150]
Aplysia Cy        ............................T.....................  [150]
Crassostrea GIA   ..................................................  [150]
Mytilus           ............................A.....................  [150]
Dreissena         ............................T.....................  [150]
B. glabrata       ............................T.....................  [150]
B. alexandrina    ............................T.....................  [150]
B. tenagophila    ............................T.....................  [150]
B. obstructa      ............................T.....................  [150]
B. pfeifferi      ............................T.....................  [150]
Helisoma          ............................T.....................  [150]
Patella 1         ..................................................  [147]
Patella 2         ..................................................  [147]
Patella 3         ..................................................  [147]
                                             *

H.irisA1          GIVLDSGDGVTHTVPIYEGYALPHAIMRLDLAGRDLTDYLMKILTERGYS [197]
H.irisA2          ..........S.....................................    [197]
H.irisA3          ........................L............M.........    [197]
H. rufescens      ................X--------------------------------    [166]
H. discus         ..................................................  [200]
Crassostrea 2     ........................L.........................  [200]
Placopecten       ........................L.........................  [200]
Aplysia M         ........................L............S.........    [200]
Aplysia Cy        ..................................................  [200]
Crassostrea GIA   ..........S.....................................    [200]
Mytilus           ........................LC........S.NW..........    [200]
Dreissena         ...M.....S..............L.........................  [200]
B. glabrata       ...M.....................................          [200]
B. alexandrina    ...M.....................................          [200]
B. tenagophila    ..................................................  [200]
B. obstructa      ..................................................  [200]
B. pfeifferi      ..................................................  [200]
Helisoma          ...I.....................................          [200]
Patella 1         ..........S.....................................    [197]
Patella 2         ..........................................V.......  [197]
Patella 3         ..........S..............V.........................  [197]
                     *       *                *          *      *
```

Figure 2.25

```
H.irisA1          FTTTAEREIVRDIKEKLCYVALDFEQEMSTAASSSSLEKSYELPDGQVIT  [247]
H.irisA2          ................A.I.......Q.....................  [247]
H.irisA3          ................A.........Q.....................  [247]
H. rufescens      --------------------------------------------------  [166]
H. discus         ................D.........A.....................  [250]
Crassostrea 2     ............................A.....................  [250]
Placopecten       ........................N..A.....................  [250]
Aplysia M         ............................A.....................  [250]
Aplysia Cy        ................D.......Q.....................  [250]
Crassostrea GIA   ............................T.....................  [250]
Mytilus           ..................................................  [250]
Dreissena         ............................Q.....................  [250]
B. glabrata       ............................Q..ST.................  [250]
B. alexandrina    ............................Q..ST.................  [250]
B. tenagophila    ............................Q..ST.................  [250]
B. obstructa      ............................Q..T..................  [250]
B. pfeifferi      ............................H...T.................  [250]
Helisoma          ............................Q..ST.................  [250]
Patella 1         ............................G...N.................  [247]
Patella 2         ................A.......A.....................  [247]
Patella 3         ..................................................  [247]
                                            *      *+

H.irisA1          IGNERFRCPESLFQPSFLGMESAGIHETTYNSIMKCDVDIRKDLYANTVL  [297]
H.irisA2          ..........A.......................................  [297]
H.irisA3          ..........A.......................................  [297]
H. rufescens      --------------------------------------------------  [166]
H. discus         ..........A.......................................  [300]
Crassostrea 2     ..................................................  [300]
Placopecten       ..................................................  [300]
Aplysia M         ..............IL..................................  [300]
Aplysia Cy        ..........AM............V....F.....................  [300]
Crassostrea GIA   ..........AM..........S.....S..................I..  [300]
Mytilus           ..................................................  [300]
Dreissena         ..........AM......................................  [300]
B. glabrata       ..........AT.........A............................  [300]
B. alexandrina    ..........AM.........A............................  [300]
B. tenagophila    ..........AM.........A............................  [300]
B. obstructa      ..........A..........A............................  [300]
B. pfeifferi      ..........AV......................................  [300]
Helisoma          ..........AMY.....................................  [300]
Patella 1         ..........A.......................................  [297]
Patella 2         ..........AV..........S...........................  [297]
Patella 3         ..........T......I....S...........................  [297]
                            *      **    ** *    *+
```

```
H.irisA1          SGGTTMFPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASL  [347]
H.irisA2          ......................A.......................... [347]
H.irisA3          ......................A.......................... [347]
H. rufescens      -------------------------------------------------- [166]
H. discus         ......Y.....................V..................... [350]
Crassostrea 2     .................................................. [350]
Placopecten       .................................................. [350]
Aplysia M         ...................S.............................. [350]
Aplysia Cy        ...S.............V....P.......................... [350]
Crassostrea GIA   ................V.....P.....V..................... [350]
Mytilus           .................................................. [350]
Dreissena         ...S.............................................. [350]
B. glabrata       ...S.....................P........................ [350]
B. alexandrina    ...S.....................P........................ [350]
B. tenagophila    ...S.....................P........................ [350]
B. obstructa      ...S.....................P........................ [350]
B. pfeifferi      ...S.....................P........................ [350]
Helisoma          ...S.....................P........................ [350]
Patella 1         ......Y...........S...A.....V..................... [347]
Patella 2         ...S..Y..................A.......................... [347]
Patella 3         ....................S...P.......................... [347]
                     *   +           *+*    +      *

H.irisA1          STFQQMWISKQEYDESGPSIVHRKCF                         [375]
H.irisA2          .........................                          [375]
H.irisA3          .........................                          [375]
H. rufescens      -------------------------                          [166]
H. discus         .........................                          [376]
Crassostrea 2     .........................                          [376]
Placopecten       .........................                          [376]
Aplysia M         .........................                          [376]
Aplysia Cy        .........................                          [376]
Crassostrea GIA   .........................                          [376]
Mytilus           .........................                          [376]
Dreissena         ..............A..........                          [376]
B. glabrata       .........................                          [376]
B. alexandrina    .........................                          [376]
B. tenagophila    .........................                          [376]
B. obstructa      .........................                          [376]
B. pfeifferi      .........................                          [376]
Helisoma          .........................                          [376]
Patella 1         .........................                          [375]
Patella 2         ................A........                          [375]
Patella 3         ................A........                          [375]
                              *   *
```

Figure 2.25. Alignment of predicted amino acid sequences of *H.irisA1*, *H.irisA2*, *H.irisA3* and other molluscs. Identical amino acids are indicated by dots (.), changes within strongly conserved amino acid groups are indicated by an asterisk (*), changes within weakly conserved amino acid groups are indicated by pluses (+), as defined by the Gonnet PAM250 substitution matrix. The first seven and last six amino acids were not included in the percentage similarity calculation as they were derived from the primer sequences. The amino acid numbers are given to the right.

COMPARISON OF *H.IRISA1* TO *H. RUFESCENS ACTIN*

H.irisA1 was aligned to the partial *H. rufescens* actin sequence, as the 5′ non-coding and intron B sequences of both genes were available (Figure 2.26). The percentage sequence similarity between the two genes was calculated for non-coding and coding regions. The similarities between the non-coding regions of the two genes varied from 52.3% to 78.6%: the 5′ non-coding regions shared 473 of 621 nt (76.2% similarity/3.8% difference), the 5′ UTRs shared 55 of 70 nt (78.6% similarity/21.4% difference), the intron B sequences shared 83 of 174 nt (52.3% similarity/47.7% difference). The similarities between the coding regions were high. For exon 2 there were 120/123 nt shared (97.6% similarity/2.4% difference), of the three changes all were synonymous substitutions. For exon 3 there were 332/354 nt shared (93.8% similarity/5.9% difference, A3R primer derived sequence removed), of the 22 changes 18 were synonymous substitutions (81.8%).

Table 2.7. Variable actin amino acid sites. Variant amino acids are shaded in order of preference, the most common variant is red, less common variants are shaded lighter colours.

Mollusc Species	Variant Amino Acid Site																											
	2	3	4	5	6	11	23	25	27	33	77	87	88	98	102	130	134	154	161	177	178	187	189	190	191	193	218
H.irisA1	-	D	D	D	V	C	A	D	A	P	V	W	H	A	P	S	Y	L	T	M	R	T	Y	L	M	I	C
H.irisA2	-	D	D	D	V	I	V	D	P	S	V	W	H	A	P	S	Y	L	S	M	R	T	Y	L	M	I	A
H.irisA3	-	D	D	D	V	I	A	D	A	P	V	W	H	A	P	S	Y	L	T	L	R	T	Y	M	M	I	A
Crassostrea 2	G	D	E	E	V	V	A	D	A	P	V	W	H	A	P	A	Y	L	T	L	R	T	Y	L	M	I	C
Placopecten	C	D	D	E	V	V	A	D	A	P	V	W	H	A	P	A	Y	L	T	L	R	T	Y	L	M	I	C
Aplysia M	C	D	D	D	V	V	A	D	A	P	V	W	H	P	P	A	Y	L	T	L	R	T	Y	L	S	I	C
Aplysia Cy	C	D	D	E	V	V	A	D	A	P	V	W	H	A	P	T	Y	L	T	M	R	T	Y	L	M	I	C
H. rufescens	-	D	D	D	V	C	A	D	A	P	A	R	H	A	X	S	C	L	T	-	-	-	-	-	-	-	-
H. discus	C	D	E	D	V	I	A	D	A	P	V	W	Q	A	P	S	Y	L	T	M	R	T	Y	L	M	I	D
Crassostrea GIA	G	D	E	D	I	V	A	D	A	P	V	W	H	A	P	S	Y	L	S	M	R	T	Y	L	M	I	C
Mytilus	C	D	D	K	V	V	A	N	A	P	V	W	H	A	P	A	Y	L	T	L	C	S	N	W	M	I	C
Dreissena	C	D	E	D	V	I	A	D	A	P	V	W	H	A	P	T	Y	M	S	L	R	T	Y	L	M	I	C
B. glabrata	C	D	E	D	V	V	A	D	A	P	V	W	H	A	P	T	Y	M	T	M	R	T	Y	L	M	I	C
B. alexandrina	C	D	E	D	V	V	A	D	A	P	V	W	H	A	P	T	Y	M	T	M	R	T	Y	L	M	I	C
B. tenagophila	C	D	E	D	V	V	A	D	A	P	V	W	H	A	P	T	Y	L	T	M	R	T	Y	L	M	I	C
B. obstructa	C	D	E	D	V	V	A	D	A	P	V	W	H	A	P	T	Y	L	T	M	R	T	Y	L	M	I	C
B. pfeifferi	C	D	E	D	V	V	A	D	A	P	V	W	H	A	P	T	Y	L	T	M	R	T	Y	L	M	I	C
Helisoma	C	D	E	D	V	V	A	D	A	P	V	W	H	A	P	T	Y	I	T	M	R	T	Y	L	M	I	C
Patella pPA2	-	D	E	D	V	I	A	D	A	P	V	W	H	A	P	S	Y	L	T	M	R	T	Y	L	M	V	C
Patella pPA3	-	D	E	D	V	I	A	D	A	P	V	W	H	A	P	S	Y	L	S	V	R	T	Y	L	M	I	C
Patella pPA1	-	E	D	E	V	I	A	D	A	P	V	W	H	A	P	S	Y	L	S	M	R	T	Y	L	M	I	C

Mollusc Species	Variant Amino Acid																										
	220	226	229	232	233	261	262	263	266	267	268	272	273	275	279	280	298	304	307	318	319	320	324	330	366	369
H.irisA1	V	Q	S	A	S	S	L	F	S	F	L	S	A	I	T	Y	T	T	F	I	T	A	S	I	S	S
H.irisA2	I	Q	Q	A	S	A	L	F	S	F	L	S	A	I	T	Y	T	T	F	I	T	A	A	I	S	S
H.irisA3	V	Q	Q	A	S	A	L	F	S	F	L	S	A	I	T	Y	T	T	F	I	T	A	A	I	S	S
Crassostrea 2	V	Q	A	A	S	S	L	F	S	F	L	S	A	I	T	Y	T	T	F	I	T	A	S	I	S	S
Placopecten	V	N	A	A	S	S	L	F	S	F	L	S	A	I	T	Y	T	T	F	I	T	A	S	I	S	S
Aplysia M	V	Q	A	A	S	S	L	F	I	L	L	S	A	I	T	Y	T	T	F	I	T	S	S	I	S	S
Aplysia Cy	D	Q	Q	A	S	A	M	F	S	F	L	S	A	V	T	F	T	S	-	I	V	A	P	I	S	S
H. rufescens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H. discus	V	Q	A	A	S	A	L	F	S	F	L	S	A	I	T	Y	T	S	F	I	T	A	S	V	S	S
Crassostrea GIA	V	Q	T	A	S	A	M	F	S	F	L	S	S	I	S	Y	I	S	F	V	T	A	P	V	S	S
Mytilus	V	Q	S	A	S	S	L	F	S	F	L	S	A	I	T	Y	T	S	F	I	T	A	S	I	S	S
Dreissena	V	Q	Q	A	S	A	M	F	S	F	L	S	A	I	T	Y	T	S	F	I	T	A	S	I	A	S
B. glabrata	V	Q	Q	S	T	A	T	F	S	F	L	A	A	I	T	Y	T	S	F	I	T	A	P	I	S	S
B. alexandrina	V	Q	Q	S	T	A	M	F	S	F	L	A	A	I	T	Y	T	S	F	I	T	A	P	I	S	S
B. tenagophila	V	Q	Q	S	T	A	M	F	S	F	L	A	A	I	T	Y	T	S	F	I	T	A	P	I	S	S
B. obstructa	V	Q	Q	T	S	A	L	F	S	F	L	A	A	I	T	Y	T	S	F	I	T	A	P	I	S	S
B. pfeifferi	V	Q	H	A	T	A	V	F	S	F	L	S	A	I	T	Y	T	S	F	I	T	A	P	I	S	S
Helisoma	V	Q	Q	S	T	A	M	Y	S	F	L	S	A	I	T	Y	T	S	F	I	T	A	P	I	S	S
Patella pPA2	A	Q	A	A	S	A	V	F	S	F	L	S	S	I	T	Y	T	S	Y	I	T	A	S	I	S	A
Patella pPA3	V	Q	S	A	S	T	L	F	S	F	I	S	S	I	T	Y	T	T	F	I	T	S	P	I	S	A
Patella pPA1	V	Q	G	A	N	A	L	F	S	F	L	S	A	I	T	Y	T	T	Y	I	T	S	S	V	S	S

```
H. rufescens ggaacagtgtcaaacatatatacaagcttgatggtgagaaatattagcattaata [55]
H.irisA1     ---------------------------------------------------- [0]

H. rufescens ctgtcacttgttgtttagcatattatttctgatatataaatacttaggaaattct [110]
H.irisA1     ---------------------------------------------------- [0]

H. rufescens atttttctcatgcaaaaagcccattaacttcatcaataaaatcgttatctgcacc [165]
H.irisA1     ---------------------------------------------------- [0]

H. rufescens taagaatgatcttccatcatctctatgttgacatacgttttatcatccagattat [220]
H.irisA1     ---------------------------------------------------- [0]

H. rufescens caatgaaagtcgaaaagattacaactccctctgaaatatgaatattcacagtttta [275]
H.irisA1     ---------------------------------------------------- [0]

H. rufescens gagggcaagtaaagccaacacaactactttgcacg----gcgagaacaagcaata [326]
H.irisA1     ------------aacaacatagcaactttgcatgcatggtgagaacaaacattg [42]

H. rufescens tcagggtgttacattctgtacattcgttagtattttactt--ctgggtccaatcc [379]
H.irisA1     gcagggaattccatt-tgtatattcattagtattttacttta ctgtatcgaatcc [96]

H. rufescens attattacacactgcatttcttccaggaagtttggatataactttacaatgcttt [434]
H.irisA1     agtattacatactgcatttctttttctgaattatggttgtaacattacaa-----c [146]

H. rufescens ataactgatgagtagtaagcgttgtatggattttagaatctatgcatttcccaat [489]
H.irisA1     attactaaggaggcgtaagcgttgtatgctgtttagagtctatgcatttccaaat [201]

H. rufescens aatgctgatatattcatttaacttaattt--attacc-----agcatcacattct [537]
H.irisA1     aattctgaaatatgcatttaacttatttttattacggaacaagcatcatttgca [256]

H. rufescens tgcattcatgctcgtcagctcgagaagcgcttccccattctccgcgcacatacca [592]
H.irisA1     tgcatacatactagtcaactcaaggagcgcttccc-attctccg-gaatatatca [309]

H. rufescens cgtggcgttgtccttgcttc-gggaacg----ggggtaggggc----gtacttac [638]
H.irisA1     tgtggcgtt--cactgcttgtgggagcgtttgggggggacgggctctggtacatcc [362]

H. rufescens ttctaggggaaagagagatcagtg---------cagatcacccccgactgtgaca [684]
H.irisA1     tattaggggaaa-agagttcagtgagctggatgcagttgacccctgactgtgacg [416]

H. rufescens cattcttccacatgtacacatgaaaggttgttatgcaatataatacattagaagg [739]
H.irisA1     cattcttccacatgtacacatgaaaggctcttatgcaatataatacattaggagg [471]

H. rufescens gtatatttat-------tacaattacaatggttacgtttctattattctcaaaca [787]
H.irisA1     gtatagttatcattacctacaattacattggttacgtttcaaytattcttaaaca [526]

H. rufescens caatctgattggtcgcctactaatggggtatgtataaaagacgcctgggtcagaa [842]
H.irisA1     caatctgattggtcttctactgatgggggcagtataaaagacgcatggatgctra [581]

H. rufescens catcgatattgcatccgctttcagtcttcagctgacac--atcgtctttcccgtt [895]
H.irisA1     cctcgatagtgcattcgctttcagtcttcatttgactcgcatcttgtttcctgtt [636]

H. rufescens ctctcacacagcaactacaaccATGGATGATGATGTTGCTGCATTGGTCTGTGAC [950]
H.irisA1     ctcycacacaactactacagtcATGGATGATGATGTTGCTGCATTGGTCTGTGAC [691]

H. rufescens AACGGCTCCGGCATGTGCAAGGCCGGTTTTGCCGGTGACGACGCTCCCAGAGCTG [1005]
H.irisA1     AACGGCTCCGGCATGTGCAAAGCCGGTTTCGCCGGTGACGACGCTCCCAGAGCTG [746]
```

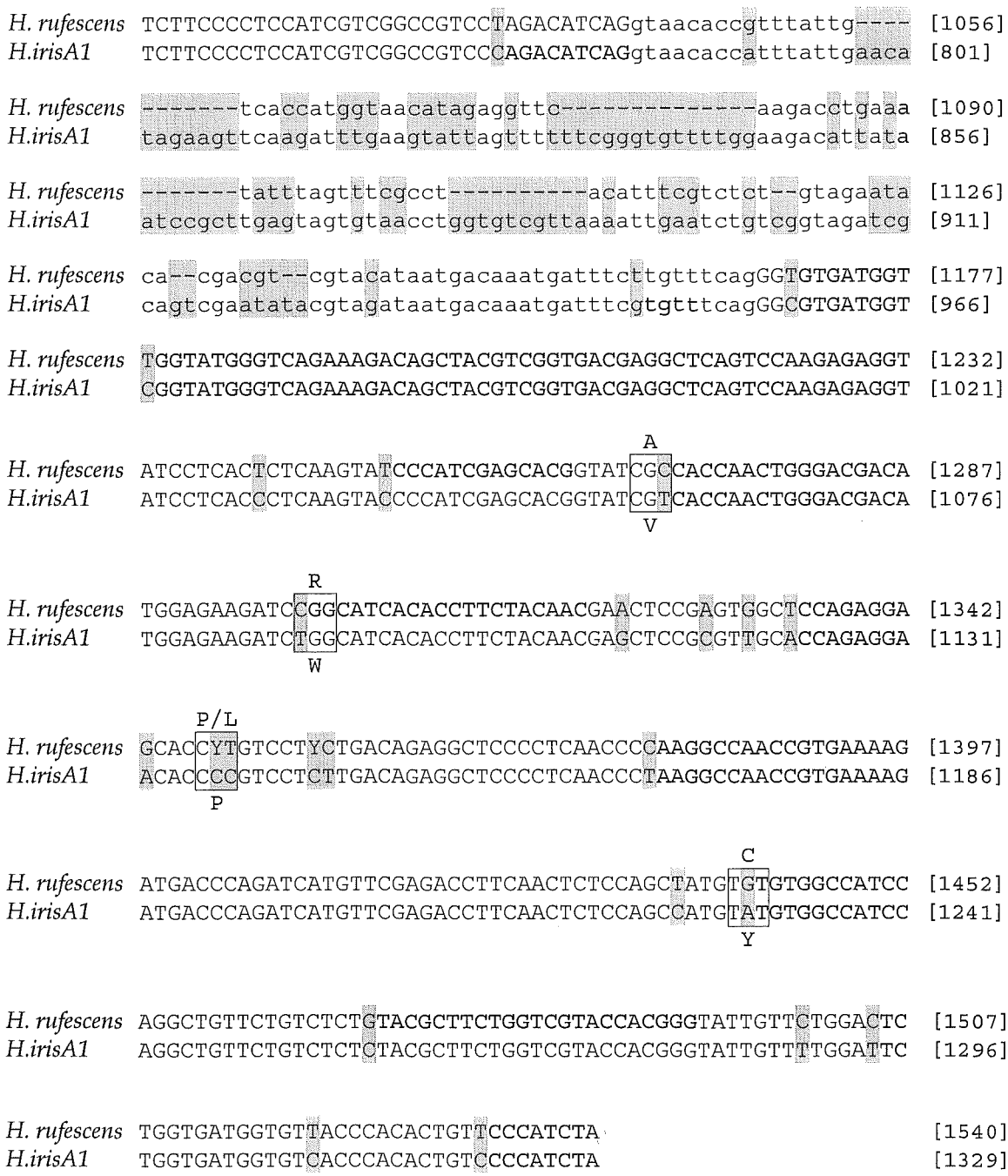


Figure 2.26. Alignment of *H.irisA1* to the partial *H. rufescens* nucleotide sequence. Lower case letters represent untranslated regions, upper case letters represent coding nucleotides. Variant nucleotides are shaded in grey. Nonsynonymous substitutions are boxed, with alternative amino acids shown. Dashes (-) represent lack of sequence data due to variation in intron size. The nucleotide positions are given to the right.

H.IRISA1 RELATED SEQUENCES

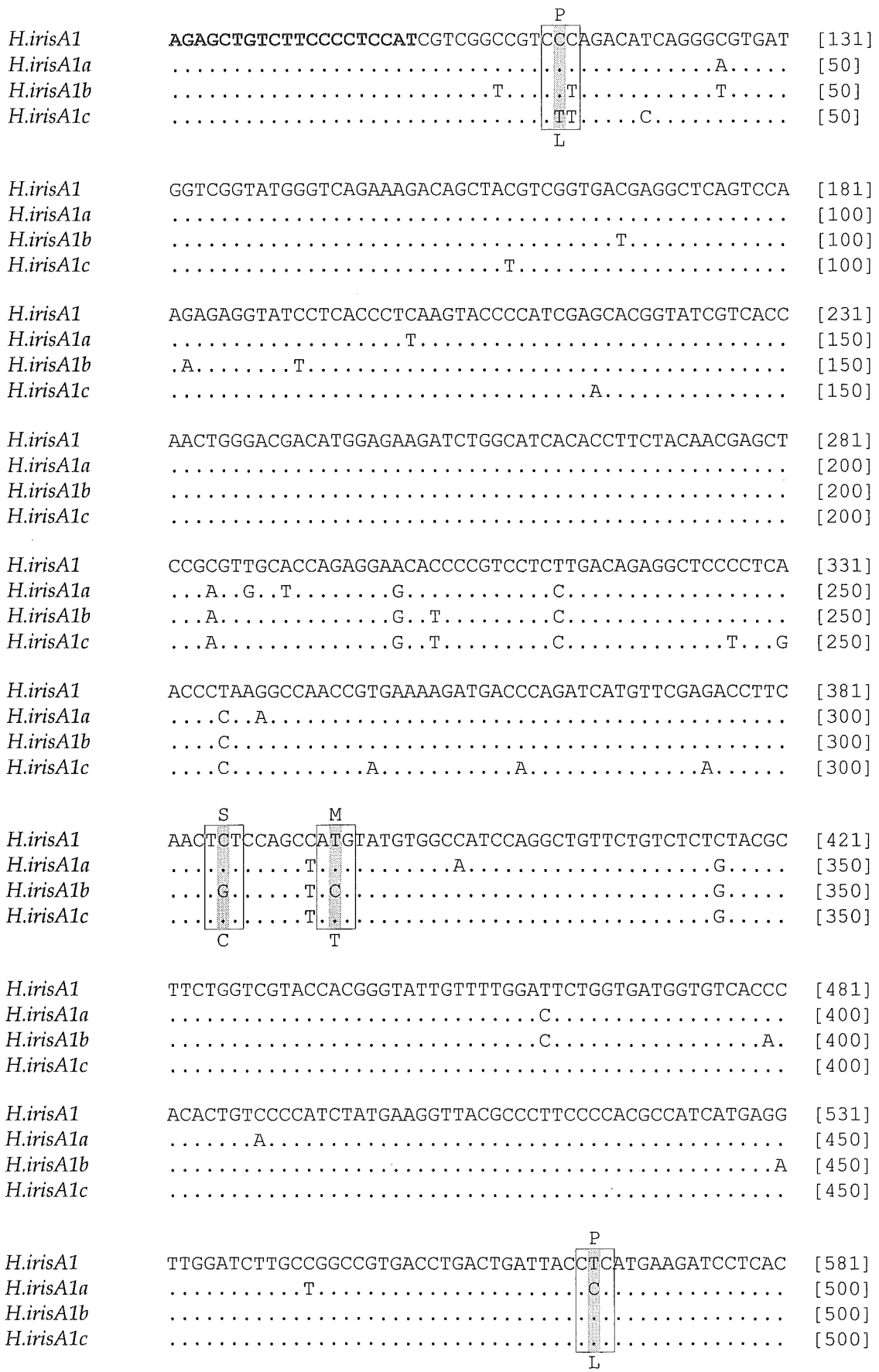
The partial *H.irisA1*-type cDNA sequences were aligned to the *H.irisA1*, *H.irisA2* and *H.irisA3* cDNA sequences. The percentage sequence similarities between *H.irisA1*, *H.irisA2* and *H.irisA3* and the three partial *H.irisA1*-type genes were calculated (the primer derived sequence data was excluded).

Between *H.irisA1* and the three partial *H.irisA1*-type sequences there were 39 variable sites (Figure 2.27). *H.irisA1* and *H.irisA1a* showed 96.5% sequence similarity. There were 19 substitutions, one of which was non-synonymous, changing a leucine at aa 189 to a proline in *H.irisA1a*. *H.irisA1* and *H.irisA1b* showed 96.3% sequence similarity. There were 20 substitutions, two of which were non-synonymous, changing a serine to a cysteine at aa 129 and a methionine at to a threonine aa 132 in *H.irisA1b*. *H.irisA1* and *H.irisA1c* showed 95.9% sequence similarity. There were 24 changes, one of which was a non-synonymous substitution, changing a proline at aa 38 to a leucine in *H.irisA1c*.

When compared to *H.irisA2*, *H.irisA1a*, *H.irisA1b* and *H.irisA1c* showed 89.9%, 90.4% and 90.2% sequence similarity, respectively. There were three non-synonymous substitutions from *H.irisA2* to *H.irisA1a* that changed three amino acids: a proline to a leucine at aa 38, an asparagine to an aspartate at aa 111 and a serine to a threonine at aa 160. There were three non-synonymous substitutions from *H.irisA2* to *H.irisA1b* that changed three amino acids: a serine to a cysteine at aa 129, a methionine to a threonine at aa 132 and a serine to a threonine at aa 160. There were two non-synonymous substitutions from *H.irisA2* to *H.irisA1c* that changed two amino acids: a serine to a threonine at aa 160 and a leucine to a proline at aa 189.

The percentage sequence similarities between *H.irisA3* and *H.irisA1a*, *H.irisA1b* and *H.irisA1c* were 92.4%, 91.9% and 91.3%, respectively. There were five non-synonymous substitutions from *H.irisA3* to *H.irisA1a*, which altered four amino acids: a proline to a leucine at aa 38, an asparagine to an aspartate at aa 111, a leucine to a methionine at aa 176 and a methionine to a leucine at aa 189. There were five non-synonymous substitutions from *H.irisA3* to *H.irisA1b*, which altered four amino acids: a serine to a cysteine at aa 129, a methionine to

a threonine at aa 132, a leucine to a methionine at aa 176 and a methionine to a leucine at aa 189. There were three non-synonymous substitutions from *H.irisA3* to *H.irisA1c* that changed two amino acids: a leucine to a methionine at aa 176 and a methionine to a proline at aa 189.


P
H.irisA1 AGAGCTGTCTTCCCCTCCATCGTCGGCCGTCCCAGACATCAGGGCGTGAT [131]
H.irisA1a .....A..... [50]
H.irisA1b .....T.....T.....T..... [50]
H.irisA1c .....TT.....C..... [50]
L
H.irisA1 GGTCGGTATGGGTCAGAAAGACAGCTACGTCGGTGACGAGGCTCAGTCCA [181]
H.irisA1a ..... [100]
H.irisA1b .....T..... [100]
H.irisA1c .....T..... [100]
H.irisA1 AGAGAGGTATCCTCACCCTCAAGTACCCCATCGAGCACGGTATCGTCACC [231]
H.irisA1a .....T..... [150]
H.irisA1b .A.....T..... [150]
H.irisA1c .....A..... [150]
H.irisA1 AACTGGGACGACATGGAGAAGATCTGGCATCACACCTTCTACAACGAGCT [281]
H.irisA1a ..... [200]
H.irisA1b ..... [200]
H.irisA1c ..... [200]
H.irisA1 CCGCGTTGCACCAGAGGAACACCCCGTCCTCTTGACAGAGGCTCCCCTCA [331]
H.irisA1a ...A..G..T.....G.....C..... [250]
H.irisA1b ...A.....G..T.....C..... [250]
H.irisA1c ...A.....G..T.....C.....T...G [250]
H.irisA1 ACCCTAAGGCCAACCGTGAAAAGATGACCCAGATCATGTTCGAGACCTTC [381]
H.irisA1a ....C..A..... [300]
H.irisA1b ....C..... [300]
H.irisA1c ....C.....A.....A.....A..... [300]
S
M
H.irisA1 AACTCTCCAGCCATGTATGTGGCCATCCAGGCTGTTCTGTCTCTCTACGC [421]
H.irisA1a .....T.....A.....G..... [350]
H.irisA1b ....G.....T.C.....G..... [350]
H.irisA1c .....T.....G..... [350]
C
T
H.irisA1 TTCTGGTCGTACCACGGGTATTGTTTTGGATTCTGGTGATGGTGTCACCC [481]
H.irisA1a .....C..... [400]
H.irisA1b .....C.....A. [400]
H.irisA1c ..... [400]
H.irisA1 ACACTGTCCCCATCTATGAAGGTTACGCCCTTCCCCACGCCATCATGAGG [531]
H.irisA1a .....A..... [450]
H.irisA1b .....A [450]
H.irisA1c ..... [450]
P
H.irisA1 TTGGATCTTGCCGGCCGTGACCTGACTGATTACCTCATGAAGATCCTCAC [581]
H.irisA1a .....T.....C..... [500]
H.irisA1b ..... [500]
H.irisA1c ..... [500]
L

```
H.irisA1     TGAGCGTGGCTACTCCTTCACCACCACCGCTGAGAGAGAAATCGTCAGAG  [631]
H.irisA1a    ..................................................  [550]
H.irisA1b    ......C..T..............T.......................  [550]
H.irisA1c    ..................................................  [550]

H.irisA1     ACATCAAGGAGAAGCTCTGCTACGTCGCTCTCG                   [664]
H.irisA1a    ....................A.GC.C..T.                      [583]
H.irisA1b    ........................C..T.                      [583]
H.irisA1c    ........................C..T.                      [583]
```

Figure 2.27. Alignment of *H.irisA1a*, *H.irisA2b*, *H.irisA1c* and *H.irisA1* cDNA sequences. Identical nucleotides are indicated by dots (.). Codons which encode different amino acids are indicated by a shaded box with alternative amino acids shown (the amino acid for *H.irisA1* is shown above the alignment, the amino acid of the variant partial sequences are shown below the alignment). The nucleotide positions are given to the right, for *H.irisA1* the nucleotide numbers refer to the number within the full cDNA sequence. Primers binding sites are shown in bold. The first and last 20 nt were excluded from percentage sequence similarity calculations, as the sequence was determined from primers.

SOUTHERN HYBRIDISATION

Gene number was investigated by Southern hybridisation using the cloned 954 nt actin sequence as a probe (Flint, 2000). This probe contains an actin coding gene fragment, situated down stream of intron B, which was used to detect *H. iris* actin gene sequences.

Analysis of restriction enzyme cut sites (Appendix E) predicted that *Cla* I and *Pst* I would not cut the known *H. iris* actin genes within the probe binding region, therefore each hybridisation signal within these digests are assumed to represent an individual actin gene. It was predicted that *Eco*R I would cut *H.irisA1*, but within the 5′ non-coding region of the gene, so each hybridisation signal would represent an individual gene. *Hin*d III was predicted to cut *H.irisA1c* upstream of the probe binding region, within intron B, so each hybridisation signal would represent an individual gene.

Hybridisation revealed several restriction products with similarity to *H. iris* actin at a stringency of 0.1 × SSC/0.1% SDS at 65 °C (Figure 2.28). At least nine DNA fragments hybridised to the actin probe. The *Cla* I digest contained 10 strong hybridisation bands and one faint hybridisation band in individuals 1 and 2. The *Eco*R I digest contained eight strong hybridisation bands and two faint hybridisation bands. The *Hin*d III digest contained eight strong hybridisation bands and one faint hybridisation band in individual 3. The *Pst* I digest contained eight strong hybridisation bands in all individuals.

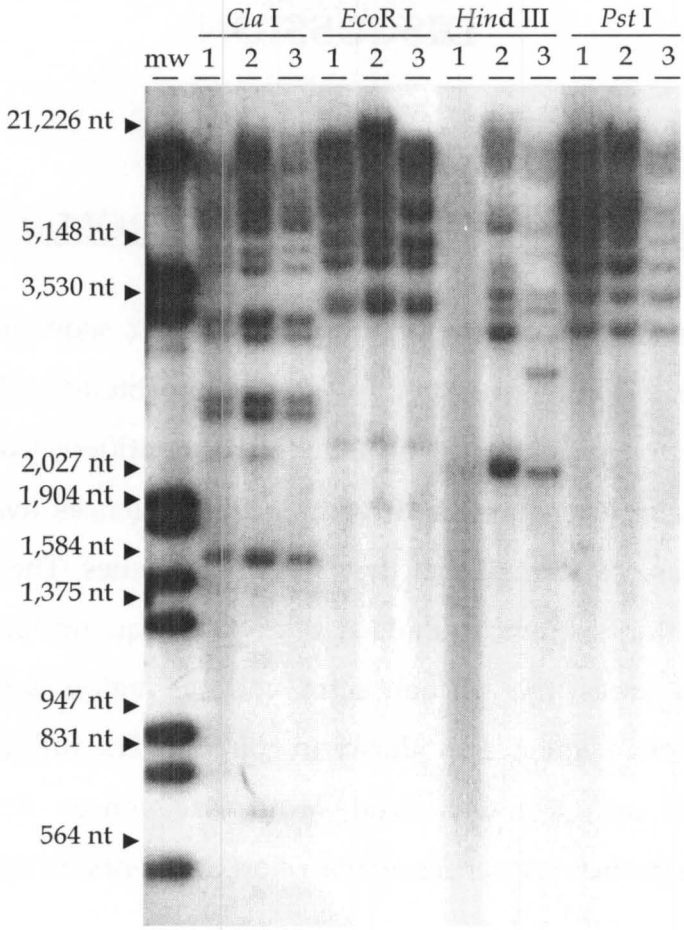


Figure 2.28. Estimation of *Haliotis* actin gene number by Southern hybridisation. The molecular weight marker (mw) was λ genomic DNA digested with *Hin*d III and *Eco*R I. Five micrograms of genomic abalone DNA was digested with *Cla* I, *Eco*R I, *Hin*d III, *Pst* I or *Pvu* II. For each restriction enzyme the three lanes were: 1, *H. iris* individual 1; 2, *H. iris* individual 2; and 3, *H. iris* individual 1. Hybridisation was at 65°C, using a probe corresponding to a generic *H. iris* actin coding region. The blot was washed to a stringency of 0.1 × SSC/0.1% SDS at 65°C.

DISCUSSION

ISOLATION OF ACTIN GENES FROM *H. IRIS*

PCR and sequencing isolated putative actin coding sequences from *H. iris*. Three sequences encoding predicted full size actin proteins and three sequences encoding predicted partial actin proteins were identified. Comparison of the putative *H. iris* actin sequences to actin gene sequences available from the GenBank database confirmed that they were actin genes. The three full length actin coding region sequences showed over 85% sequence similarity to other actin genes, whereas the partial actin coding region sequences showed approximately 96% sequence similarity to *H.irisA1*. The full and partial *H. iris* genes predicted appropriately sized protein sequences that showed high identity to actin protein sequences available on the GenBank database.

CONSERVATION OF ACTIN GENES

As expected from previous analyses of invertebrate actins, the actin gene sequences of *H. iris* showed high sequence similarity to invertebrate and vertebrate actins due to the functional constraint of actin proteins (DesGroseillers *et al.*, 1994; Patwary *et al.*, 1996; Lardans *et al.*, 1997; Gomez-Chiarri *et al.*, 1999).

Three partial actin sequences were isolated by PCR using the *H.irisA1* subtype-specific reverse primer (A1717R1) in combination with the generic primer A989F(T3). The sequences were distinguishable due to changes in the size of intron B, so they could be excised and purified. These sequences were assumed to be related to *H.irisA1* ('*H.irisA1* types') as they were amplified with *H.irisA1* specific primers and were therefore called *H.irisA1a*, *H.irisA1b* and *H.irisA1c*. When the partial cDNA sequences of the putative *H.irisA1* type actins were compared to *H.irisA1*, *H.irisA2* and *H.irisA3* they showed greater

percentage sequence similarity to *H.irisA1* (96.5, 96.3, 95.9%, respectively) than to *H.irisA2* (89.9, 90.4, and 90.2%, respectively) or *H.irisA3* (92.4, 91.9, 91.3%, respectively), suggesting they are more closely related to *H.irisA1* than they are to the *H.irisA2* or *H.irisA3*.

The sequence similarity between the coding regions of *H.irisA1* and the *H.irisA1* subtypes was so high that the partial sequences may not have been detectable if PCR across the intron site had not been performed. *H.irisA2* and *H.irisA3* subtypes were not detectable from PCRs spanning the intron site. If subtypes were present for these genes they may not have been amplified due to greater specificity at the subtype-specific primer binding sites for *H.irisA2* and *H.irisA3* as compared to *H.irisA1*.

In an attempt to obtain more sequence data for the *H.irisA1* subtype genes, PCR using A1717R1 and *H.irisA1* 5' UTR based primers was tried. PCR failed to amplify PCR products representing the *H.irisA1* subtype genes. Amplification with A266F/A1717R1 did not amplify *H.irisA1* subtype products, presumably due to divergence in the *H.irisA1* subtype genes' 5' UTRs which prevented A266F from binding. A623F/A1717R1 produced faint products from 50.3 to 58.1°C, but these products were not of the expected sizes. These PCR products were not analysed further, although it is possible that these faint products could represent *H.irisA1* subtype PCR products that have increased in size due to the expansion in the 5'UTR (isolation and sequencing of the extra PCR products would confirm this). The ATG based primers A911F and ATG(T3) were not tried, but probably would work due to high conservation in the coding regions of actin genes. Alternatively, the sequence information available for *H.irisA1a*, *H.irisA1b* and *H.irisA1c* could be extended by PCR with primers designed to regions of high diversity within the introns of the respective genes (this approach would be the only way to obtain sequence data downstream of the intron).

H. IRIS ACTIN GENE NUMBER

Southern hybridisation confirmed that there is a large actin gene family in *H. iris*, which is composed of more than the six genes sequenced during this research. Hybridisation banding patterns suggest the actin gene family of *H. iris* contains at least nine members. This number is in agreement with estimates of gene number from other molluscs, there are greater than four to seven genes in *Patella vulgata* (van Loon *et al.*, 1993), three to five genes in *Aplysia californica* (DesGroseillers *et al.*, 1994), at least five genes in freshwater pulmonate snails (Adema, 2002) and between 12-15 genes in *Placopecten magellanicus* (Patwary *et al.*, 1996).

Non-specific PCR products may be evidence of uncharacterised *H. iris* actin genes. The *H.irisA1* A266F/A1717R1 PCR (Figure 2.7A) amplified a strong product that was smaller than the expected amplicon and amplified non-specific PCR products that persist through the annealing temperature gradient and amplify (faintly) at the highest annealing temperature (58.1°C). The *H.irisA1* A623/A1717R1 PCR (Figure 2.9A) also shows non-specific PCR products which amplify throughout the annealing temperature gradient. These non-specific products could represent amplicons from other actin genes, like the extra PCR products from the A989F(T3)/A1717R1 PCR which led to the discovery of *H.irisA1a*, *H.irisA1b* and *H.irisA1c*.

Further evidence of another *H. iris* actin gene may be provided by the comparison of the *H.irisA1* sequence to the 381 nt sequence of Flint (2000) (Figure 2.16). The PCR primers used to sequence the *H.irisA1* gene were designed from the 381 nt sequence and were expected to amplify a full coding region sequence of the partial 381 nt sequence. Alignment of the two sequences suggests that they may not represent the same sequence as they differ by 7.6%, whereas the sequences of *H.irisA2* and *H.irisA3* show less than 1% difference to the corresponding partial sequences of Flint and therefore probably do represent the same genes.

H.irisA1 and the *H. rufescens* actin were compared as 5′ non-coding sequence data was available for both sequences. These sequences showed high sequence similarity between the 5′ non-coding regions, between 76.2 and 78.6%. The intron sequences showed on average 52.3% sequence similarity, but contained blocks of 100% alignment (up to 18 nt in length). The high sequence similarity in the non-coding regions suggests that these genes are closely related and may represent actins of the same type within different species (orthologous genes).

Of the multiple genes within *H. iris*, some are assumed to be cytoplasmic actins while others are assumed to be muscle actins, however, assignment to these groups requires functional analysis.

ACTIN INTRONS

Isolation of *H. iris* actin genes from genomic DNA showed that the six genes identified contained an intron between codon positions 41/42. Although the intron positions were conserved, the size and the sequences of the introns were very divergent. The intron sizes ranged from 174 nt in *H.irisA1*, to 1,078 nt in *H.irisA2*, to 581 nt in *H.irisA3*. The introns of *H.irisA1a*, *H.irisA1b* and *H.irisA1c* were similarly sized to *H.irisA1* (301 nt, 282 nt and 229 nt, respectively, compared to 174 nt), but could not be aligned due to sequence divergence.

The size similarity between the introns of these sequences may represent recent common ancestry; the introns of *H.irisA1* subtype genes may have diverged quickly by substitution, but may be diverging more slowly by expansion. Similar intron sizes are found in the closely related actin genes of *Strongylocentrotus franciscanus* (Foran *et al.*, 1985). The *S. franciscanus* genes, *SfA 15A* and *Sfa 15B*, are thought to have arisen from a recent duplication event as they show high sequence similarity and are closely linked on the same chromosome. The intron lengths of these genes are 236 and 224 nt at intron 1, and 172 and 191 nt at intron 2, respectively.

Where data is available, all mollusc actin sequences contained an intron between aa 42/43 (or the equivalent position aa 41/42 for *H. rufescens* and *H. iris* actins). This intron position is also found in the echinoderms *P. ochraceus* and *S. purpuratus*. Comparison of actin genes from a range of organisms has indicated that this position is a commonly occurring intron site that is present in deuterostome and protostome actin genes (Weber and Kabsch, 1994).

There are two theories that attempt to explain the role of introns in genes. The 'intron early' hypothesis suggests introns are the remnants of ancestral gene sequences and that intron pattern diversity is the result of intron loss in lineages, whereas the 'intron late' hypothesis suggests introns arose during the evolution of genes (Dibb and Newman, 1989; Nyberg and Cronhjort, 1992). The small number of introns in mollusc actin genes, considered to be a primitive invertebrate, does not support the hypothesis of an ancient origin of introns.

STRUCTURE OF THE *H. IRIS* ACTIN GENE FAMILY

The actin gene sequences of *H. iris* appear to represent three subtypes based on sequence divergence. Data from other molluscs suggests there are three actin subtypes in cephalopods (Carlini *et al.*, 2000) and four actin subtypes in *Patella vulgata* (van Loon *et al.*, 1993).

Organisation of actin families into subtypes has been documented for invertebrates, such as dipterans, echinoderms and molluscs. In dipterans actins are sorted by functional analysis, two genes form a cytoplasmic group, four genes form two muscle groups which differ by expression patterns (Fyrberg *et al.*, 1983). In the mollusc *P. vulgata* (van Loon *et al.*, 1993) and in the sea urchin *S. purpuratus* (Lee *et al.*, 1984) four actin subtypes are discernible by relatedness between the 3'UTRs of the genes. In cephalopods, actin gene sequences separate into three actin groups during phylogenetic analysis (Carlini *et al.*, 2000).

As for the sea urchin *S. purpuratus* and the limpet *P. vulgata*, multiple genes within actin subtypes appear to exist in *H. iris*. Within the *H.irisA1*

subtype there is evidence of four genes, *H.irisA1*, *H.irisA1a*, *H.irisA1b* and *H.irisA1c*.

Confirmation of *H. iris* subtypes and the number of genes within each actin subtype requires further characterisation. Isolation and sequencing of the 3′ UTRs of *H. iris* genes would provide subtype-specific probes which could be used to screen *H. iris* genomic DNA Southern blots and confirm *H. iris* actin subtype organisation. In the absence of *H. iris* actin 3'UTR sequences, expression analysis and phylogenetic analysis of the actin gene sequences could confirm the three putative *H. iris* actin subtypes.

The existence of actin subtypes and the sequence divergence between actin subtypes may allow the actins of each subtype to perform specialised functions.

ACTIN PROTEINS

H. iris actin genes differed from other actin genes by predicting proteins of 375 aa rather than the more usual 376 aa. The only other actin sequences that predict a 375 aa protein are from abalone *H. rufescens* and the limpet *P. vulgata*. Although the *H. rufescens* actin gene sequence is only a partial one, it does not contain the second amino acid (usually a cysteine or glycine) found in other mollusc actin proteins (Gomez-Chiarri *et al.*, 1999). The other haliotid actin sequence available, that of *H. discus hannai*, encodes a protein of 376 aa, dismissing the conclusion that the haliotid group have a divergent protein.

The amino acid sequences at the beginning of *H.irisA2* and *H.irisA3* were determined from the primers used to amplify them, which dictated a predicted protein of 375 aa as the primers were based on *H. rufescens*. However, the *H.irisA1* sequence was determined from PCR products that spanned the beginning of the coding region, so the missing second amino acid was determined from the gene sequence. The amino-terminal part of the actin protein is responsible for myosin binding, so the absence of the second amino acid in the *H. iris*, *H. rufescens* and *P. vulgata* actins may reflect co-evolution

with the myosin protein of these species. Also, as the 5′ sequences of *H.irisA2* and *H.irisA3* were determined by the primer sequences used to amplify these genes, the possibility that differences in these parts of the sequence cannot be overlooked. As an important myosin binding site is found in the amino-terminus of the actin protein, changes in this region may exist in response to actin functional specialisation. This possibility cannot be confirmed without sequencing across the *H.irisA2* and *H.irisA3* genes.

Chapter 3:

EXPRESSION OF THE *HALIOTIS IRIS* ACTIN GENE FAMILY

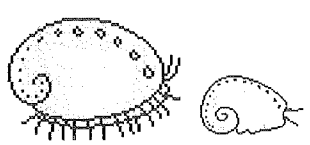


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

This chapter will study the expression of *H.irisA1*, *H.*irisA2 and *H.irisA3* during *Haliotis iris* development and in adult *H. iris* tissues. The differential expression of *H. iris* actin genes will be compared to the known differential developmental expression patterns of other invertebrates. A preliminary study of the expression of *H. iris* actin genes in adult tissues will be used to investigate the presence of cytoplasmic and muscle type actins in molluscs.

ACTIN GENE EXPRESSION

The expression of actin genes is of interest for two reasons. Firstly, investigating the expression of multiple actin genes within an organism may help explain the functional significance of actin gene families. Actin gene families are thought to have evolved in response to a high demand for the actin protein, but also to allow functional specialisation of family members (Rubenstein, 1990). Expression patterns may support these hypotheses. Secondly, in invertebrates analysis of the expression of multiple actin genes is the best way to determine the actin functional type (cytoplasmic or muscle) (Fyrberg *et al.*, 1983; White and Crother, 1999).

The multiplicity of actin genes with the same expression pattern suggests high levels of the actin protein are required. In *Drosophila melanogaster* there are six actin genes, which form three pairs, that show co-expression in particular tissues (Fyrberg *et al.*, 1983). Pairs of co-expressed genes are also found in the sea urchin, *Strongylocentrotus purpuratus* (Cox *et al.*, 1986).

Functional specialisation between actins is supported by the different binding affinities and kinetics of actins and by the specific localisation of actins within cells and tissues (Reisler, 1993; Sheterline and Sparrow, 1994). Mounier and Sparrow (1997) mapped mammalian muscle-specific amino acid residues to the three dimensional structure of actin and found that the majority of amino

acid changes were on the external surface of the actin protein. This suggests that gene-specific amino acid sequences may facilitate isoform-specific interactions with actin-binding proteins. They also showed that the N-terminus, which is involved in myosin binding, is different between muscle and cytoplasmic forms and that these differences may lead to a rigid N-terminus in muscle actins which may be better suited to binding muscle myosin. A study which compared the ability of non-muscle and muscle actins to activate myosin ATPase showed that non-muscle actins were 75% less effective than muscle actins (Gordon *et al.*, 1977).

In lower invertebrates the sequence of actin genes are highly conserved; cytoplasmic and muscle forms cannot be distinguished based on sequence alone, as they can be in insects and vertebrates (Mounier *et al.*, 1992). The only way to determine actin type in these cases is to analyse expression patterns, cytoplasmic actins will be ubiquitously expressed whereas muscle actins will be restricted to tissues containing muscle.

DIFFERENTIAL DEVELOPMENTAL EXPRESSION

Differential expression and functional specialisation has been demonstrated in several invertebrate species, including the insects *Drosophila, Bactrocera dorsalis* and *Bombyx mori*, the echinoderm *S. purpuratus*, and the mollusc *Patella vulgata.* In insects, muscle actins show differential expression, whereas cytoplasmic actins were expressed in all tissues at all developmental stages. In echinoderms cytoplasmic and muscle actin genes show differential expression.

Insects

Fyrberg *et al.* (1983; 1984) investigated *Drosophila* actin gene expression by northern hybridisation of RNA gels and RNA dot blots using actin gene specific 3′ UTR probes. Fyrberg *et al.* (1983) deduced actin function from expression patterns (Table 3.1). They found *act57B* and *act87E* were expressed concurrently at very low levels during larval stages, with expression levels increasing during late pupal development. In the late pupae *act57B* and *act87E* expression was high in tissues comprised of restructured larval muscle, the head, abdomen and

abdominal wall. *Act87E* was also expressed in the leg and faintly in the ovary. The expression pattern of these genes correlates to the differentiation and restructuring of larval musculature to form adult muscle.

These authors further found that *act79B* and *act88F* transcripts occurred at high levels during mid to late pupal stages and remained high in adults. The timing and location of *act79B* and *act88F* correlated with differentiation of adult leg and thoracic musculature. *Act79B* was also expressed at high levels in the legs of late pupae. *Act79B* and *act88F* were the only transcripts detected in the late pupal thorax, which was mainly composed of the indirect flight muscles.

The expression patterns in *Bactrocera* and *Bombyx* were also examined and used to infer actin gene function in these species (Table 3.1). The four isolated genes of *Bactrocera, BdA1, BdA2, BdA3* and *BdA5*, showed expression patterns consistent with these genes encoding muscle-type actins. Two genes of *Bombyx, A1* and *A2*, showed expression patterns consistent with these genes also encoding muscle-type actins.

Table 3.1. Differential expression patterns in Diptera and Lepidoptera. The expression patterns and the inferred actin functional type are given.

Species	Gene	Expression Pattern	Actin Type
Drosophila melanogaster	*act5C*	All larval stages All adult tissues	cytoplasmic
	act42A	All larval stages All adult tissues	cytoplasmic
	act57B	Larval stages, late pupal Adult head, abdomen, abdominal wall	muscle
	act87E	Larval stages, late pupal stages Adult head, abdomen, abdominal wall, ovary	muscle
	act79B	Mid-late pupal stages Adult leg and thorax	muscle
	act88F	Mid-late pupal stages Adult leg and thorax	muscle
Bactrocera dorsalis	*BdA1*	Late pupal stages Adult thorax	muscle
	BdA2	Late larval stages, late pupal stages Adult head, abdomen, leg, thorax	muscle
	BdA3	Late pupal stages Adult thorax and leg	muscle
	BdA4	Late larval stages, late pupal stages Adult head, thorax, leg	muscle
Bombyx mori	*A1*	Larval midgut, body wall and gut, late pupal stages Adult thorax and abdomen	muscle
	A2	Late pupal stages Adult thorax	muscle

Echinoderms

Expression of *S. purpuratus* actin genes was investigated by probing embryonic and adult tissue RNA blots with 3′ UTR gene-specific probes. Hybridisation analysis showed that *CyI, CyIIa, CyIIb* and *M* are expressed in the embryo and adult tissues, whereas *CyIIIa* and *CyIIIb* expression is confined to the embryo. Results indicated that the linkage group *CyI, CyIIa* and *CyIIb* are not co-ordinately expressed (Shott *et al.*, 1984).

The spatial pattern of *S. purpuratus* embryonic expression was analysed by *in situ* hybridisation of embryos (Cox *et al.*, 1986). *M* actin appeared late in development and was confined to cells associated with the forming coelom. The cytoplasmic genes form three groups which show distinct expression patterns. *CyI* and *CyIIb* are mainly expressed in the oral ectoderm and parts of the gut of the pluteus larva. *CyIIa* is only expressed in the mesenchyme lineages throughout the late gastrula and accumulates in parts of the gut. *CyIIIa* and *CyIIIb* are expressed in aboral ectoderm. After the late gastrula stage *CyIII* genes are the only detectable actins.

Molluscs

In *Patella* the expression of three actin subtypes, *pPA2, pPA3* and *pPA17,* was investigated by probing poly(A) RNA from different developmental stages with 3′ UTR probes specific for each actin subtype (van Loon *et al.*, 1993). The expression level was determined by densitometry of the hybridisation signal and was plotted over time. *Patella* actin subtypes showed diverse expression patterns. pPA2 was almost undetectable following fertilisation until 10 h post-fertilisation, after which expression levels increased rapidly. pPA3 was moderately expressed following fertilisation, with an increase in expression occurring after 10 h post-fertilisation. pPA17 expression was almost undetectable until 16 h post-fertilisation, after which expression levels increased steadily.

ACTIN FUNCTIONAL TYPE

Expression patterns are used to infer actin functional type. The functional division of actins is supported by muscle-specific amino acids in certain regions of the actin molecule found in insect actins. Muscle-specific amino acids are not found in other invertebrates.

In insects, actin type was characterised by expression patterns. Cytoplasmic gene expression was found through development and in all tissues examined. Muscle gene expression was correlated with the development of musculature in larvae and adults; in adults, expression occurred in tissues containing muscular structures, such as the abdomen, thorax and legs. In *Drosophila, act5C* and *act42A* were found to be cytoplasmic actins, whereas *act57A, act87E, act79B* and *act88F* were found to be muscle actins (Fyrberg *et al.*, 1983; Fyrberg, 1984). In *Bactrocera* the four identified genes showed muscle-specific expression (He and Haymer, 1995). In *Bombyx* expression analysis identified *A3* as a cytoplasmic actin, *A1* as a larval muscle actin and *A2* as an adult muscle-specific actin (Mounier and Prudhomme, 1991; Mounier *et al.*, 1991).

Expression analysis of actin genes in echinoderms identified one muscle gene and multiple cytoplasmic genes in the sea urchin *S. purpuratus* (Lee *et al.*, 1984) and one muscle gene and one cytoplasmic gene in the sea star *Pisaster ochraceus* (Kowbel and Smith, 1989).

In molluscs, the actin gene of *Placopecten magellanicus* is considered a muscle actin as it was isolated from an adductor muscle cDNA library (Patwary *et al.* 1996), although the expression of *Placopecten* actin in non-muscle tissues was not investigated. In *Patella* the ubiquitous expression of one actin suggests it is a cytoplasmic gene, whereas the expression of the second actin gene in the muscular sheath of neurons suggests that it is a muscle actin (DesGroseillers *et al.*, 1990; DesGroseillers *et al.*, 1994).

In cephalopods, actin type has been inferred from phylogenetic analysis of sequences (Carlini *et al.*, 2000). Carlini *et al.* (2000) showed that cephalopod actins form three clades and suggested that these represent functionally distinct

actins. The *Actin I* type genes were identified as muscle actins as they grouped with *Placopecten* and *Aplysia* muscle actins, *Actin II* genes were identified as cytoplasmic actins as they grouped with *Aplysia* cytoplasmic actin, whereas *Actin III* type genes were thought to be pseudogenes as they grouped distinctly from other actins and had low codon bias (which is thought to reflect the lack of neutral evolution in non-functional genes). The functional classifications made by Carlini *et al.* (2000) were not confirmed by expression studies.

AIMS

The expression of actin can identify functionally specialised types that may clarify the presence of multiple actins in a species. In insects and echinoderms actins can be classified as cytoplasmic and muscle types depending on function, and are differentially expressed. In molluscs the presence of muscle actins has been inferred, but has not been well studied. The aims of this study are to investigate actin expression in *H. iris* and to determine whether expression of actin can give clues to their functional specialisation in *H. iris*.

The specific aims of this chapter are to:

1. Show that *H.irisA1, H.irisA2* and *H.irisA3* genes are functional
2. Show that *H.irisA1, H.irisA2* and *H.irisA3* are differentially regulated during larval development
3. To examine *H. iris* actin gene expression in adult tissue to investigate actin type

METHODS

LARVAL SAMPLE COLLECTION

H. iris larval samples were collected from an abalone spawning facility at Pacifica Fisheries Ltd (Kaikoura, South Island, New Zealand). Four developmental stages were sampled: fertilised egg (1 h post-fertilisation), blastula embryo (10 h post-fertilisation), trochophore larvae (20 h post-fertilisation) and veliger larvae (40 h post-fertilisation) (Tong and Moss, 1992). Samples were snap frozen in liquid nitrogen and stored at -80°C until required.

RNA EXTRACTION

RNA was purified from larval samples using TriReagent (Molecular Research Center Inc, Progenz Limited). Tissue was homogenised in 10 volumes of TriReagent and incubated at room temperature for 5 min to allow cellular dissociation. RNA was separated from DNA and protein by extraction with chloroform and purified by precipitation with a half volume of isopropanol. RNA samples were dissolved in diethyl pyrocarbonate (DEPC)-treated water.

RNA was treated with RNase-free DNase to remove contaminating genomic DNA. Reactions contained 10 mM Tris pH 7.5, 6 mM $MnCl_2$ and 1.5 U DNase I (Roche) and were incubated at 37°C for 1 h. Reactions were stopped by the addition of 10 mM EDTA, 0.2% SDS and 0.3 M sodium acetate (pH 5.2). The RNA was purified by phenol/chloroform extraction, followed by precipitation with two volumes of cold ethanol. RNA samples were dissolved in DEPC-treated water. Samples were stored in two volumes of ethanol at -20°C and precipitated with 0.3 M sodium acetate (pH 5.2) prior to use.

The concentration, yield and purity of RNA were determined by spectrophotometry at absorbances of 260 and 280 nm (using a Ultrospec II, LKB Biochrom Ltd).

RT-PCR

The expression of actin genes was detected by RT-PCR. Complementary DNA (cDNA) was synthesised from total RNA using SuperScript II RNase H$^-$ Reverse Transcription (Invitrogen). Five micrograms of RNA was mixed with 500 ng of oligo $(dT)_{12-18}$ and made up to volume with deionised/distilled water. The RNA mix was denatured at 70°C for 10 min then chilled on ice. First strand buffer (50 mM Tris-HCl (pH 8.3), 75 mM KCl, 15 mM $MgCl_2$), 10 mM dithiothreitol and 2 mM dNTPs were added to the RNA. The reaction was heated at 42°C for 2 min before 200 U of SuperScript II was added. Reverse transcription was allowed to proceed at 42°C for 50 min and was stopped by heating at 70°C for 15 min.

H.irisA1, *H.irisA2* and *H.irisA3* cDNA was detected by PCR actin gene specific primers. Expression of *H.irisA1a*, *H.irisA1b* and *H.irisA1c* could not be performed as the sequences were too similar for specific primers to be designed. As the *H.irisA1* specific primer, A1717R1, also amplifies *H.irisA1a*, *H.irisA1b* and *H.irisA1c*, analysis with this primer includes '*H.irisA1* subtype' expression, rather than solely *H.irisA1* expression.

PCR was performed as described in Chapter 2 (page 38). Gene-specific reverse primers were used to amplify *H.irisA1*, *H.irisA2* and *H.irisA3*; *H.irisA1* was amplified with A1F and A1717R1 at 62.2°C, *H.irisA2* was amplified with A1F and A1719R2 at 54.4°C and *H.irisA3* was amplified with A1F and A1721R3 at 62.2°C (primer details are given in Table 2.2). The internal control was amplification of 18S rRNA with 18S rRNA F and 18S rRNA R at 68.1°C (primer details in Chapter 2, page 37).

RT-PCR products were analysed by electrophoresis on a 1.5-2% agarose/0.5 × TBE gels and visualised by ethidium bromide staining (0.5 µg/ml) and transillumination with UV (300 nM, UltraLum).

DEVELOPMENTAL EXPRESSION

Temporal expression during larval development of *H. iris* was examined by RT-PCR of RNA purified from fertilised eggs, blastula, trochophore larvae and veliger larvae. Following electrophoretic analysis, RT-PCR products were transferred to Hybond-XL membrane by Southern hybridisation (see Chapter 2, page 43 for details).

The specificity of the RT-PCR assay was confirmed by Southern hybridisation of the RT-PCR membrane with actin gene-specific probes. The gene-specific probes were purified PCR products, which were obtained by amplification using gene-specific forward primers (A1686F1, A1686F2 or A1686F3 in combination with A3R). The probes share 32-36 nt overlap with the gene-specific reverse primers used in the RT-PCR assay, over a region of high diversity between the three genes (nt 806-838, nt 1,710-1,744 and nt 1,213-1,249 of *H.irisA1*, *H.irisA2* and *H.irisA3*, respectively), which allows the probes to hybridise to specific actin sequences.

ADULT EXPRESSION

Preliminary analysis of actin gene expression in adult tissues was determined by RT-PCR using RNA extracted from adult gill, gonad, hepatopancreas, foot and mantle. Adult *H. iris* RNA was kindly donated by Alice Johnstone. This RNA was not DNase I treated, but PCR with RNA templates confirmed the absence of contaminating DNA (data not shown).

RESULTS

RNA EXTRACTION

Total RNA was isolated from egg, blastula, trochophore larvae and veliger larvae. Yields ranged from 0.47-3.23 µg/mg of starting tissue. RNA purity was indicated by the A_{260}/A_{280} ratios, which were between 1.8 and 2 for different preparations using the manufacturer's protocols. Electrophoresis showed that RNA ranged in size from 100-2,800 nt. Highly expressed RNA molecules were evidenced by the presence of distinct bands within mRNA smears. The number of bands, sizes of bands and band intensities in RNA varied between developmental stages.

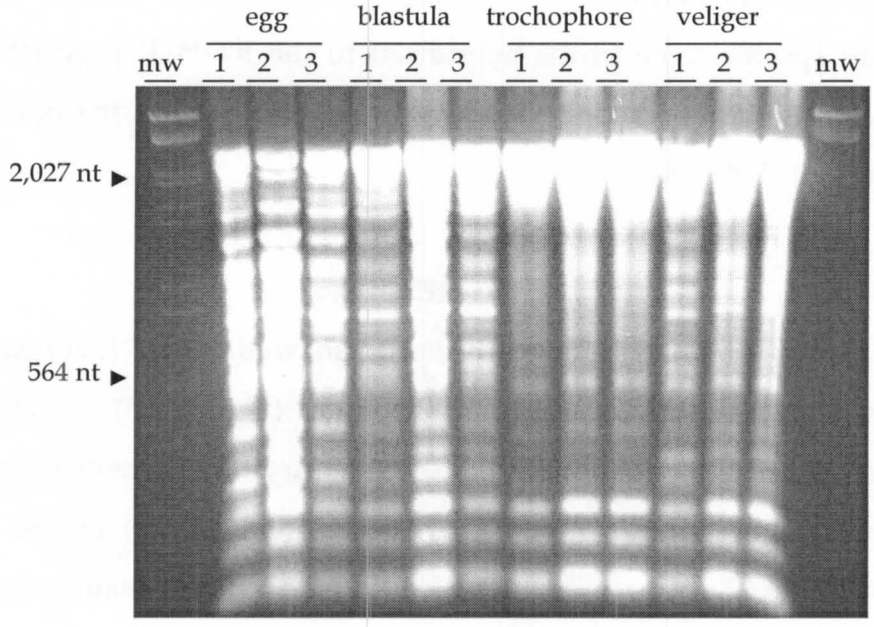


Figure 3.1. *H. iris* RNA. RNA was analysed by electrophoresis through a 1.75% agarose/0.5 x TBE buffer gel. The molecular weight markers (mw) were λ genomic DNA digested with *Eco*R I and *Hin*d III.

RT-PCR

DEVELOPMENTAL EXPRESSION

H.irisA1 (Figure 3.2A and 3.2B) and *H.irisA2* (Figure 3.2A and 3.2C) showed similar expression patterns. Both genes were expressed in all stages, ranging from low levels in the fertilised egg and blastula stages, to high levels in the trochophore and veliger larval stages. *H.irisA3* (Figure 3.2A and 3.2D) was not detected in the fertilised eggs, but was expressed at low levels in the blastula (detected by Southern hybridisation) and at high levels in the trochophore and veliger larvae.

The specificity of the RT-PCR was confirmed by Southern hybridisation using *H. iris* gene specific probes. The hybridisation probes overlapped each of the RT-PCR products by only 32-36 nt, over a region of high diversity (nt 806-838, nt 1,710-1,744 and nt 1,213-1,249 of *H.irisA1*, *H.irisA2* and *H.irisA3*, respectively). Hybridisation with gene specific probes showed that the RT-PCR products were specific, each probe hybridised to the RT-PCR product of the same gene, with no cross-reactivity to the RT-PCR products of the other actin genes (Figure 3.2B, 3.2C and 3.2D).

ADULT EXPRESSION

Preliminary expression analysis in adult tissues showed *H.irisA1*, *H.irisA2* and *H.irisA3* were expressed differently in adult tissues (Figure 3.3). *H.irisA1* was expressed in all tissues examined, though at varying intensities between developmental stages when compared to the internal control of 18S rRNA amplification. *H.irisA2* was expressed in the gonad, foot and mantle, and was expressed very weakly in gill. *H.irisA2* was not expressed in the hepatopancreas. *H.irisA3* was expressed in the foot and weakly in the gonad and mantle. *H.irisA3* was not expressed in the gill or hepatopancreas. Further analysis by Southern hybridisation with gene-specific probes is required to confirm absence of expression.

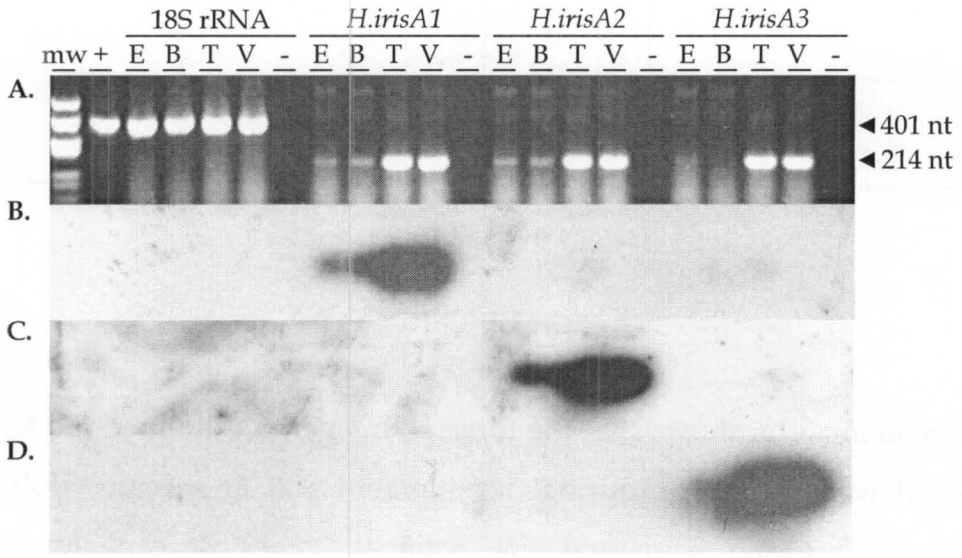


Figure 3.2. Temporal expression of three *H. iris* actin genes during larval development. *H. iris* actin expression was determined by RT-PCR in four developmental stages, fertilised egg (E), blastula (B), trochophore larvae (T) and veliger larvae (V). Specificity of RT-PCR was confirmed by Southern hybridisation of gene specific probes to RT-PCR products. Gene probes overlapped the PCR products by 32-36 nt over a region of high diversity. Post-hybridisation washes were at 65°C to a stringency of 0.05 × SSC/0.1% SDS. **A. Electrophoresis.** RT-PCR products were analysed on a 2% agarose/0.5 × TBE gel. The molecular weight marker (mw) was pBluescript M13+ genomic DNA cut with *Hae* III. The positive control (+) was amplification of 18S rRNA from *H. iris* genomic DNA. Amplification of 18S rRNA from larval samples was used as an internal standard. For each primer pair, amplification of ddH_2O was used as a negative control (-). **B. Hybridisation with an *H.irisA1* gene-specific probe. C. Hybridisation with an *H.irisA2* gene-specific probe. D. Hybridisation with an *H.irisA3* gene-specific probe.**

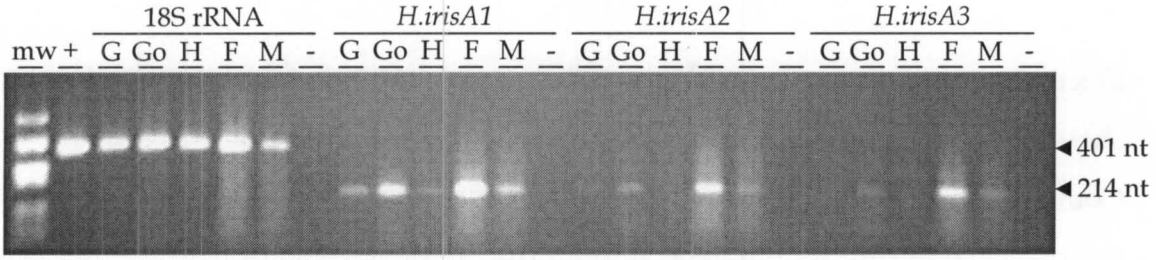


Figure 3.3. Spatial expression of three *H. iris* actin genes in adult tissues. *H. iris* actin expression was determined by RT-PCR in five adult tissues, gill (G), gonad (Go), hepatopancreas (H), foot (F) and mantle (M). RT-PCR products were analysed on a 1.5% agarose/0.5 × TBE buffer gel. The molecular weight (mw) marker was pBluescript M13+ genomic DNA digested with *Hae* III. The positive control (+) was amplification of 18S rRNA from *H. iris* genomic DNA. Amplification of 18S rRNA from larval samples was used as an internal standard. For each primer pair, amplification of ddH_2O was used as a negative control (-).

DISCUSSION

ACTIN FUNCTION

Expression analysis shows that the three *H. iris* genes examined are functional genes, all were expressed during development and in specific adult tissues. Carlini *et al.* (2002) suggested the *Actin III* genes of cephalopods were pseudogenes, based on distinct phylogenetic grouping and low codon bias. There is no evidence of pseudogenes in *H. iris*. The function of only three *H. iris* actin genes was examined by expression analysis, however, codon usage and sequence conservation of *H.irisA1a, H.irisA1b* and *H.irisA1c* does not suggest loss of functional selection in genes, as would be expected for pseudogenes (see Chapter 4).

The redundant expression of *H.irisA1* and *H.irisA2* during development suggests that these two genes may be co-expressed in *H. iris*, as observed in actin genes of insects and *S. purpuratus* (Fyrberg *et al.*, 1983; Cox *et al.*, 1986). This redundant co-expression may reflect a mechanism to increase the level of actin protein available.

H. IRIS DIFFERENTIAL EXPRESSION

Electrophoretic analysis of *H. iris* RNA from various developmental stages shows that there are a number of genes expressed throughout development. Comparison of RNA patterns from different developmental samples indicates the presence of differential gene activity during development. Differential gene expression was evidenced by changes in the number and size of bands expressed, but also in the changing relative intensities of bands expressed in all developmental stages.

The actin genes of *H. iris* display differential expression; the presence and degree of expression changes during development and the sites of

expression changes in adults. Similar differential expression patterns have been documented in insects, echinoderms and molluscs and may reflect functional specialisation of actin genes in response to changing physiological conditions during development or different physiological conditions in tissues.

The expression levels of the putative cytoplasmic actins, *H.irisA1* and *H.irisA2*, vary during development. Fertilised eggs and blastula express these actin genes at low levels, whereas the trochophore and veliger larval stages express actin at a much higher levels. The pattern of differential expression observed for these genes is probably related to developmental changes occurring in *H. iris*. The increase in actin gene expression during *H. iris* development is correlated with the development of larval forms. Larvae undergo drastic morphological changes, developing feeding and locomotory cilia in the trochophore stage, and forming adult body plans in the veliger stage (Tong and Moss, 1992). These structural changes would explain an increased demand for actin.

H.irisA3 is expressed in the trochophore and veliger larval stages. This pattern of expression may be linked to muscle formation in these stages. The putative correlation between *H.irisA3* expression and myogenesis may suggest that *H.irisA3* is a muscle-type actin. In *Drosophila* and *Bactrocera* the expression of muscle genes is correlated with the formation of muscle in the late pupal stage and continues in adult muscle tissues (Fyberg *et al.* 1983; Fyrberg 1984). *H.irisA3* expression also continues in muscular adult tissues, the foot and mantle. If *H.irisA3* is a muscle-type actin gene, differential expression of *H.irisA3* may reflect a conservation mechanism that ensures *H.irisA3* is expressed only when and where needed.

The trends in *H. iris* actin gene expression are similar to those found in *Patella* (Van Loon *et al.*, 1993). In *Patella* the expression of actin genes increases as development progresses, with some actin genes not expressed until 10 hr post-fertilisation. Expression of *pPA3* was detected from fertilisation and increased at 10 post-fertilisation. This expression pattern is similar to the expression patterns of *H.irisA1* and *H.irisA2*. Expression of *pPA2* and *pPA17*

was detected after 10 and 16 h, respectively. These expression patterns are similar to that of *H.irisA3*, whose expression was not detected until the trochophore larval stage (16-24 h post-fertilisation).

In *S. purpuratus*, *CyI* and *CyIIa*, *CyIIb* and *M* are expressed in embryonic and adult tissues, whereas *CyIIIa* and *CyIIIb* are expressed in the embryo only. The *H.irisA1*, *H.irisA2* and *H.irisA3* genes were expressed in embryonic and adult stages, no evidence of larval-specific actin gene expression was found. The presence of larval-specific actin genes requires further investigation, by analysis of the expression of the partially characterised actin genes, *H.irisA1a*, *H.irisA1b* and *H.irisA1c*, or by screening larval cDNA for the presence of currently unidentified actin genes (by cDNA library screening or PCR of cDNA with degenerate *H. iris* actin primers).

Further analysis of spatial expression is required to determine the fine details of differential actin function in *H. iris*. *In situ* hybridisation of whole mount larvae was used to study the spatial pattern of expression in *S. purpuratus* and *Patella* (Cox *et al.*, 1986; Van Loon et al., 1993). This technique could be used to investigate specific functions for *H.irisA1* and *H.irisA2* in *H. iris* larvae. The relative levels of actin gene expression could be measured using real-time PCR.

The differential expression of specific actins may be facilitated by transcriptional regulation. In *S. purpuratus* the spatial and temporal expression of *CyIIIa* is regulated by binding of at least nine transcription factors to 20 binding sites in a 2.3 kb *cis*-proximal regulatory region (Kirchhamer and Davidson, 1996; Kirchhamer *et al.*, 1996). Deletion and mutagenesis studies of the *cis*-regulatory region have identified particular functions for binding sites. Binding sites control timing and amplitude of expression and involve positive and negative regulation.

As the 5′ regulatory regions of the *H. iris* actin genes were not determined, except in *H.irisA1*, differences in the regulatory regions of genes which may be responsible for differential regulation of expression can not be inferred. However, the inability of regulatory region primers of *H. rufescens* and

H. iris to amplify PCR products with *H.irisA2* and *H.irisA3* gene-specific primers (see Chapter 2) suggests the regulatory regions of *H. iris* actin genes are divergent, supporting the hypothesis of transcriptional regulation.

ACTIN FUNCTIONAL TYPE

Expression analysis of *H.irisA1*, *H.irisA2* and *H.irisA3* in adult tissues and larvae suggests that cytoplasmic and muscle type actins exist in *H. iris*.

H.irisA1 and *H.irisA2* may represent cytoplasmic-type actins. *H. irisA1* and *H.irisA2* were ubiquitously expressed during larval development. *H.irisA1* was expressed in all adult tissues examined. *H.irisA2* expression was not detected in hepatopancreas tissue by RT-PCR, but may be expressed at low levels which could not be visualised. Southern hybridisation of adult RNA RT-PCR samples would confirm the absence of *H.irisA2* expression in hepatopancreas. The widespread expression of *H.irisA1* and *H.irisA2* support a cytoplasmic function for this gene.

H.irisA3 may represent a muscle-type actin. *H.irisA3* was expressed in late developmental larval stages, but not in early developmental stages such as fertilised egg and blastula. In adult tissues *H.irisA3* expression occurs at high levels in the muscular foot, at low levels in the muscular mantle and is faintly detected in the gonad. Expression of the *Drosophila* muscle type actin, *act87E*, occurs predominantly in the head and abdomen, but its expression is faintly detected in the ovary (Fyrberg 1984). Muscle tissue may be present in the structural tissue of gonad. In a broadcast spawner, such as *H. iris*, muscle tissue may contract the gonad and assist in gamete release.

The presence of distinct cytoplasmic and muscle type actins in molluscs has been debated due to the lack of sequence differences that characterise cytoplasmic and muscle actins in insects and vertebrates (Vanderkerckhove and Weber, 1984; Mounier *et al.* 1992). Mollusc actins have often been designated as muscle types based on evidence other than expression patterns (Patwary *et al.*, 1996; Carlini *et al.*, 2002). However, in studies where actins have been

characterised by expression analysis, the presence of a muscle-specific function for actins has been confirmed (DesGroseillers *et al.*, 1990; DesGroseillers *et al.*, 1994).

The lack of 'muscle-characteristic amino acids' in mollusc actin genes specifically expressed in muscle tissue may suggest a lack of structural specialisation within mollusc muscle actins. This may be due to a lack of complexity in the actin-binding proteins which are present to interact with actin. A large number of actin-binding proteins have been characterised in insects and vertebrates; it is thought that cytoplasmic and muscle-characteristic amino acids may have evolved to facilitate specialist interactions to occur between actins and these diverse actin-binding proteins (Sheterline and Sparrow, 1994). The diversity of actin-binding proteins in molluscs may be smaller than that in insects and vertebrates, reducing the selective pressure for mollusc actins to change and evolve binding sites specific to particular actin-binding proteins. However, in *H. iris* the presence of a large muscular foot that expresses multiple actin genes predicts the requirement of multiple associated actin-binding proteins, such as myosin, for muscle formation and function. Investigation of mollusc myosins may elucidate whether lack of actin protein diversification is related to lack of diversification in associated actin-binding proteins.

Alternatively, muscle-characteristic amino acids may be present in mollusc muscle actins, but as yet unidentified. The lack of a large number of muscle actin sequences in other molluscs may also make the detection of mollusc muscle-characteristic amino acids more difficult. The number of muscle-characteristic amino acids in insect muscle actins is smaller than that in vertebrates, insects have 7-9 whereas vertebrates have 17-26 muscle-characteristic amino acids (Mounier *et al.*, 1992). It may be that the number of mollusc muscle-specific amino acids is even smaller and have been overlooked.

In *H.irisA3*, a putative *H. iris* muscle actin, there are two amino acid changes that are not found in the cytoplasmic *H. iris* genes, a leucine at aa 177 and a methionine at aa 190. Leucine 177 is also found in *Placopecten* muscle actin

(Patwary *et al.*, 1996) and the *Aplysia* muscle actin gene (DesGroseillers *et al.*, 1990), which may suggest it could be a muscle-specific amino acid. Leucine 177 is also found in the sequences of *Crassostrea gigas* 2 actin, *Mytilus galloprovincialis* actin and *Dreissena polymorpha* actin, for which expression type has not been established. Isolation of *Crassostrea* 2 actin from a mantle cDNA library suggests it could be a muscle-type actin. The GenBank entry for *Dreissena* actin lists it as a cytoplasmic actin. The basis for cytoplasmic classification was probably the isolation of the gene from testis, but the presence of muscle actins in *Drosophila* gonad, and possibly *H. iris* muscle actin in gonad, has been shown. Whether the function of this gene was checked by expression analysis is unknown.

The key force behind the evolution of muscle-specific amino acids in muscle actins is the interaction with myosin. Modelling the atomic structure of actin has allowed the prediction of myosin-binding amino acids (Rayment *et al.*, 1993). Rayment *et al.* (1993) identified several binding sites for myosin and highlighted the importance of the acidic N-terminus in the interaction with myosin. In vertebrates the N-terminus of actins is variable enough to allow the identification of actin based on the sequencing of the N-terminal tryptic peptides (Vandekerckhove and Weber, 1978). There are muscle-specific amino acids in the variable N-terminal region, which are presumably linked to muscle function and myosin binding. Unfortunately variation in the N-terminal sequences of *H. iris* actins could not be determined due to the methodology employed to isolate sequences; the N-terminal sequences were predetermined due to the isolation of actin sequences by PCR using primers that bound to the beginning of the genes' coding regions. Because of this the presence of any possible muscle-specific amino acids in the N-terminus of *H. iris* actins could not be determined.

Chapter 4:

HALIOTIS IRIS ACTIN GENE FAMILY EVOLUTION

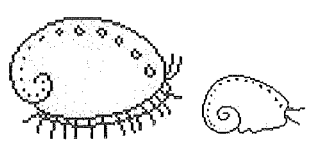


TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

GENE FAMILY EVOLUTION

THE ROLE OF PHYLOGENETICS

Gene families are defined as groups of genes which show at least 50% similarity at the protein level (Li, 1997). Gene families arise by ancestral gene duplication and divergence (Ohno, 1970; Efstratiadis *et al.*, 1980; Doolittle, 1989). The process of gene evolution is analogous to the process of speciation by lineage splitting and divergence, therefore phylogenetic approaches are applicable to the study of gene families. Phylogenetic techniques can be used to trace the evolutionary history of gene families within and between genomes.

Gene family members can be classified based on their evolutionary history, which can be inferred through phylogenetic studies (Abouheif *et al.*, 1997). Homologous genes are related by descent from a common ancestor. Homologous genes can be further divided into orthologous and paralogous genes. Orthologous genes are homologous genes in different species, which have arisen from a common ancestor due to speciation. Paralogous genes are homologous genes within a species, which have been created by gene duplication within the species (Figure 4.1).

Gene family phylogenies can be used to examine evolutionary processes leading to the functional diversity of gene families (Fryxell, 1996; Thornton and DeSalle, 2000). Phylogenetics has been used to trace the functional diversification of mammalian opsins responsible for colour vision (Yokoyama and Radlwimmer, 1999) and to illustrate the functional convergence of the genetically unrelated lysozymes (Kornegay *et al.*, 1994). Identification of sequence divergence rates within phylogenies has inferred selection via evolutionary rate disparities in sea urchin actins (Kissinger *et al.*, 1997). Similarities between phylogenies of functionally related gene families suggests related gene families expand and diversify by coevolution (Fryxell, 1996).

MECHANISMS OF GENE FAMILY EVOLUTION

There are three general mechanisms of gene family evolution: gene duplication, lateral gene transfer and concerted evolution (Figure 4.2). These mechanisms leave genomic evidence which can be detected by phylogenetic studies (Thornton and DeSalle, 2000).

Gene duplication can involve parts of genes, whole genes, chromosomes or genomes, and may occur by several mechanisms, such as tandem duplication, duplication of partial or whole genomes (polyploidy), DNA transposition or RNA transposition. Tandem duplication occurs by unequal recombination or replication slippage and duplicates a gene. Tandem duplication is evidenced by the creation of closely linked paralogues on a chromosome. Duplication of parts of a genome creates clusters of duplicated genes, whereas duplication of whole genomes results in an increased number of genes which are spread throughout the genome. DNA transposition can be detected by the presence of conserved sequences at the ends of the duplicated gene, the remnants of mobile genetic elements which facilitated transposition. RNA transposition can be detected by the absence of introns in the duplicate gene.

Lateral gene transfer can occur by horizontal transfer between paralogous genes or by domain shuffling. Horizontal transfer can occur between genes or gene groups and is inferred by incongruities in phylogenies. Domain shuffling, by transposition of gene fragments or nonhomologous recombination, involves the reorganisation of protein domains within genes to form a new gene. This can be analysed by studying the phylogenies of individual protein domains and looking for incongruities when different domains are used (Gerdes *et al.*, 2000).

Concerted evolution can occur by unequal recombination, gene conversion and replication slippage and results in the homogenisation of paralogues within a genome (Elder and Turner, 1995).

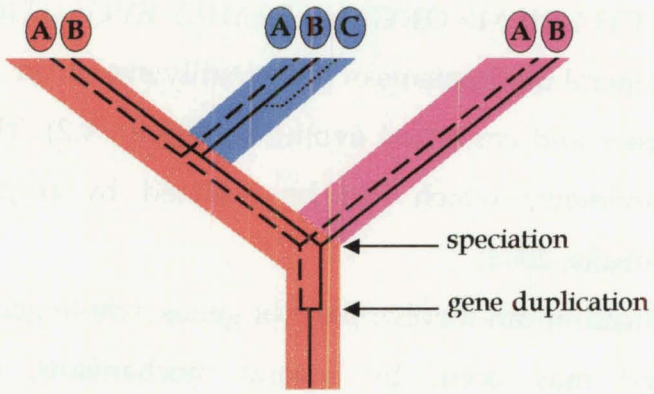


Figure 4.1. Homology, orthology and paralogy. Each colour represents a species. Genes A, B and C are homologous as they are all related by descent from a common ancestor. Genes A and B are orthologous as they existed together in a common ancestor and arose by speciation. Genes B and C of the blue species are paralogous as they arose by duplication within one species. (Based on Page and Holmes (1998)).

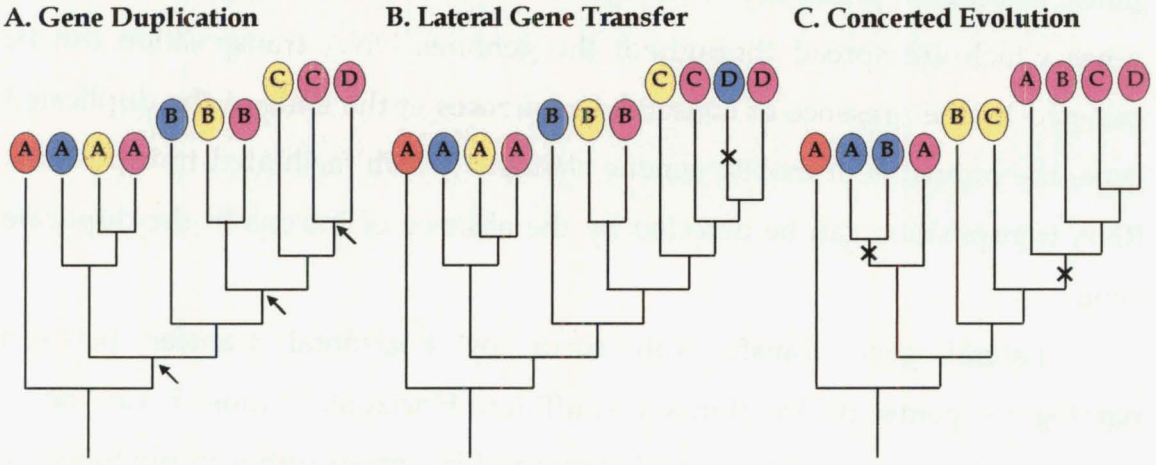


Figure 4.2. Mechanisms of gene family evolution. For each phylogeny, colours represent species and letters represent genes. Incongruities in phylogenies, indicating evolution by mechanisms other than gene duplication, are marked with crosses (×). **A. Gene duplication.** Orthologous genes group together. Origin of genes by duplication are indicated by arrows **B. Lateral gene transfer.** Genes from unrelated species group together. **C. Concerted evolution.** Paralogous genes group together. (Based on Thornton and DeSalle (2000)).

ACTIN GENE FAMILY EVOLUTION

ORTHOLOGY AND PARALOGY

Investigation of the orthologous relationships of insect actin genes has been facilitated by characterisation of gene families in *Drosophila melanogaster* and *Bombyx mori* and the isolation of gene family members in *Bactrocera dorsailis* and *Ceratitis capitata.*

The evolution of insect actins has been analysed by pairwise sequence comparisons between species and by phylogenetic techniques (Mounier *et al.*, 1992). Pairwise comparison of *Bombyx* and *Drosophila* actin proteins identified two clades based on sequence similarity greater than 98%, one group contained the cytoplasmic genes, the other contained the muscle genes. Phylogenetic reconstruction of metazoan taxa, including arthropods (*Drosophia, Bombyx, Artemia*), deuterostomes (*Strongylocentrotus purpuratus, Pisaster ochraceus)* and a mollusc (*Aplysia californica*), was performed using pairwise difference of predicted amino acid and parsimony. Phylogenetic trees showed that arthropod and deuterostome muscle actins formed two unrelated monophyletic groups, suggesting two muscle actin origins, once in the protostome lineage and once in the deuterostome lineage. The arthropod muscle actin clade was rooted within the arthropod cytoplasmic actins. Mounier *et al.* (1992) speculate that phylogenetic reconstruction and the absence of vertebrate or insect muscle-specific amino acids in the actins of the molluscs *Aplysia, Patella vulgata* and *Pecten* sp. is evidence that muscle actins do not exist in molluscs.

Within the Insecta, the phylogenetic relationships of dipteran actins has been studied. Sequence similarity, shared nonsynonymous changes, common intron positions and similar expression patterns divided dipteran sequences into three orthologous groups. These orthologous groupings were supported by phylogenetic analysis of non-synonymous substitution data and protein sequences by neighbour-joining (He and Haymer, 1995).

The results of Mounier *et al.* (1992) and He and Haymer (1995) suggest the following orthologous insect actin groups: 1. Cytoplasmic actins:

Drosophila act42A and *act5C*, *Bombyx A3*. (*Bombyx A4* was not included in either of these studies, but is expected to group here as it is a cytoplasmic actin). 2. Larval-adult muscle actins: *Drosophila act57A* and *act87E*, *Bactrocera A2* and *A3* and *B. mori A1*. 3. Two adult muscle actins: *Drosophila act88F*, *Bactrocera A1* and *Bombyx A2*; and *Drosophila act79B* and *Bactrocera A1*.

Orthologous relationships have been detected in mollusc actins (Carlini *et al.*, 2000; Adema, 2002). In a cephalopod study, maximum parsimony (MP) was used to analyse 82 actin sequences isolated from 44 taxa. Cephalopod actins divided into three clades of orthologous genes, arbitrarily named *Actin I*, *Actin II* and *Actin III*. The monophyly of the clades was well supported for *Actin I* and *Actin III* and moderately supported for *Actin II* (99, 96 and 70% bootstrap support, respectively) (Carlini *et al.*, 2000).

MP was used to analyse the predicted amino acid sequences of *Actin I*, Actin *II* and *Actin III* from seven cephalopod and the actin sequences from a range of Metazoans. The Metazoan actins included six non-cephalopod mollusc sequences, the cytoplasmic and muscle actins of *Aplysia*, the muscle actin of *Placopecten magellanicus*, the cytoplasmic actin of *Biomphalaria glabrata*, *GIA* from *Crassostrea gigas* and an actin from *C. virginica*. The results showed that mollusc muscle actins formed a group which included the cephalopod *Actin I* sequences, suggesting that these genes encode muscle actins and that orthologous actin groups exist in molluscs (Carlini *et al.*, 2000).

In a gastropod (*Biomphalaria* and *Helisoma* species) study, MP, minimum evolution and maximum likelihood (ML) were used to analyse the relationship between five *Biomphalaria* species, a *Helisoma* species and other molluscan actins. Phylogenetic analysis showed that cytoplasmic and muscle actins of molluscs formed distinct groups, supporting orthology of functionally specialised actins (Adema, 2002).

Paralogous genes have been identified in the sea urchin *S. franciscanus* (Foran *et al.*, 1985). Characterisation of *SfA 15A* and *SfA 15B* suggests they have arisen by a recent duplication event: the genes are only 5.7 kb apart, they are in

the same transcriptional orientation, they show 98.3% identity at the nucleotide level and contain introns at identical positions.

NUCLEOTIDE AND CODON BIAS

Base frequency and codon biases have been shown to obscure phylogenetic relationships of dipterans (He and Haymer, 1995). Phylogenetic reconstruction using non-synonymous substitutions or protein sequences grouped sequences into orthologous groups, whereas analysis with synonymous substitutions grouped sequences into species. Synonymous substitutions failed to reveal the evolutionary history and led to a homogenisation effect. The possible causes of the homogenisation effect, gene conversion, variation in the rate of substitution and codon biases, were investigated. Homogenisation was caused by codon biases, which were correlated with G+C content in *Drosophila*. Codon bias in *Drosophila* has been detected previously (Shields *et al.*, 1988). Comparison of substitution rates at synonymous and nonsynonymous sites within and between species showed that the substitution rate was constant at nonsynonymous sites but varied at synonymous sites, presumably due to the G+C preference. No evidence of gene conversion was detected.

GENE CONVERSION

Investigation of muscle actin evolution led researchers to suggest a gene conversion event had occurred in *Pisaster* muscle actins, which created conflicting results during phylogenetic analyses (White and Crother, 2000). When different echinoderm species were used in phylogenetic reconstruction, sea urchin (*Heliocidaris tuberculata*) actins placed echinoderms as a sister group to the chordate muscle actins, but when sea star (*Pisaster*) actins were used echinoderms were placed basally on the tree and showed little relationship to chordate muscle actins.

Previous characterisation of *Pisaster* genes highlighted features that serve as evidence to support the hypothesis of a gene conversion event: the N-terminal domains of the cytoplasmic and muscle gene were apparently

conserved (7.3% difference), there were less four-fold degenerate silent substitutions in the N-terminal domain and muscle-specific amino acids were restricted to the C-terminal domain (Kowbel and Smith, 1989).

White and Crother (2000) analysed the *Pisaster* cytoplasmic and muscle actins for gene conversion. They found that the region from codons 41-121, between two introns of the genes, showed less nucleotide change than other regions of the gene (4.12% compared to 12.29%). When this region was excluded from the sequence used in phylogenetic analyses the *Pisaster* muscle-actin was placed as a chordate muscle-actin sister group. Statistical analyses suggested that *Pisaster* muscle actin is significantly more related to chordate muscle actins along much of its length. These results pointed to a gene conversion in the N-terminal part of the gene, causing part of the *Pisaster* muscle actin to resemble the cytoplasmic actin, which obscured the true phylogenetic relationship of deuterostome muscle actins.

Changes in the sequence similarity along genes has also be used as evidence for gene conversion in the muscle actin of the sea urchin *S. purpuratus* (Crain *et al.*, 1987). Crain *et al.* (1987) compared the *S. purpuratus* muscle actin, *M*, to two *S. purpuratus* cytoplasmic actins, *CyI* and *CyIIa*. They found a region of high sequence similarity between codons 61-120 of the muscle and cytoplasmic genes, that showed only 7.9% synonymous substitutions compared to the 43.3% found in the rest of the gene.

Further evidence for the gene conversion events in *Pisaster* and *S. purpuratus* was provided by Drouin (2002). Drouin employed the GENECONV program (Sawyer, 1989) to test whether the sequence similarity used to infer gene conversion was significant. The results supported relatively old gene conversion events in *Pisaster* and *S. purpuratus*.

POSITIVE SELECTION

Another evolutionary process which can be tested for is positive selection. Positive selection has not been detected in actin gene families, but investigation for its presence is important as it may point to functional specialisation of actins and must be considered when performing evolutionary analyses as it may influence the phylogenies produced.

In abalone, positive selection has been detected in the sperm lysin, the sperm protein which facilitates species-specific fertilisation (Lee and Vacquier, 1992). Lee and Vacquier (1992) compared the lysin amino acid sequences of seven conspecific Californian abalone. They found that in 20 of the 21 pairwise comparisons possible for seven species, the number of nonsynonymous substitutions exceeds the number of synonymous substitutions, indicating the presence of positive selection. Further studies showed that positive selection was a general phenomenon in abalone lysin, nonsynonymous substitutions exceeded synonymous substitutions in an analysis of 20 lysin sequences isolated from abalone from Japan, Taiwan, Australia, New Zealand, South Africa and Europe (Lee *et al.*, 1995).

Positive selection of lysin is hypothesised to be driven by the evolution of the vitelline envelope receptor for lysin (VERL). VERL contains repetitive protein motifs which are homogenised by concerted evolution, but the VERL protein as a whole evolves by neutral drift. Neutral drift in VERL could drive adaptive sweeps in lysin, to maintain fertilisation potential (Swanson and Vacquier, 1998; Swanson *et al.*, 2001).

AIMS

The evolution of actin genes in insects and echinoderms have been studied. The evolutionary relationships between mollusc actins have been partially characterised, but are hindered by the small number of actin gene sequences available for mollusc species. The aim of this study is to investigate the evolution of the actin gene family of *H. iris*.

The specific aims of this chapter are to:

1. Identify orthologous actins between haliotids and molluscs
2. Examine nucleotide and codon bias in the haliotids, and the potential effect of bias on phylogenetic analysis
3. To investigate evolutionary process in the *H. iris* gene family

METHODS

PHYLOGENETIC DATA

A data set of 11 *Haliotis* actin sequences were used for phylogenetic analysis. The haliotid data set consisted of six *H. iris* sequences (three full and three partial), three *H. virginea* sequences (partial), a *H. rufescens* sequence (partial) and a *H. discus hannai* sequence (full). The *H. virginea* sequences were determined as part of this research (sequence data Appendix F).

For gastropod analyses a data set of 22 sequences was used. This contained the 11 *Haliotis* sequences above, as well as three *Patella vulgata* actin sequences, two *Aplysia californica* sequences, five *Biomphalaria* spp. sequences and a *Helisoma* sequence. The details of these sequences are listed in Table 4.1.

Analyses were performed with a set of cephalopod actin sequences. This set contained 21 sequences isolated from seven species. This set was composed of *Actin I, Actin II* and *Actin III* sequences from *Spirula spirula, Sepia opipara, Sepia officinalis, Idiosepius pygmaeus, Chtenopteryx sicula, Histioteuthis* and *Vampyroteuthis*. The details of these sequences are listed in Table 4.1.

Analyses performed with mollusc actins contained the 22 gastropod sequences and five bivalve actin sequences.

The species used as outgroups for phylogenetic analyses are listed in Table 4.2. For analysis of the haliotid data set, various mollusc species were used as outgroups. For gastropod and cephalopod studies two actin sequences from *Tetrahymena* were used.

Table 4.1. Actin sequences used for phylogenetic analyses. For each species the gene name, gene size and references are given.

Species	Taxonomic Group	Abbreviation	Type	Nt	Accession Number	Reference
H. iris	Haliotid	*H.irisA1*		1128	-	This thesis
		H.irisA1a		583	-	This thesis
		H.irisA1b		583	-	This thesis
		H.irisA1c		583	-	This thesis
		H.irisA2		1128	-	This thesis
		H.irisA3		1128	-	This thesis
H. virginea		*H.virgA1*		583	-	This thesis
		H.virgA2		583	-	This thesis
		H.virgA3		583	-	This thesis
H. rufescens		*H. rufes*		497	AF032125	Gomez-Chiarri et al., 1999
H. discus hannai		*H. disc*		1131	AY380809	Ma, Mai, Liufu, Chen, Xu, Zhang and Tan (unpublished)
Patella vulgata	Gastropod	*Patella 1*		1128	-	Van Loon *et al.*, 1993
		Patella 2		1128	-	Van Loon *et al.*, 1993
		Patella 3		1128	-	Van Loon *et al.*, 1993
Aplysia californica		*Aplysia* C	Cy	1131	U01352	DesGroseillers *et al.*, 1994
		Aplysia M	M	1131	X52868	DesGroseillers *et al.*, 1990
Biomphalaria glabrata M		*B. glabrata*		1131	AF329436	Adema, 2002
Biomphalaria alexandrina		*B. alex*		1131	AF329437	Adema, 2002
Biomphalaria tenagophila		*B. tena*		1131	AF329440	Adema, 2002
Biomphalaria obstructa		*B. obstructa*		1131	AF329439	Adema, 2002
Biomphalaria pfeifferi		*B. pfeifferi*		1131	AF329438	Adema, 2002
Helisoma trivolvis		*Helisoma*		1131	AF329441	Adema, 2002
Crassostrea gigas	Bivalve	*Crass GIA*		1131	AF026063	Cadoret *et al.*, 1999
		Crass 2		1131	AB071191	Miyamoto *et al.*, 2002
Mytilus galloprovincialis		*Mytilus*		1131	AF157491	Mitta and Cadoret (unpublished)
Dreissena polymorpha		*Dreissena*	Cy?	1131	AF082863	Lamers, Heiney and Ram (unpublished)
Placopecten magellanicus		*Placopecten*	M	1131	U55046	Patwary *et al.*, 1996

Sepia officinalis	Cephalopod	*Se. officinalis Act I*	M?	784	AF234905	Carlini *et al.*, 2000
		Se. officinalis Act II	Cy?	784	AF234906	Carlini *et al.*, 2000
		Se. officinalis Act III	ψ?	784	AF234907	Carlini *et al.*, 2000
Sepia opipara		*Se. opipara Act I*	M?	784	AF234902	Carlini *et al.*, 2000
		Se. opipara Act II	Cy?	784	AF23903	Carlini *et al.*, 2000
		Se. opipara Act III	ψ?	784	AF234904	Carlini *et al.*, 2000
Spirula spirula		*Spirula Act I*	M?	784	AF234913	Carlini *et al.*, 2000
		Spirula Act II	Cy?	784	AF234914	Carlini *et al.*, 2000
		Spirula Act III	ψ?	784	AF234915	Carlini *et al.*, 2000
Idiosepius pygmaeus		*Idiosepius Act I*	M?	784	AF234918	Carlini *et al.*, 2000
		Idiosepius Act II	Cy?	784	AF234919	Carlini *et al.*, 2000
		Idiosepius Act III	ψ?	784	AF234920	Carlini *et al.*, 2000
Chtenopteryx sicula		*Chtenopteryx Act I*	M?	784	AF234931	Carlini *et al.*, 2000
		Chtenopteryx Act II	Cy?	784	AF234932	Carlini *et al.*, 2000
		Chtenopteryx Act III	ψ?	784	AF234933	Carlini *et al.*, 2000
Histioteuthis hoylei		*Histioteuthis Act I*	M?	784	AF234948	Carlini *et al.*, 2000
		Histioteuthis Act II	Cy?	784	AF234949	Carlini *et al.*, 2000
		Histioteuthis Act III	ψ?	784	AF234950	Carlini *et al.*, 2000
Vampyroteuthis infernalis		*Vampyroteuthis Act I*	M?	784	AF234980	Carlini *et al.*, 2000
		Vampyroteuthis Act II	Cy?	784	AF234981	Carlini *et al.*, 2000
		Vampyroteuthis Act III	ψ?	784	AF234982	Carlini *et al.*, 2000
Tetrahymena thermophila	Ciliate	*T. thermophila*		1131	M13939	Cupples and Pearlman, 1986
Tetrahymena pyriformes		*T. pyriformes*		1131	X05195	Hirono *et al.*, 1987

EVOLUTIONARY TESTS

Models for phylogenetic inference are based on several assumptions which may not hold for biological data. These assumptions are nucleotide site independence, substitution rate equality between lineages and substitution rate equality within a sequence. When these assumptions are violated they may lead to inconsistencies in the reconstructed trees, therefore these assumptions must be tested to evaluate the reliability of phylogenetic inference.

Nucleotide sites are assumed to be independent, that is substitutions at one site do not influence the probability of substitutions at another site. This phenomenon is particularly important in genes encoding for products with secondary structures, such as the rRNA genes. When secondary structures are present, a change at one site may be matched by a compensatory change at another site. This is known as the covarion hypothesis. Independence was not considered to be a complicating factor for actin genes.

Violation of the assumption of equality of substitution rates between lineages is called non-stationarity. Non-stationarity can be caused by nucleotide frequency and codon biases within lineages. When non-stationarity is not accounted for sequences may group together based on shared biases which may not reflect a shared evolutionary origin, for example, sequences from unrelated species which tend towards a G+C bias may group together based on G+C content rather than phylogenetic signal (Stiller and Hall, 1999). When using distance methods, non-stationarity can be accounted for by using LogDet or paralinear distances (Lake, 1994; Lockhart *et al.*, 1994; Gu and Li, 1998). When using ML methods, non-stationarity can be accounted for by removing invariant sites from the data or by incorporating the proportion of invariant sites into the maximum likelihood algorithm (Steel *et al.*, 2000).

Violation of the assumption of equality of substitution rates within a sequence is known as amongst site rate variation or positional rate heterogeneity. This may be caused by functional constraints within a sequence leading to sites which do not change. Ignoring amongst site variation leads to

underestimation of evolutionary distance between sequences, underestimation of branch length and inconsistent tree reconstruction (Yang, 1996; Yang and Kumar, 1996). Amongst site variation is accounted for by using the gamma distribution to describe substitution rates within a sequence (Yang, 1996). The gamma distribution relies on a shape parameter, α, which can change the shape of the rate distribution. The α value allows a range of substitution rates to be accommodated; when α is less than one the rate distribution is L-shaped, when α is greater than one the rate distribution is bell shaped.

Nucleotide/codon bias, transition/transversion ratio and saturation, stationarity and amongst site rate variation were investigated to determine the suitability of the data for phylogenetic inference and the appropriate phylogenetic approach. The presence of gene conversion and adaptive evolution were also tested for, allowing greater confidence in phylogenetic analyses.

NUCLEOTIDE AND CODON BIAS

Base composition was estimated using MEGA 2.1 (Kumar *et al.*, 2001). The base frequencies for the first, second and third positions and the overall frequency for each codon were determined.

The degree of codon bias was determined by the relative synonymous codon usage statistic (RSCU) (Sharp and Li, 1988; Nei and Kumar, 2000). For each amino acid, the RSCU is the observed frequency of a codon divided by the expected frequency under the assumption of equal codon usage. RSCU equals one when there is no bias in codon usage. Methionine and tryptophan were not considered in codon bias analysis as they are only encoded by one codon. Stop codons were not analysed. In this study, RSCU values greater than 1.5 were considered biased.

TRANSITIONS/TRANSVERSIONS

The transition/transversion ratio (R) was estimated using MEGA 2.1 (Kumar *et al.*, 2001).

Substitution saturation was tested by plotting the number of transitions and transversion against the total evolutionary distance. The number of transitions and transversions were calculated using MEGA 2.1. The total evolutionary distance was estimated using the LogDet model available in Paup*4.0 (Swofford, 1999). The LogDet model was used due to sequence heterogeneity caused by nucleotide and codon bias.

STATIONARITY

Pattern homogeneity was tested by the disparity index test available in MEGA 2.1 (Kumar and Gadagkar, 2001). This test measures the observed difference in evolutionary patterns for pairs of sequences and compares them to the patterns expected under homogeneity. The difference between the observed and the expected patterns are tested for significance. The null hypothesis is that the substitution patterns are homogeneous. The null hypothesis is rejected when the probability is less than 0.05, i.e., the sequences have different substitution patterns. Modeltest also detects non-stationarity and estimates the proportion of invariant sites for incorporation into the maximum likelihood tree building algorithm.

AMONGST SITE RATE VARIATION

Amongst site rate variation was detected by Modeltest. This program analyses the sequence data and detects the presence of rate variation and estimates the gamma distribution shape value (α) to use in maximum likelihood inference.

GENE CONVERSION

The presence of gene conversion was assessed using GENECONV Version 1.81 (Sawyer, 1999). This program analyses aligned sequences and detects possible gene conversion events as regions with a high alignment score. As the number of aligned sequences increases, the probability of finding an erroneously significant result by pairwise comparison increases, therefore a global comparison is preferred as it is more conservative. The significance of the

possible gene conversion event is calculated as a permutation P-value or a Karlin-Altschul (KA) P-value. Permutation and KA P-values are corrected for the number of sequences compared and the length of the sequence, for KA values a Bonferroni correction is used (BC-KA).

The six actin sequences of *H. iris* were analysed using GENECONV. Two approaches were used, the default option, which compares polymorphic nucleotide sites, and the coding sequence silent option, which compares polymorphic codon sites. The silent option is more conservative for coding sequences, as the silent sites within a codon are expected to be correlated. This option ignores amino acid polymorphisms, which may increase significance values incorrectly. A G-scale of 0 was used, which sets the program to ignore internal mismatches, as there were no mismatches in the sequences. *H. iris* sequences were edited to be of the same length.

POSITIVE SELECTION

The number of synonymous substitutions per synonymous site (d_S), the number on nonsynonymous substitutions per nonsynonymous site (d_N) and the d_N/d_S ratio (ω) were calculated using the approximate method of Yang and Nielsen (2000), implemented in the yn00 program available in PAML Version 3.12 (Yang, 1997). This method is an improvement on other approximate methods as it incorporates nucleotide/codon bias and transition/transversion bias, which may bias the dN and dS values. Yang and Nielsens's method is more appropriate than ML methods when estimating d_S and d_N values for small sequences, less than 300 codons, due to a positive bias (Yang and Nielsen, 2000). The value of ω indicates what type of evolution is acting on the genes of interest, $\omega>1$ indicates positive selection, $\omega=1$ indicates neutral evolution, if $\omega<1$ indicates purifying selection.

PHYLOGENETIC TREE BUILDING

Three phylogenetic methods were considered, maximum parsimony, distance and maximum likelihood. Method selection was made on the basis of advantages and disadvantages of each approach, which are well reviewed in Nei (1996), Page and Holmes (1998), Nei and Kumar (2000) and Holder and Lewis (2003). A summary of the three methods and evaluation of their appropriateness in this study are given below.

Maximum parsimony (MP) assumes that the sequences of interest have evolved by duplication and divergence from common ancestors and is therefore suitable for the evolutionary analysis of gene families (Thornton and DeSalle, 2000). MP selects the most parsimonious tree (i.e., the tree which requires the least number of substitutions) which explains the distribution of extant sequences as a result of common ancestry. However, as MP assumes unidirectional changes it will often give erroneous tree topologies due to the presence of backward and parallel substitutions and the inequality of nucleotide substitution rates amongst lineages leading to long branch attraction.

MP was chosen for phylogenetic inference. Results should be considered in conjunction with the disparity index due to the presence of nonstationarity in some lineages, which may create erroneous trees.

Distance methods assume that sequences which have diverged more recently than other sequences will show a higher level of similarity, due to inheritance from their common ancestor. Therefore, similarities can be used to measure relatedness. Distance methods convert sequences to evolutionary distance measures which are compared and sorted on the basis of level of similarity. Distance methods are simple and fast and have been used extensively in the past. However, distance methods are inappropriate for the analysis of gene families where duplicated members may undergo rapid diversification and loss of sequence similarity, despite relatedness by descent

(Thornton and DeSalle, 2000). Distance methods were not used during this study.

Maximum likelihood (ML) was chosen for phylogenetic inference as it analyses sequences given an evolutionary model and is not subject to long branch attraction, making it more suitable for analysis of species with long divergence times (Nei and Kumar, 2000; Holder and Lewis, 2003). The selection of an appropriate evolutionary model is critical to the success of ML and has been simplified by the creation of computer software, such as Modeltest, to analyse data sets and suggest the most appropriate substitution model (Rzhetsky and Nei, 1995; Posada and Crandall, 1998). The main disadvantage of ML is that it is computationally expensive and may take a long time depending on the computer used for analyses.

ML was used to analyse relationships within the haliotid actins (allowing comparison to MP trees) and the relationships of the haliotid actins to the mollusc actins. ML was used to analyse haliotids and a subset of cephalopod actin sequences. Actins from only three cephalopods species were used as analysis of the full cephalopod data set would have been to computationally intensive.

MAXIMUM PARSIMONY

MP trees were constructed using MEGA 2.1 (Kumar *et al.*, 2001). Gaps were handled by pairwise deletion. First, second and third codon position data was used. CNI searching was used. Statistical support for branching was estimated by bootstrapping 500 times. The 11 haliotid sequences were analysed with the actins from different mollusc species as outgroups (Table 4.2).

The haliotid actin sequences were analysed with a cephalopod actin data set. Twenty one cephalopod actin sequences were used, corresponding to seven species for which *Actin I*, *Actin II* and *Actin III* sequences were available (Carlini *et al.*, 2000).

MAXIMUM LIKELIHOOD

Substitution Model

The appropriate substitution model was determined using ModelTest 3.06 PPC (Posada and Crandall, 1998). ModelTest uses the Akaike information criteria to compare 56 evolutionary models (this includes gamma and invariance versions of each model) and find the model which best fits the data. ModelTest estimates the substitution rate matrix, gamma distribution α shape value and proportion of invariance for inclusion into the ML algorithm used to analyse the data set.

Tree Construction

Trees were constructed using Paup*4.0Beta10 (Swofford, 1999). Heuristic searching with general search options was used. The statistical significance of trees were examined by bootstrapping 100 times. Haliotid trees were constructed using different mollusc outgroups. A gastropod tree was constructed using actins from *Tetrahymena thermophila* (accession number M13939.1 (Cupples and Pearlman, 1986)) and *T. pyriformes* (accession number X05195.1 (Hirono *et al.*, 1987)) as outgroups. Species used as outgroups are listed in Table 4.2.

Table 4.2. Species used as outgroups in phylogenetic analyses.

Species	Taxonomic Group	Number of Genes Used	Outgroup For:
Patella vulgata	Gastropod	3	Haliotids
Aplysia californica	Gastropod	2	Haliotids
Biomphalaria glabrata	Gastropod	1	Haliotids
Placopecten magellanicus	Bivalve	1	Haliotids
Crassostrea gigas	Bivalve	2	Haliotids
Mytilus galloprovincialis	Bivalve	1	Haliotids
Dreissena polymorpha	Bivalve	1	Haliotids
Tetrahymena thermophila	Ciliate	1	Gastropods
Tetrahymena pyriformes	Ciliate	1	Gastropods

RESULTS

EVOLUTIONARY STATISTICS

NUCLEOTIDE AND CODON BIAS

Haliotid Nucleotide Bias

The nucleotide frequencies of *Haliotis* actins are given in Table 4.3. In haliotids, nucleotide frequency for the overall codon favoured G+C, but showed no strong bias. The G+C content ranged from 51-55%, the average was 54%.

At the first position G+C content ranged from 52-60%. The average G+C content was 56%. Within *H. iris* the G+C content varied from 52-57%. In *H. virginea* actins the G+C contents were 56-57%. *H. rufescens* showed 54% G+C content. *H. discus hannai* had the highest G+C preference at 60%. At the first position the G composition ranged between 34-39%.

At the second codon position the average haliotid G+C content was 40%. The G+C content of *H. iris* varied between 42-39%, the G+C content of *H. virginea* was 40% and the G+C contents of *H. rufescens* and *H. discus hannai* were both 38%.

The average haliotid G+C content at the third position was 66%. In *H. iris* the G+C content ranged from 60-68%. The G+C content of *H. virginea* ranged between 63-68%. The G+C content of *H. rufescens* and *H. discus hannai* was 68 and 54%, respectively. The C composition at the third position ranged from 37-47%.

Table 4.3. Nucleotide frequency in haliotid actins. For each species the nucleotide frequency for the overall codon and the nucleotide frequencies at each codon position are given. The nucleotide frequency (mean ± the standard error of the mean (SE)) for the *Haliotis* species are shown. (Random substitution predicts equal nucleotide frequencies i.e. 0.25)

Codon Position	Nucleotide	*H.irisA1*	*H.irisA1a*	*H.irisA1b*	*H.irisA1c*	*H.irisA2*	*H.irisA3*	*H.virgA1a*	*H.virgA1b*	*H.virgA1c*	*H. rufescens*	*H. discus*	*Haliotis*	A+T/G+C
Overall	A	0.23	0.23	0.23	0.23	0.24	0.23	0.23	0.23	0.24	0.23	0.21	0.23±0.01	0.46
	T	0.23	0.22	0.23	0.23	0.25	0.24	0.22	0.23	0.24	0.22	0.24	0.23±0.01	
	G	0.24	0.24	0.24	0.24	0.24	0.24	0.25	0.25	0.24	0.25	0.26	0.24±0.01	0.54
	C	0.30	0.31	0.30	0.30	0.27	0.29	0.30	0.30	0.29	0.30	0.29	0.30±0.01	
1st	A	0.29	0.30	0.29	0.29	0.29	0.29	0.30	0.30	0.30	0.29	0.27	0.29±0.01	0.44
	T	0.19	0.14	0.14	0.14	0.18	0.17	0.14	0.14	0.14	0.17	0.13	0.15±0.02	
	G	0.34	0.34	0.35	0.35	0.34	0.35	0.34	0.34	0.34	0.35	0.39	0.35±0.02	0.56
	C	0.18	0.23	0.22	0.22	0.19	0.20	0.23	0.22	0.22	0.19	0.21	0.21±0.02	
2nd	A	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.30	0.30	0.31	0.30	0.31±0.00	0.60
	T	0.28	0.29	0.29	0.30	0.29	0.29	0.29	0.30	0.30	0.28	0.28	0.29±0.01	
	G	0.16	0.16	0.16	0.16	0.15	0.15	0.16	0.16	0.16	0.16	0.17	0.16±0.01	0.40
	C	0.26	0.24	0.24	0.23	0.26	0.26	0.24	0.24	0.24	0.26	0.25	0.24±0.01	
3rd	A	0.11	0.09	0.09	0.09	0.13	0.10	0.09	0.10	0.10	0.11	0.07	0.10±0.02	0.35
	T	0.22	0.23	0.26	0.24	0.27	0.26	0.23	0.24	0.27	0.22	0.30	0.25±0.03	
	G	0.21	0.23	0.22	0.22	0.23	0.23	0.24	0.23	0.21	0.24	0.23	0.23±0.01	0.66
	C	0.47	0.45	0.44	0.45	0.37	0.40	0.44	0.43	0.42	0.44	0.41	0.43±0.03	

Mollusc Nucleotide Bias

The mean nucleotide composition of haliotids was compared to the nucleotide composition of molluscs (Table 4.4).

At the first codon position G+C content ranged from 50-56%. The highest G+C content was found in *Haliotis*. The average G+C content was 52%.

At the second codon position G+C content ranged from 40% in *Haliotis* actins to 58% in *Aplysia* muscle and *Helisoma* actins. The average G+C content was 41%.

At the third codon position the G+C content ranged from 52% in *Dreissena* to 80% in *Aplysia* muscle actin. The average G+C content was 57%. The G+C contents of *Aplysia Cy* and *M* and *Placopecten actin* were high, 72, 80 and 72%, respectively. The C composition in *Aplysia Cy* and *M* and *Placopecten actin* were 48, 53 and 50%, respectively.

H. iris Intron Nucleotide Bias

The nucleotide frequency in *H. iris* introns was calculated (Table 4.5). *H. iris* actin introns had very similar nucleotide frequencies. The average G+C content was 32%.

Table 4.4. Nucleotide frequency in mollusc actins. For each species the nucleotide frequency for the overall codon and the nucleotide frequencies at each codon position are given. The nucleotide frequencies of *Haliotis* are the mean values from Table 4.3, the nucleotide frequencies of *Biomphalaria* are mean values for *B. alexandrina, B. glabrata, B. obstructa, B. pfeifferi* and *B. tenagophila* (the SE across *Biomphalaria* species was <0.00 for all codon positions). The nucleotide frequency (mean ± the SE) for molluscs is shown. (Random substitution predicts equal nucleotide frequencies i.e. 0.25).

Codon Position	Nucleotide	*Haliotis*	*Patella 1*	*Patella 2*	*Patella 3*	*Aplysia Cy*	*Aplysia M*	*Crassostrea GIA*	*Crassostrea 2*	*Placopecten*	*Mytilus*	*Dreissena*	*Biomphalaria*	*Helisoma*	Mollusca	A+T/G+C
Overall	A	0.23	0.29	0.29	0.29	0.23	0.21	0.26	0.24	0.24	0.29	0.26	0.28	0.28	0.26±0.01	0.50
	T	0.23	0.26	0.28	0.28	0.22	0.20	0.22	0.22	0.21	0.23	0.25	0.26	0.26	0.24±0.01	
	G	0.24	0.20	0.21	0.21	0.25	0.26	0.23	0.24	0.24	0.23	0.24	0.23	0.23	0.23±0.00	0.50
	C	0.29	0.25	0.22	0.23	0.31	0.33	0.28	0.30	0.31	0.26	0.25	0.23	0.24	0.27±0.01	
1st	A	0.29	0.29	0.29	0.30	0.29	0.28	0.29	0.29	0.31	0.31	0.30	0.30	0.31	0.30±0.00	0.48
	T	0.15	0.20	0.21	0.21	0.17	0.17	0.17	0.17	0.18	0.19	0.16	0.18	0.18	0.18±0.01	
	G	0.35	0.35	0.35	0.34	0.35	0.34	0.35	0.35	0.35	0.34	0.34	0.34	0.34	0.34±0.00	0.52
	C	0.21	0.17	0.15	0.16	0.20	0.21	0.19	0.19	0.19	0.16	0.19	0.17	0.18	0.18±0.00	
2nd	A	0.31	0.31	0.31	0.30	0.31	0.30	0.30	0.30	0.30	0.30	0.31	0.31	0.31	0.31±0.00	0.59
	T	0.29	0.28	0.28	0.29	0.29	0.28	0.29	0.28	0.28	0.28	0.28	0.28	0.28	0.28±0.00	
	G	0.16	0.15	0.15	0.15	0.16	0.16	0.16	0.16	0.16	0.17	0.17	0.16	0.16	0.16±0.00	0.41
	C	0.24	0.26	0.26	0.26	0.25	0.26	0.25	0.25	0.25	0.25	0.24	0.25	0.26	0.25±0.00	
3rd	A	0.10	0.27	0.27	0.27	0.09	0.04	0.19	0.13	0.11	0.24	0.18	0.22	0.22	0.18±0.02	0.43
	T	0.25	0.29	0.34	0.34	0.19	0.16	0.21	0.21	0.18	0.23	0.29	0.32	0.31	0.25±0.02	
	G	0.23	0.11	0.13	0.13	0.24	0.27	0.19	0.21	0.22	0.17	0.21	0.19	0.18	0.19±0.01	0.57
	C	0.43	0.34	0.26	0.27	0.48	0.53	0.41	0.45	0.50	0.36	0.31	0.27	0.29	0.38±0.03	

Table 4.5. Nucleotide frequency in *H. iris* actin introns. The mean (±SE) nucleotide frequency for *H. iris* introns is shown.

Nt	*H.irisA1*	*H.irisA1a*	*H.irisA1b*	*H.irisA1c*	*H.irisA2*	*H.irisA3*	*H. iris*	A+T/G+C
A	0.30	0.31	0.34	0.35	0.34	0.32	0.33±0.01	0.68
T	0.35	0.35	0.38	0.36	0.32	0.37	0.35±0.01	
G	0.22	0.19	0.16	0.18	0.19	0.17	0.18±0.01	0.32
C	0.13	0.16	0.12	0.12	0.15	0.14	0.14±0.01	

Haliotid Codon Usage Bias

In *H. iris*, 14 of the 18 amino acids examined had RSCU values greater than 1.5. Three amino acids had RSCU values between 1.25-1.5. Only one amino acid, cysteine, showed no codon bias between the two alternative codons. The highest biases were found in isoleucine (AUC used 2.44 as often as expected), serine (UCC used 2.5 times as often as expected), arginine (CGU used 2.4 times as often as expected) and glycine (GGU used 2.86 times as often as expected). Three amino acids, two of which are encoded by six codons, had two codons with RSCU values greater than 1.5: serine codons UCU and UCC had RSCU values of 2.27 and 2.50, alanine codons GCU and GCC had RSCU values of 2.02 and 1.64 and arginine codons CGU and AGA had RSCU values of 2.40 and 1.93).

Haliotids showed the same codon usage preferences for phenylalanine, isoleucine, serine, proline, threonine, glutamine, lysine, arginine and glycine. The RSCU values for these amino acids varied between haliotid species, for example the RSCU value for the serine codon UCU was greater than 3.00 in *H. virginea*, *H. rufescens* and *H. discus hannai*, but only 2.27 in *H. iris*. Haliotids showed different codon usage preferences for leucine, valine, alanine, tyrosine, histidine, asparagine, aspartic acid, glutamic acid and cysteine.

The relationship between codon bias and the frequency of the amino acid in the protein in haliotids was investigated by regression analysis (Figure 4.3). There was a significant relationship between the degree of codon usage bias for an amino acid (measured as the highest RSCU value found between synonymous codons) and the frequency of that amino acid in the actin protein,

the p-values were 0.010, 0.032. 0.016 and 0.045 for *H. iris, H. virginea, H. discus hannai* and *H. rufescens,* respectively.

Mollusc Codon Bias

All molluscs favoured the UUC codon for phenylalanine. All molluscs lacked a preference for aspartic acid codons. The codon usage for the other amino acids varied between molluscs.

Table 4.6. Codon bias in haliotids and molluscs. For each species the bias given is the average for actin sequences available. RSCU values above 1.5 are shown in bold.

		Species											
Amino Acid	Codon	*H. iris*	*H. virginea*	*H. rufescens*	*H. discus*	*Patella*	*Aplysia*	*Biomphalaria*	*Helisoma*	*Placopecten*	*Crassostrea*	*Mytilus*	*Dreissena*
Phe	UUU	0.15	0.25	0.40	0.00	0.22	0.08	0.34	0.50	0.00	0.15	0.15	0.62
	UUC	**1.85**	**1.75**	**1.60**	**2.00**	**1.78**	**1.92**	**1.66**	**1.50**	**2.00**	**1.85**	**1.85**	**1.38**
Leu	UUA	0.14	0.00	0.00	0.00	1.06	0.00	0.42	0.48	0.21	0.22	0.22	0.00
	UUG	0.95	0.71	0.67	0.89	**2.28**	0.44	**1.67**	1.20	1.07	0.56	**2.00**	0.92
	CUU	1.41	1.18	0.00	0.89	0.99	1.31	**1.63**	1.44	0.43	0.78	1.11	**1.85**
	CUC	**2.14**	**2.71**	**2.67**	**2.67**	1.44	**1.75**	**1.53**	**1.92**	**3.21**	**2.44**	**1.56**	**1.62**
	CUA	0.00	0.12	0.00	0.00	0.08	0.00	0.65	0.24	0.00	0.11	0.44	0.23
	CUG	0.36	1.29	**2.67**	**1.56**	0.15	**2.51**	0.09	0.72	1.07	**1.89**	0.67	1.38
Ile	AUU	0.51	0.29	0.33	0.33	1.27	0.39	**1.51**	1.39	0.56	0.72	0.67	1.39
	AUC	**2.44**	**2.71**	**2.67**	**2.67**	**1.59**	**2.61**	1.47	**1.50**	**2.44**	**2.28**	**2.33**	**1.50**
	AUA	0.05	0.00	0.00	0.00	0.14	0.00	0.02	0.11	0.00	0.00	0.00	0.11
Val	GUU	1.02	0.88	**1.71**	**1.71**	**1.68**	0.68	1.09	**1.57**	0.52	0.51	1.04	1.09
	GUC	**2.36**	**1.95**	1.43	1.43	1.39	**2.30**	1.09	1.39	**2.61**	**2.55**	**1.57**	**1.82**
	GUA	0.04	0.10	0.00	0.00	0.93	0.17	1.05	0.87	0.70	0.43	1.22	0.36
	GUG	0.58	1.07	0.86	0.86	0.00	0.85	0.77	0.17	0.17	0.51	0.17	0.73
Ser	UCU	**2.27**	**3.13**	**3.00**	**3.00**	**2.25**	1.04	**3.00**	**2.88**	0.46	1.22	1.29	**2.08**
	UCC	**2.50**	**2.09**	**2.25**	**2.25**	**1.80**	**3.92**	1.45	**1.68**	**3.69**	**3.22**	**1.93**	1.15
	UCA	0.35	0.00	0.00	0.00	1.20	0.23	0.75	0.72	0.69	0.44	**1.50**	0.92
	UCG	0.12	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.23
	AGU	0.06	0.00	0.00	0.00	0.30	0.12	0.25	0.00	0.00	0.11	0.43	0.46
	AGC	0.70	0.78	0.75	0.75	0.15	0.69	0.55	0.72	1.15	1.00	0.86	1.15
Pro	CCU	0.41	0.13	0.44	0.44	0.76	0.40	1.12	1.40	0.21	0.41	0.00	0.42
	CCC	**2.27**	**3.23**	**2.67**	**2.67**	1.38	**2.80**	0.76	0.40	**2.53**	**2.05**	0.84	0.84
	CCA	1.27	0.65	0.89	0.89	**1.86**	0.80	**2.12**	**2.20**	1.26	**1.54**	**3.16**	**2.74**
	CCG	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Thr	ACU	0.90	0.80	0.80	0.80	0.82	0.41	0.70	0.46	0.16	0.50	0.33	1.00
	ACC	**2.13**	**2.31**	**2.40**	**2.40**	**2.25**	**3.18**	**1.59**	**1.69**	**3.20**	**2.67**	**2.67**	**2.17**
	ACA	0.67	0.71	0.40	0.40	0.77	0.41	**1.68**	**1.85**	0.64	0.83	1.00	0.67
	ACG	0.30	0.18	0.40	0.40	0.16	0.00	0.03	0.00	0.00	0.00	0.00	0.17

Ala	GCU	**2.02**	**2.15**	**2.40**	**2.40**	**1.91**	1.40	**2.28**	**1.93**	**1.73**	**1.79**	**1.66**	1.33
	GCC	**1.64**	**1.56**	1.33	1.33	**1.68**	**2.60**	0.93	1.19	**2.13**	**2.00**	**1.93**	**1.60**
	GCA	0.28	0.29	0.27	0.27	0.36	0.00	0.79	0.74	0.13	0.21	0.28	1.07
	GCG	0.06	0.00	0.00	0.00	0.05	0.00	0.00	0.15	0.00	0.00	0.14	0.00
Tyr	UAU	0.43	0.62	0.50	0.25	1.11	0.21	1.07	0.88	0.27	0.27	0.43	1.20
	UAC	**1.57**	1.38	**1.50**	**1.75**	0.89	**1.79**	0.93	1.13	**1.73**	**1.73**	**1.57**	0.80
His	CAU	0.58	0.76	0.67	0.50	1.04	0.33	1.35	1.33	0.22	0.67	0.67	0.67
	CAC	1.42	1.24	1.33	**1.50**	0.96	**1.67**	0.65	0.67	**1.78**	1.33	1.33	1.33
Gln	CAA	0.26	0.13	0.00	0.15	1.39	0.32	0.72	0.46	0.18	0.33	0.33	0.15
	CAG	**1.74**	**1.87**	**2.00**	**1.85**	0.61	**1.68**	1.28	**1.54**	**1.82**	**1.67**	**1.67**	**1.85**
Asn	AAU	0.00	0.00	0.00	0.00	1.00	0.44	0.67	0.44	0.20	0.56	0.73	0.44
	AAC	**2.00**	**2.00**	**2.00**	**2.00**	1.00	**1.56**	1.33	**1.56**	**1.80**	1.44	1.27	**1.56**
Lys	AAA	0.50	0.44	0.29	0.32	**1.51**	0.16	0.93	0.95	0.53	0.63	1.20	0.53
	AAG	**1.50**	**1.56**	**1.71**	**1.68**	0.49	**1.84**	1.07	1.05	1.47	1.37	0.80	1.47
Asp	GAU	0.70	0.76	0.67	0.61	1.48	0.57	0.82	0.82	0.82	0.88	0.95	0.91
	GAC	1.30	1.24	1.33	1.39	0.52	1.43	1.18	1.18	1.18	1.12	1.05	1.09
Glu	GAA	0.60	0.43	0.44	0.37	**1.75**	0.19	1.04	1.11	0.59	0.87	1.38	1.04
	GAG	1.40	**1.57**	**1.56**	**1.63**	0.22	**1.81**	0.96	0.89	1.41	1.13	0.62	0.96
Cys	UGU	0.89	0.00	1.33	1.20	1.07	0.33	1.40	1.33	0.67	0.20	1.14	0.67
	UGC	1.11	**2.00**	0.67	0.80	0.93	**1.67**	0.60	0.67	1.33	**1.80**	0.86	1.33
Arg	CGU	**2.40**	**2.67**	**2.25**	**3.33**	0.67	**2.33**	**1.67**	**1.67**	**3.00**	**2.33**	1.06	**2.33**
	CGC	0.40	0.17	0.00	0.00	0.00	1.00	0.00	0.00	0.33	0.17	0.00	0.00
	CGA	0.47	0.17	0.75	0.00	**1.89**	0.00	0.33	0.33	0.00	0.17	0.00	0.67
	CGG	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AGA	**1.93**	**2.50**	**2.25**	**1.67**	**3.11**	1.33	**2.53**	**2.67**	0.67	**2.50**	**3.53**	**2.00**
	AGG	0.80	1.00	0.00	1.00	0.33	1.33	1.47	1.33	**2.00**	0.83	1.41	1.00
Gly	GGU	**2.86**	**2.88**	**3.20**	**3.20**	**2.92**	**1.86**	**2.59**	**2.57**	**2.29**	**2.28**	**2.71**	**2.57**
	GGC	0.76	0.47	0.80	0.80	0.42	1.36	0.72	0.86	0.71	0.21	0.14	0.57
	GGA	0.25	0.65	0.00	0.00	0.61	0.71	0.69	0.57	0.86	**1.52**	1.14	0.86
	GGG	0.13	0.00	0.00	0.00	0.05	0.07	0.00	0.00	0.14	0.00	0.00	0.00

TRANSITIONS/TRANSVERSIONS

Saturation plots for haliotids showed that transitions and transversions were not saturated (Figure 4.4).

Table 4.7 showed that transition/transversion ratio (R) varied between pairwise sequence comparisons. The highest R value was 5.25 for the *H.irisA1c/H.rufescens* comparison. The lowest R value was 0.84 for the *P. magellanicus/A. californica* muscle actin comparison. Within haliotids R ranged from 0.92-5.25, within *H. iris* R ranged from 1.23-4.75. For molluscs, 92.9% of pairwise sequence comparisons showed R values greater than 1.

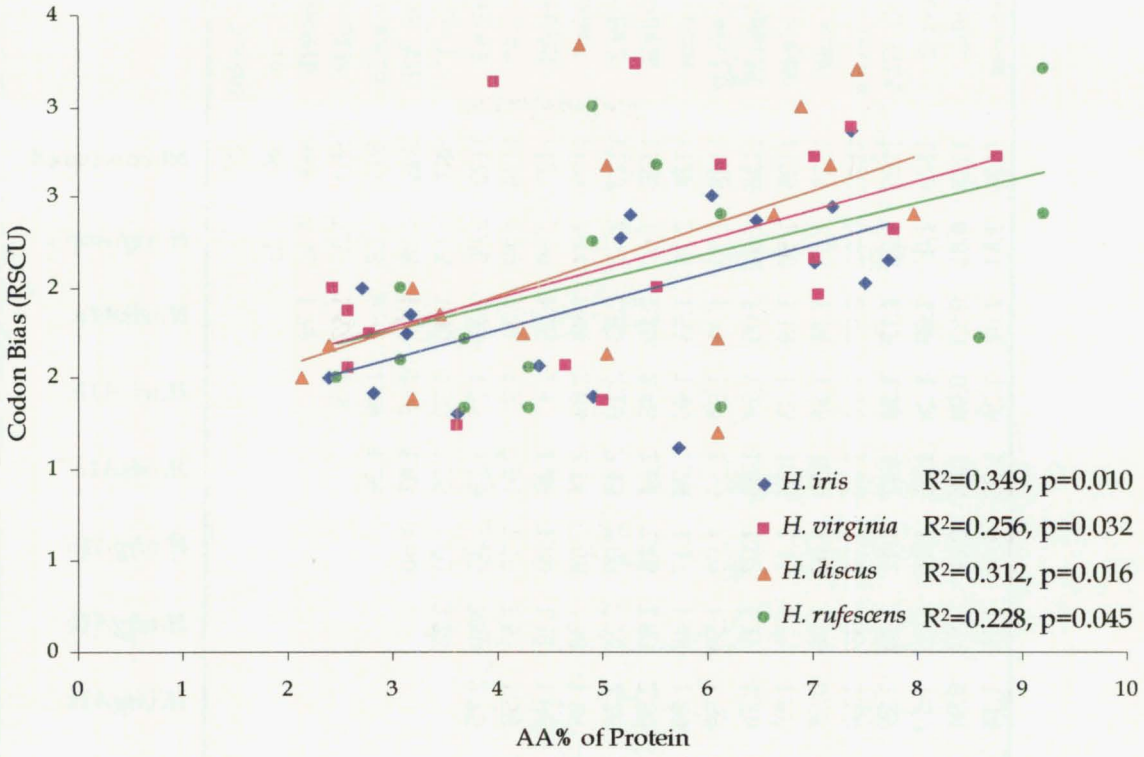


Figure 4.3. Relationship between amino acid frequency in the actin protein and codon bias for that amino acid. The occurrence of amino acid in the actin protein (AA% of Protein) is plotted against the codon bias for that amino acid. The largest RSCU value for each amino acid was used as a measure of codon bias.

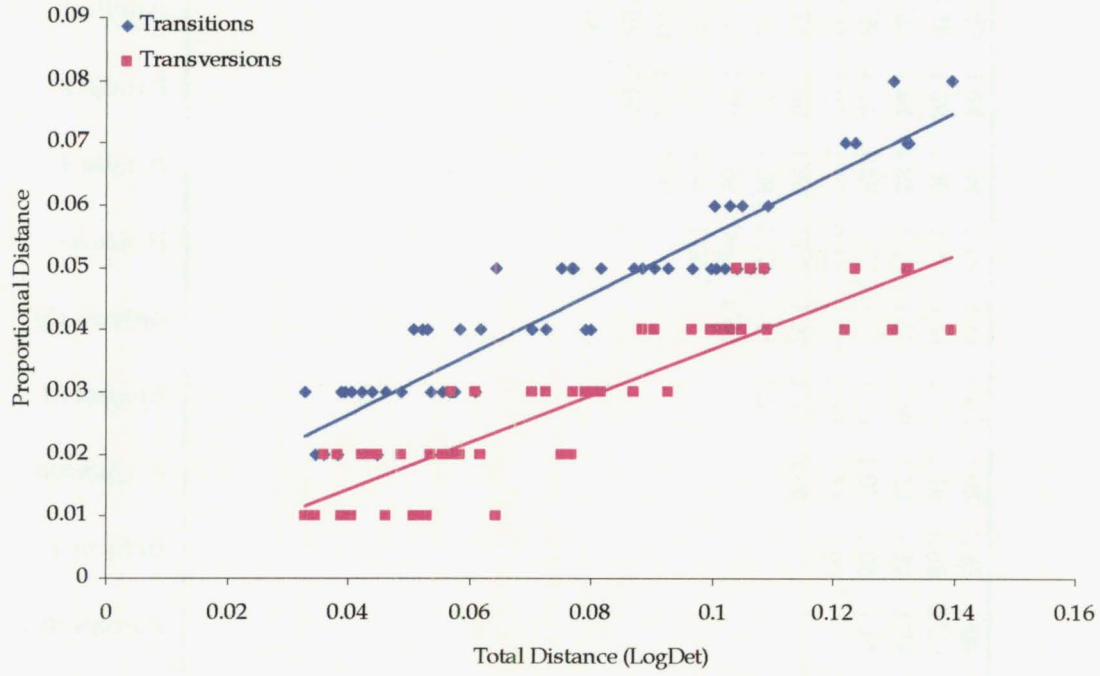


Figure 4.4. Substitution saturation in haliotids. The number of transitions and transversions are plotted against total evolutionary distance (estimated as LogDet distance) for *H. iris*, *H. virginea*, *H. discus hannai* and *H. rufescens*.

Table 4.7. Transition/transversion ratio (R) for molluscs. R= 1 shows transitions and transversions occur at equal rates, R>1 shows transitions are more frequent than transversions, R<1 shows transversions are more frequent than transitions.

	Nt compared	*H. rufescens*	*H.irisA1a*	*H.irisA1b*	*H.irisA1c*	*H.virgA1a*	*H.virgA1b*	*H.virgA1c*	*H.irisA1*	*H.irisA2*	*H.irisA3*	*Patella 1*	*Patella 2*	*Patella 3*	*H. discus*	*Aplysia Cy*	*Aplysia M*	*B. glabrata*	*Helisoma*	*Placopecten*	*Crass GIA*	*Crass 2*	*Mytilus*
H. rufescens	497																						
H.irisA1a	583	4.25																					
H.irisA1b	583	4.33	1.78																				
H.irisA1c	583	5.25	1.67	4.50																			
H.virgA1a	583	1.25	0.92	1.29	1.29																		
H.virgA1b	583	1.58	1.27	1.67	1.83	1.50																	
H.virgA1c	583	1.55	1.00	1.50	1.58	1.00	2.25																
H.irisA1	1125	5.00	1.33	4.75	3.75	1.33	2.30	2.75															
H.irisA2	1125	1.19	1.44	1.71	2.04	1.19	1.81	1.50	2.05														
H.irisA3	1125	2.50	1.23	1.17	1.43	1.00	1.11	1.03	1.65	1.68													
Patella 1	1125	2.14	1.94	2.03	2.14	2.03	2.38	1.85	1.55	1.69	1.43												
Patella 2	1125	2.35	2.03	2.19	2.45	2.00	2.26	2.26	1.69	1.63	1.67	1.78											
Patella 3	1125	2.07	2.15	2.43	2.64	2.18	2.53	2.50	1.69	1.58	1.64	1.87	2.06										
H. discus	1128	2.18	1.53	1.44	1.78	1.17	1.45	1.29	1.60	1.41	1.38	1.60	1.97	1.86									
Aplysia Cy	1128	1.75	1.31	1.34	1.59	1.52	1.59	1.59	1.35	1.22	1.14	1.75	2.21	2.10	1.00								
Aplysia M	1128	1.64	1.05	1.24	1.43	1.07	1.17	1.26	1.24	1.08	1.09	1.79	1.96	1.95	0.87	0.94							
B. glabrata	1128	1.50	1.19	1.33	1.33	1.15	1.21	1.10	1.22	1.18	1.11	1.37	1.35	1.28	1.12	1.41	1.21						
Helisoma	1128	1.33	1.14	1.16	1.20	1.02	1.06	0.98	1.14	1.20	1.07	1.51	1.38	1.35	1.04	1.54	1.22	2.94					
Placopecten	1128	0.95	0.97	1.14	1.36	1.05	1.19	1.20	1.20	1.05	1.08	1.41	1.61	1.52	0.97	1.04	0.84	1.11	1.09				
Crass GIA	1128	1.06	1.17	1.26	1.47	1.11	1.35	1.29	1.13	1.65	1.12	1.38	1.77	1.76	1.16	1.11	1.10	1.07	1.04	1.14			
Crass 2	1128	1.03	1.08	1.26	1.52	1.16	1.20	1.39	1.41	1.32	1.19	1.57	1.94	1.87	1.09	1.10	1.00	1.23	1.15	1.26	1.14		
Mytilus	1128	0.85	0.85	0.93	1.04	0.84	0.94	0.94	1.07	1.16	1.22	1.40	1.28	1.26	1.01	1.21	1.17	1.16	0.95	1.23	1.03	1.42	
Dreissena	1128	1.11	1.02	1.20	1.28	0.98	1.18	1.14	1.14	1.17	1.04	1.45	1.44	1.34	1.10	1.11	0.92	1.02	0.87	1.08	1.02	1.49	1.26

STATIONARITY

Table 4.8 shows the results of the disparity index test, asterisks indicate significant results, i.e., comparisons in which the null hypothesis of stationarity is rejected. Within haliotids non-stationarity exists in four of the 45 possible pairwise comparisons (8.9%): *H.irisA1/H.irisA2*, *H.irisA1/H.virgA1c*, *H.irisA2/H.irisA1a* and *H.virgA1a/H.virgA1c*. In molluscs non-stationarity exists between 251 of the possible 351 pairwise comparisons (71.5%).

POSITIVE SELECTION

The d_N, d_S and d_N/d_S ratios, ω, are listed in Table 4.9. Of the 27 possible pairwise comparisons for haliotid actins, none showed a d_N value greater than the d_S value. The d_N/d_S ratios, ω, were all below 0.08.

The relationship between *H. iris* d_S and d_N was investigated by regression analysis (Figure 4.5). The two substitution types are positively correlated, as d_N increases d_S also increases (P value=0.009).

Table 4.8. Disparity index test of stationarity. The null hypothesis is stationarity, significant results (*) reject the null hypothesis. Non-significant results (ns) are comparisons with P values>0.05.

	H. rufescens	*H.irisA1a*	*H.irisA1b*	*H.irisA1c*	*H.virgA1a*	*H.virgA1b*	*H.virgA1c*	*H.irisA1*	*H.irisA2*	*H.irisA3*	*Patella 1*	*Patella 2*	*Patella 3*	*H. discus*	*Aplysia Cy*	*Aplysia M*	*Crass GIA*	*Crass 2*	*Placopecten*	*Mytilus*	*Dreissena*	*B. tena*	*B. obstructa*	*B. pfeifferi*	*B. glabrata*	*B. alex*	*Helisoma*
H. rufescens																											
H.irisA1a	ns																										
H.irisA1b	ns	ns																									
H.irisA1c	ns	ns	ns																								
H.virgA1a	ns	ns	ns	ns																							
H.virgA1b	ns	ns	ns	ns	ns																						
H.virgA1c	ns	ns	ns	ns	*	ns																					
H.irisA1	ns	ns	ns	ns	ns	ns	*																				
H.irisA2	ns	*	ns	ns	ns	ns	ns	*																			
H.irisA3	ns	ns	ns	ns	ns	ns	ns	ns	ns																		
Patella 1	*	*	*	*	*	*	*	*	*	*																	
Patella 2	*	*	*	*	*	*	*	*	*	*	*																
Patella 3	*	*	*	*	*	*	*	*	*	*	*	ns															
H. discus	ns	ns	ns	ns	ns	ns	ns	ns	*	*	*	*	*														
Aplysia Cy	ns	ns	ns	ns	ns	ns	ns	ns	*	*	*	*	*	ns													
Aplysia M	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*												
Crassostrea GIA	*	*	*	*	ns	*	*	*	*	*	*	*	*	*	*	*											
Crassostrea 2	*	ns	ns	ns	ns	ns	*	ns	*	ns	*	*	*	ns	ns	*	*										
Placopecten	*	ns	*	ns	ns	*	*	ns	*	*	*	*	*	ns	ns	*	*	ns									
Mytilus	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*								
Dreissena	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*							
B. tenagophila	*	*	*	*	*	*	*	*	*	*	*	ns	*	*	*	*	*	*	*	*	ns						
B. obstructa	*	*	*	*	*	*	*	*	*	*	*	ns	*	*	*	*	*	*	*	*	*	ns					
B. pfeifferi	*	*	*	*	*	*	*	*	*	*	*	ns	ns	*	*	*	*	*	*	*	*	ns	ns				
B. glabrata	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	ns	ns	ns	ns			
B. alexandrina	*	*	*	*	*	*	*	*	*	*	*	ns	*	*	*	*	*	*	*	*	*	ns	ns	ns	ns		
Helisoma	*	*	*	*	*	*	*	*	*	*	ns	*	*	*	*	*	*	*	*	*	ns	ns	ns	ns	ns	ns	

Table 4.9. d_S and d_N statistics for haliotid actins.

Species Compared	d_N	d_S	$\omega=d_N/d_S$
H.irisA1 × *H.irisA1a*	0.000	0.202	0.000
× *H.irisA1b*	0.006	0.218	0.028
× *H.irisA1c*	0.006	0.218	0.028
× *H.irisA2*	0.009	0.613	0.015
× *H.irisA3*	0.006	0.325	0.019
× *H.virgA1a*	0.006	0.271	0.023
× *H.virgA1b*	0.000	0.409	0.000
× *H.virgA1c*	0.003	0.288	0.011
× *H. discus*	0.006	0.334	0.018
× *H. rufescens*	0.009	0.274	0.034
H.irisA2 × *H.irisA1a*	0.010	0.486	0.020
× *H.irisA1b*	0.016	0.460	0.035
× *H.irisA1c*	0.016	0.615	0.026
× *H.irisA3*	0.009	0.377	0.025
× *H.virgA1a*	0.016	0.581	0.027
× *H.virgA1b*	0.010	0.627	0.015
× *H.virgA1c*	0.013	0.604	0.021
× *H. discus*	0.013	0.452	0.028
× *H. rufescens*	0.016	0.420	0.038
H.irisA3 × *H.irisA1a*	0.006	0.265	0.023
× *H.irisA1b*	0.013	0.260	0.048
× *H.irisA1c*	0.013	0.381	0.033
× *H.virgA1a*	0.123	0.427	0.029
× *H.virgA1b*	0.006	0.438	0.014
× *H.virgA1c*	0.009	0.419	0.022
× *H. discus*	0.003	0.194	0.016
× *H. rufescens*	0.123	0.301	0.041
H.irisA1a × *H.irisA1b*	0.006	0.187	0.033
× *H.irisA1c*	0.006	0.237	0.026
× *H.virgA1a*	0.006	0.258	0.024
× *H.virgA1b*	0.000	0.380	0.000
× *H.virgA1c*	0.003	0.324	0.010
× *H. discus*	0.006	0.253	0.024
× *H. rufescens*	0.009	0.160	0.059
H.irisA1b × *H.irisA1c*	0.013	0.188	0.067
× *H.virgA1a*	0.013	0.285	0.044
× *H.virgA1b*	0.006	0.315	0.020
× *H.virgA1c*	0.009	0.230	0.041
× *H. discus*	0.012	0.265	0.047
× *H. rufescens*	0.016	0.205	0.077
H.irisA1c × *H.virgA1a*	0.013	0.324	0.039
× *H.virgA1b*	0.006	0.396	0.016
× *H.virgA1c*	0.009	0.283	0.033
× *H. discus*	0.012	0.390	0.032
× *H. rufescens*	0.016	0.275	0.057
H.virgA1a × *H.virgA1b*	0.006	0.157	0.040
× *H.virgA1c*	0.009	0.202	0.046
× *H. discus*	0.012	0.393	0.031
× *H. rufescens*	0.016	0.305	0.051
H.virgA1b × *H.virgA1c*	0.003	0.232	0.014
× *H. discus*	0.006	0.450	0.014
× *H. rufescens*	0.009	0.417	0.023
H.virgA1c × *H. discus*	0.009	0.428	0.022
× *H. rufescens*	0.009	0.359	0.026
H. discus × *H. rufescens*	0.016	0.271	0.057

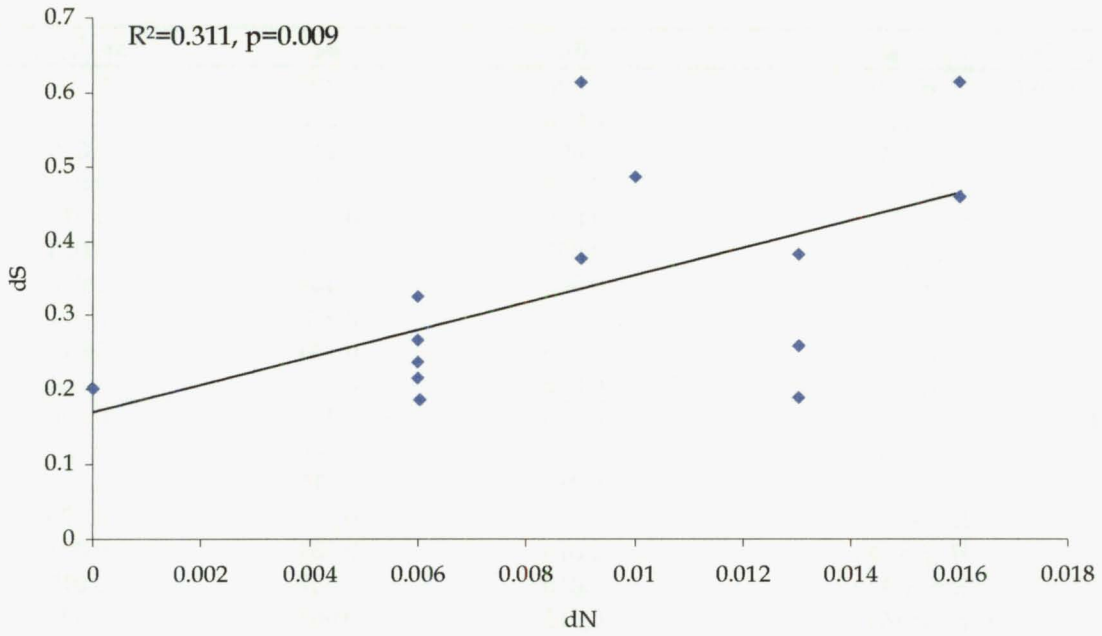


Figure 4.5. Relationship between d_N and d_S.

GENE CONVERSION

GENECONV detected one gene conversion event between *H. iris* actins. The conversion event was between *H.irisA2* and *H.irisA3* at nt 250-326, corresponding to aa 83-109. This conversion was supported by a significant permutation P-value (0.0326), but not by a significant BC-KA P-value (0.21813).

Gene conversion analysis was repeated using the conservative silent sites option for coding sequences. This analysis did not detect any gene conversion events.

PHYLOGENETIC TREE BUILDING

HALIOTIDS

Maximum Parsimony

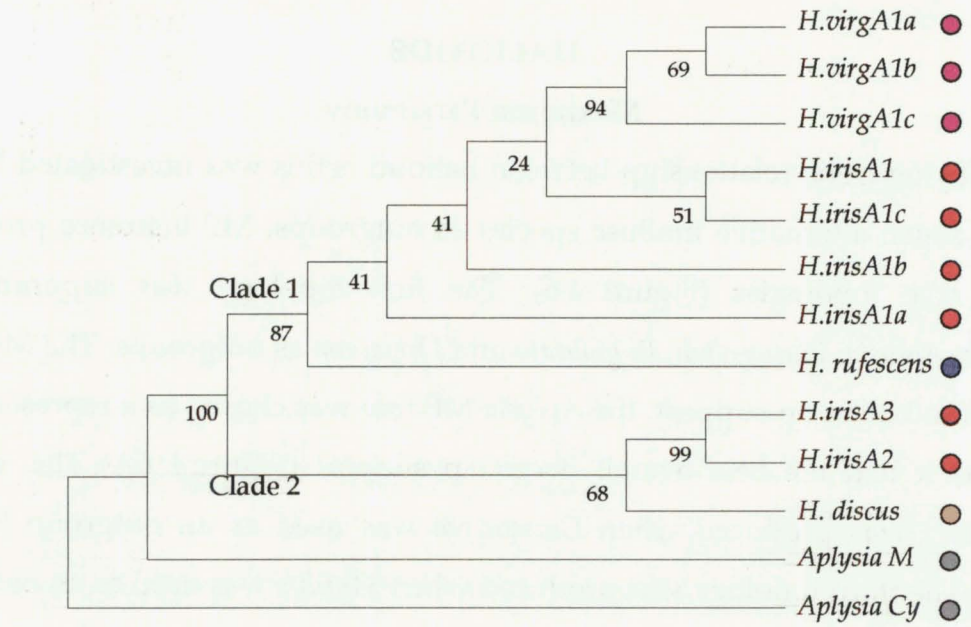
The phylogenetic relationship between haliotid actins was investigated by MP using seven alternative mollusc species as outgroups. MP inference produced three tree topologies (Figure 4.6). The first topology was supported by *Patella, Aplysia, Placopecten, B. glabrata* and *Dreissena* as outgroups. The MP trees varied in bootstrap support, the *Aplysia* MP tree was chosen as a representative tree as it had the best overall bootstrap support (Figure 4 .6A). The second topology was produced when *Crassostrea* was used as an outgroup (Figure 4.6B). The third topology was produced when *Mytilus* was used as an outgroup (Figure 4.6C).

MP topology 1 (Figure 4.6A) grouped haliotid actins into two clades: clade 1 contained *H.irisA2, H.irisA3* and *H. discus hannai,* clade 2 contained all other haliotid actins identified in this study. There was 100% bootstrap support for this division. Within each clade actins were grouped by species; in clade 1 the two *H. iris* actins were more closely related to each other than they were to the *H. discus hannai* actin, in clade 2 the *H. virginea* actins clustered together. *H. rufescens* actin was the most divergent lineage in the group. Within clade 2 *H.irisA1* and *H.irisA1c* were closely related and *H.virgA1a* and *H.virgA1b* were closely related.

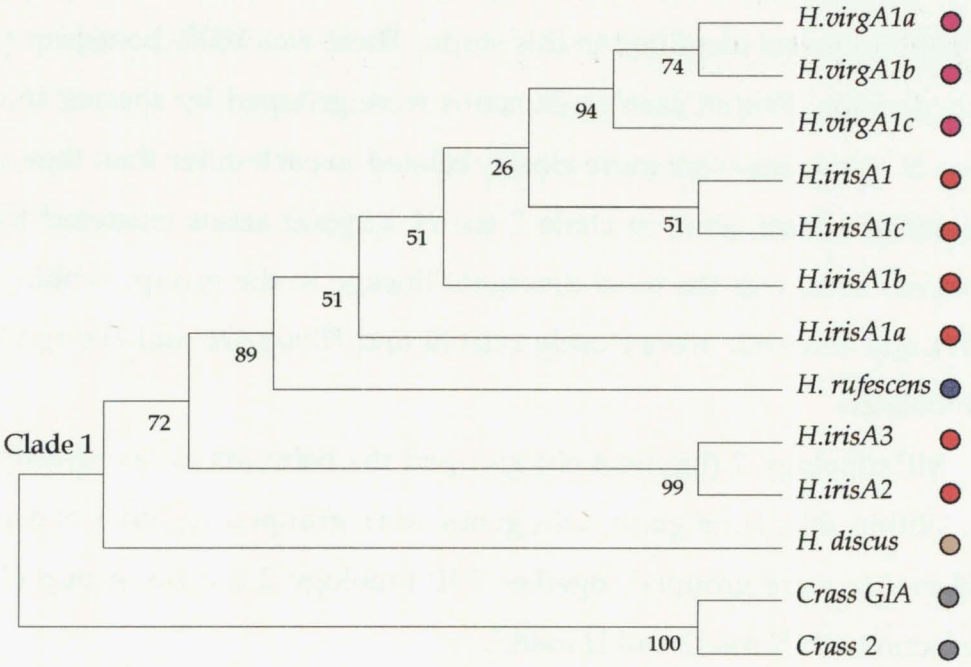
MP topology 2 (Figure 4.6B) grouped the haliotids in one monophyletic clade. Within this, *H. virginea* actin genes were grouped together and *H.irisA1* and *H.irisA1c* were grouped together. MP topology 2 did not group *H. discus hannai* actin with *H.irisA2* and *H.irisA3.*

MP topology 3 (Figure 4.6C) grouped the haliotids into two clades. Clade 1 contained the *H. discus hannai* actin gene, *H.irisA2* and *H.irisA3.* Clade 2 contained the *H. virginea* actin genes, the *H. rufescens* actin gene and the remaining *H. iris* actin genes. Within clade 2 the *H. virginea* actin genes formed a group and *H.irisA1a, H.irisA1* and *H.irisA1c* formed a group.

A. Topology 1



B. Topology 2



C. Topology 3

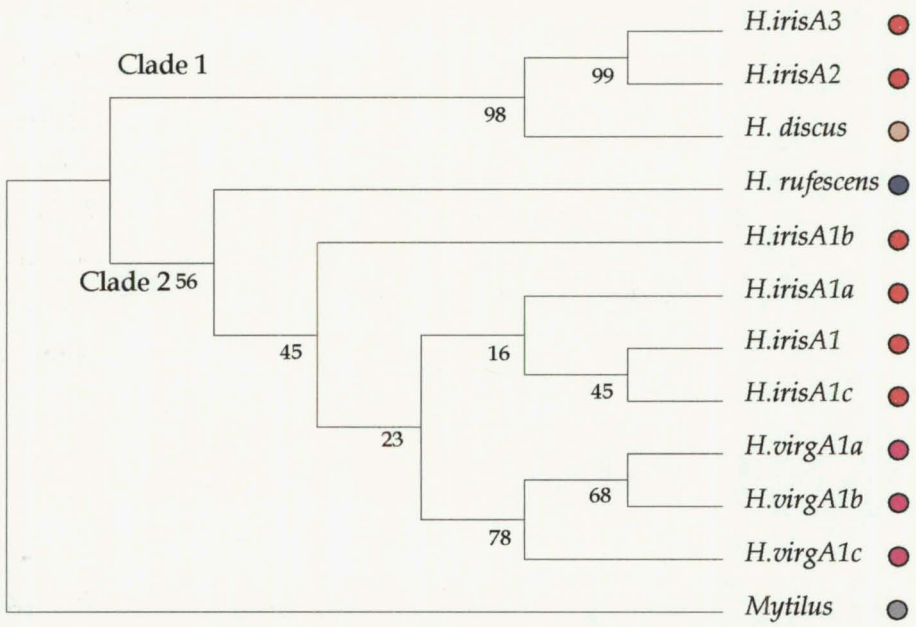


Figure 4.6. Alternative topologies of *Haliotis* actin gene MP phylogenetic trees reconstructed with different mollusc outgroups. Species are colour coded for ease of identification. The results of 500 bootstraps are shown **A. Topology 1.** MP tree obtained with *Placopecten, Aplysia, B. glabrata, Placopecten* or *Dreissena* as outgroups. A representative tree obtained with *Aplysia* is shown. **B. Topology 2.** MP tree obtained with *Crassostrea* as the outgroup. **C. Topology 3.** MP tree obtained with *Mytilus* as the outgroup.

Maximum Likelihood

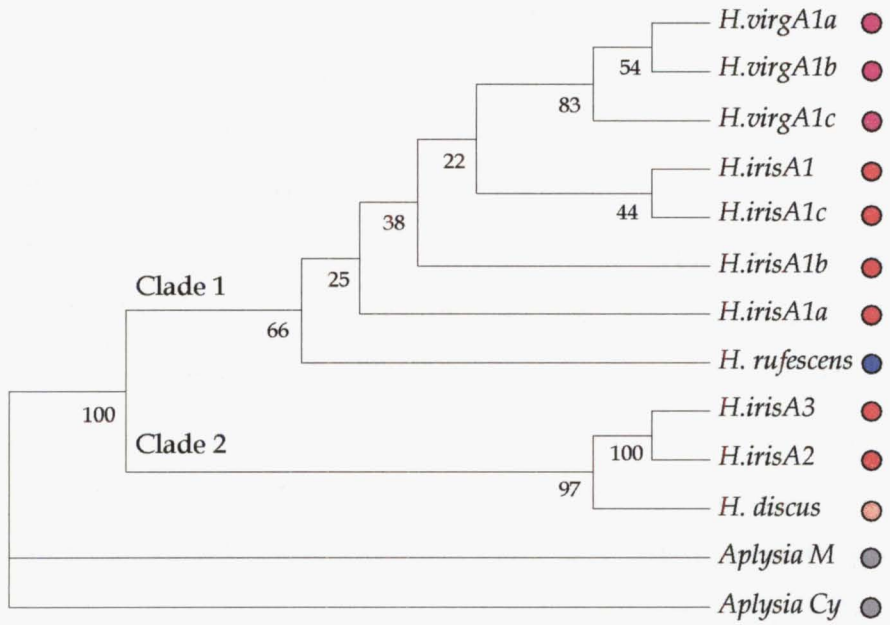
The phylogenetic relationship between haliotid actins was also investigated by ML using seven alternative mollusc species as outgroups. ML inference produced three tree topologies (Figure 4.7). The first topology was supported by *Patella, Aplysia, Placopecten* and *B. glabrata* as outgroups. The ML trees varied in bootstrap support, with different trees supporting some lineages more strongly than others, for example, supports for the *H. rufescens*/haliotid split were 58, 80, 94%, respectively, when *Patella, B. glabrata* or *Placopecten* were the outgroups. The *Aplysia* MP tree was chosen as a representative tree (Figure 4.7A). The second topology was produced when *Crassostrea* was used as the outgroup (Figure 4.7B). The third topology was produced when *Mytilus* was used as the outgroup (Figure 4.7C).

ML topology 1 (Figure 4.7A) divided the haliotid actin genes into two clades. *H.irisA2, H.irisA3* and *H. discus hannai* actin formed one clade (clade 2). The remaining *H. iris* actin genes, the *H. virginea* actin genes and the *H. rufescens* actin gene formed the second clade (clade 1). The *H. virginea* actin genes formed a group within clade 1.

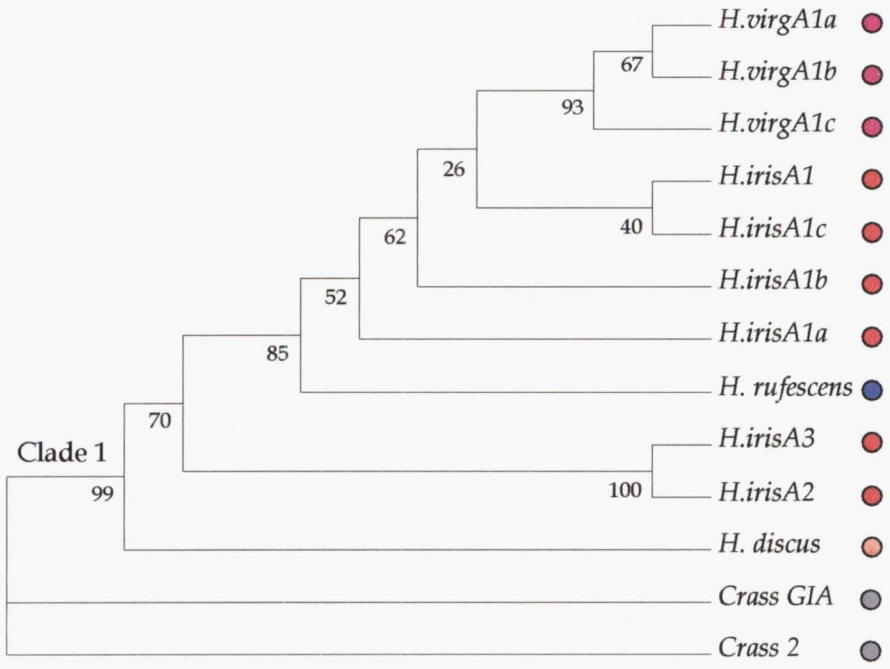
ML topology 2 (Figure 4.7B) grouped the haliotid actin genes as a monophyletic clade. Within this clade *H.irisA2* and *H.irisA3* grouped together, *H.irisA1* and *H.irisA1c* grouped together and the *H. virginea* actin genes grouped together.

ML topology 3 (Figure 4.7C) grouped the haliotid actin genes into two clades. Clade 2 contained the *H. virginea* actin genes. Clade 1 contained the remaining haliotid actin genes. Within clade 1 *H.irisA2, H.irisA3* and *H. discus hannai* actin grouped together and *H.irisA1* and *H.irisA1c* grouped together.

A. Topology 1



B. Topology 2



C. Topology 3

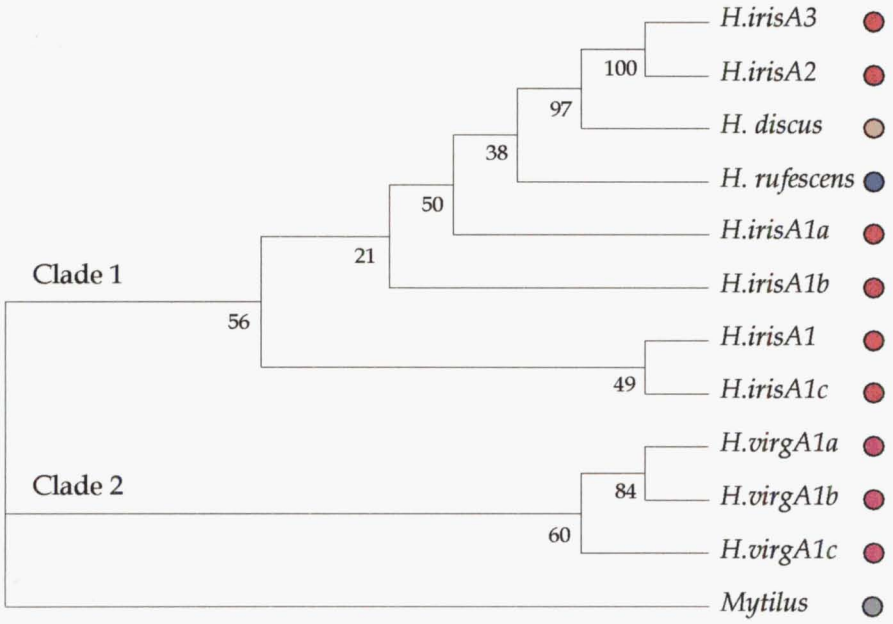


Figure 4.7. Alternative topologies of *Haliotis* actin gene ML phylogenetic trees reconstructed with different mollusc outgroups. Species are colour coded for ease of identification. The results of 100 bootstraps are shown. **A. Topology 1.** ML tree obtained with *Patella, Aplysia, B. glabrata, Placopecten* or *Dreissena* as outgroups. A representative tree obtained with *Aplysia* is shown. **B. Topology 2.** ML tree obtained with *Crassostrea* as the outgroup. **C. Topology 3.** ML tree obtained with *Mytilus* as the outgroup.

HALIOTIDS AND GASTROPODS

The phylogenetic relationships between the actin genes of haliotids and other gastropods were investigated by MP (Figure 4.8A) and ML (Figure 4.8B).

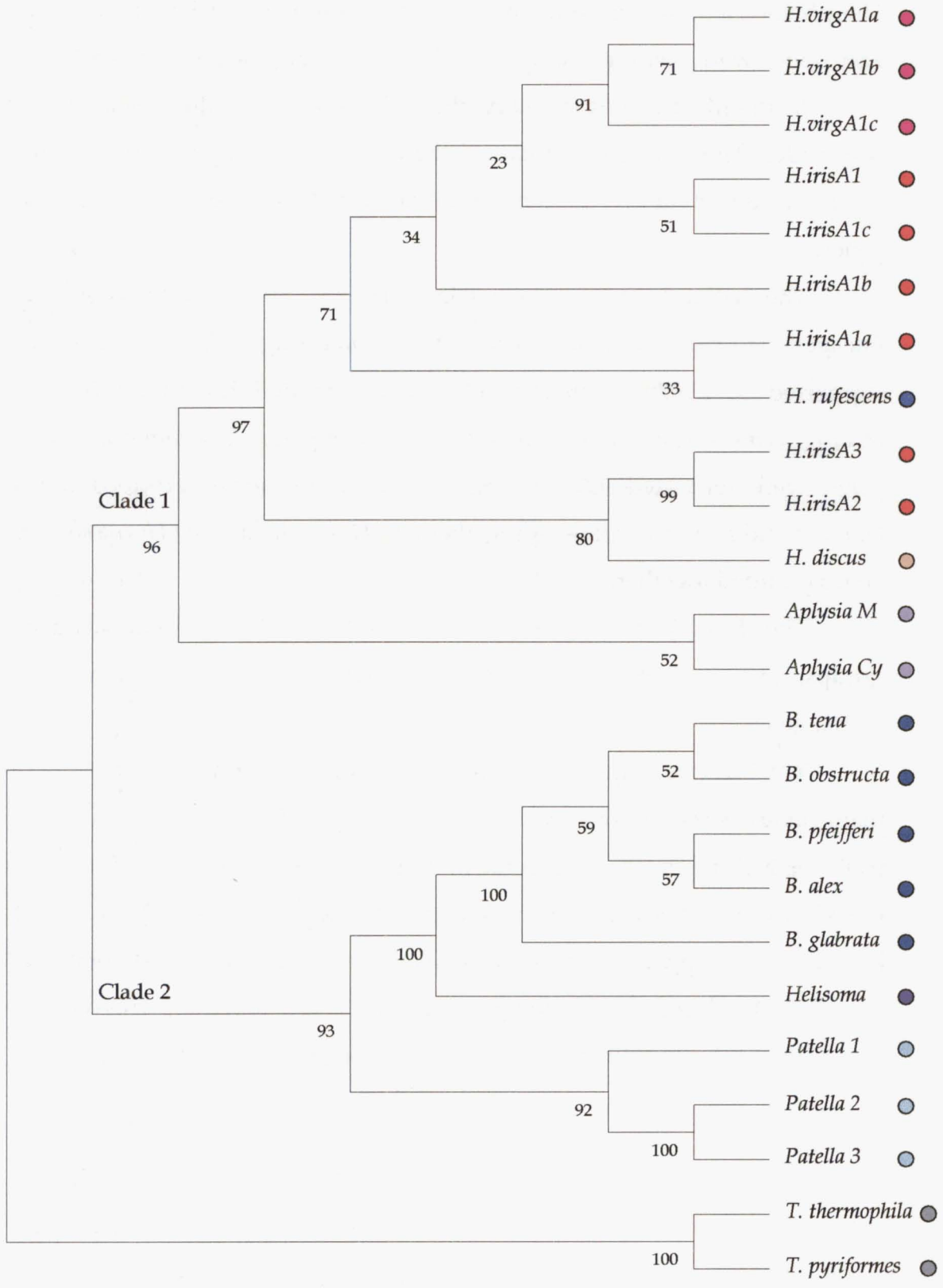
In the MP tree (Figure 4.8A) the actin genes of molluscs separated into two clades. The first clade contained the haliotid and *Aplysia* actin genes. The second clade contained of the *Biomphalaria/Helisoma* genes and the *Patella* genes.

The haliotid/*Aplysia* clade (clade 1) was further divided into three lineages, two haliotid lineages and one *Aplysia* lineage. The haliotid split was supported by a 97% bootstrap value and separated *H.irisA2*, *H.irisA3* and *H. discus hannai* from the remaining haliotid actin genes. The remaining haliotid genes were separated into groups; the *H. virginea* genes grouped together, *H.irisA1* and *H.irisA1c* grouped together and *H.irisA1a* and the *H. rufescens* actin genes grouped together.

The *Biomphalaria/Helisoma/Patella* clade (clade 2) divided into two groups, one containing the *Biomphalaria* genes, the other containing the *Patella* genes.

The ML tree (Figure 4.8B) divided the mollusc actin genes into the same main clades as the MP tree, with 100% bootstrap support. The ML tree differed from the MP tree in the arrangements of genes within clade 2. In the ML tree, the relationships between *Biomphalaria* genes observed in the MP tree were lost, but bootstrap support for these divisions were low in both the MP and the ML tree. The ML placed *Patella* actin genes 1 and 2 closer together, whereas the MP tree grouped *Patella* actin genes 2 and 3 closer together.

A. MP Consensus Tree



B. ML Consensus Tree

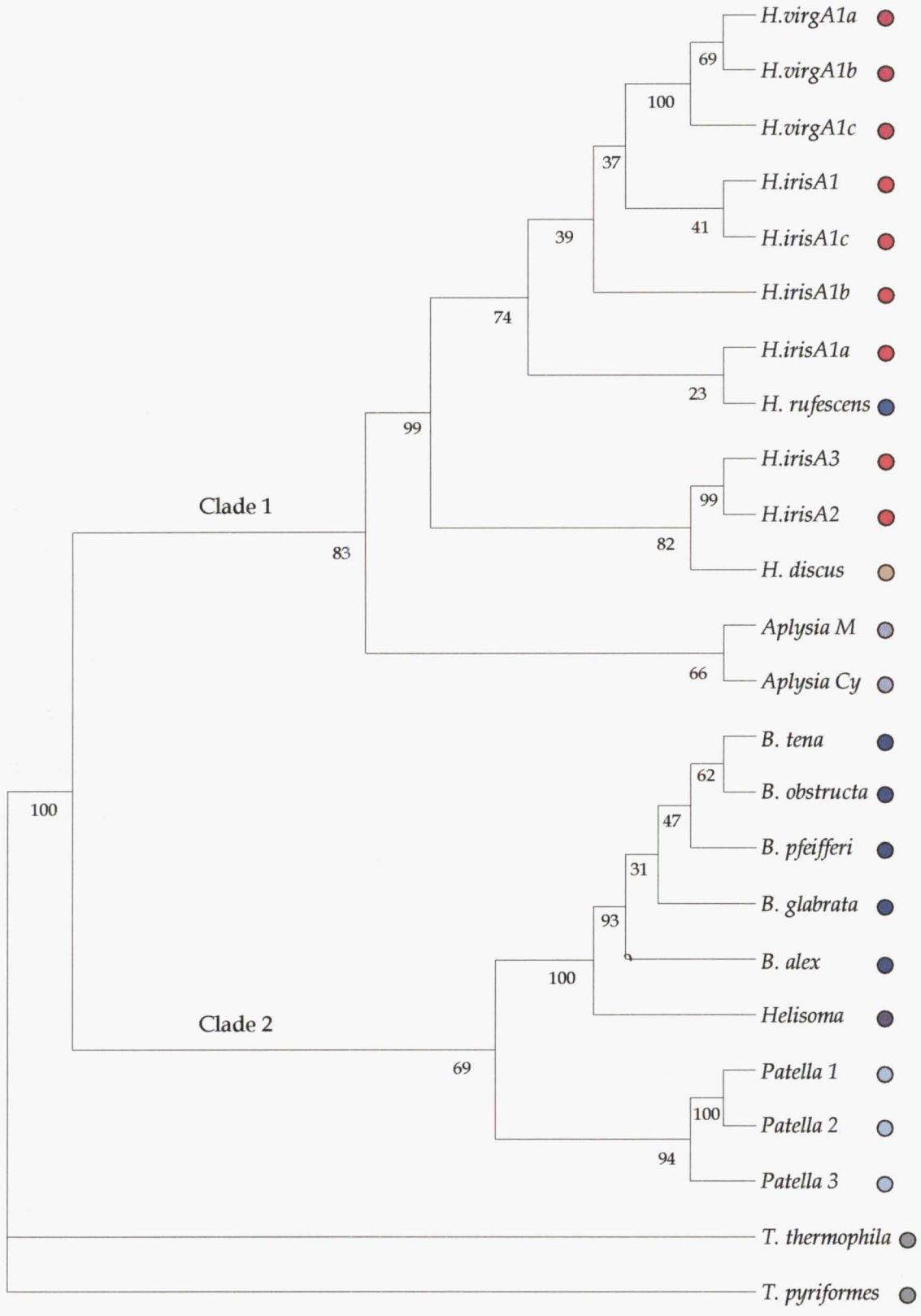


Figure 4.8. Phylogenetic reconstruction of *Haliotis* and gastropod actin genes with *Tetrahymena* spp. as outgroups. Haliotid species are colour coded for ease of identification. **A. MP Consensus Tree.** Results of 500 bootstraps are shown. **B. ML Consensus Tree.** Results of 100 bootstraps are shown.

HALIOTIDS AND CEPHALOPODS

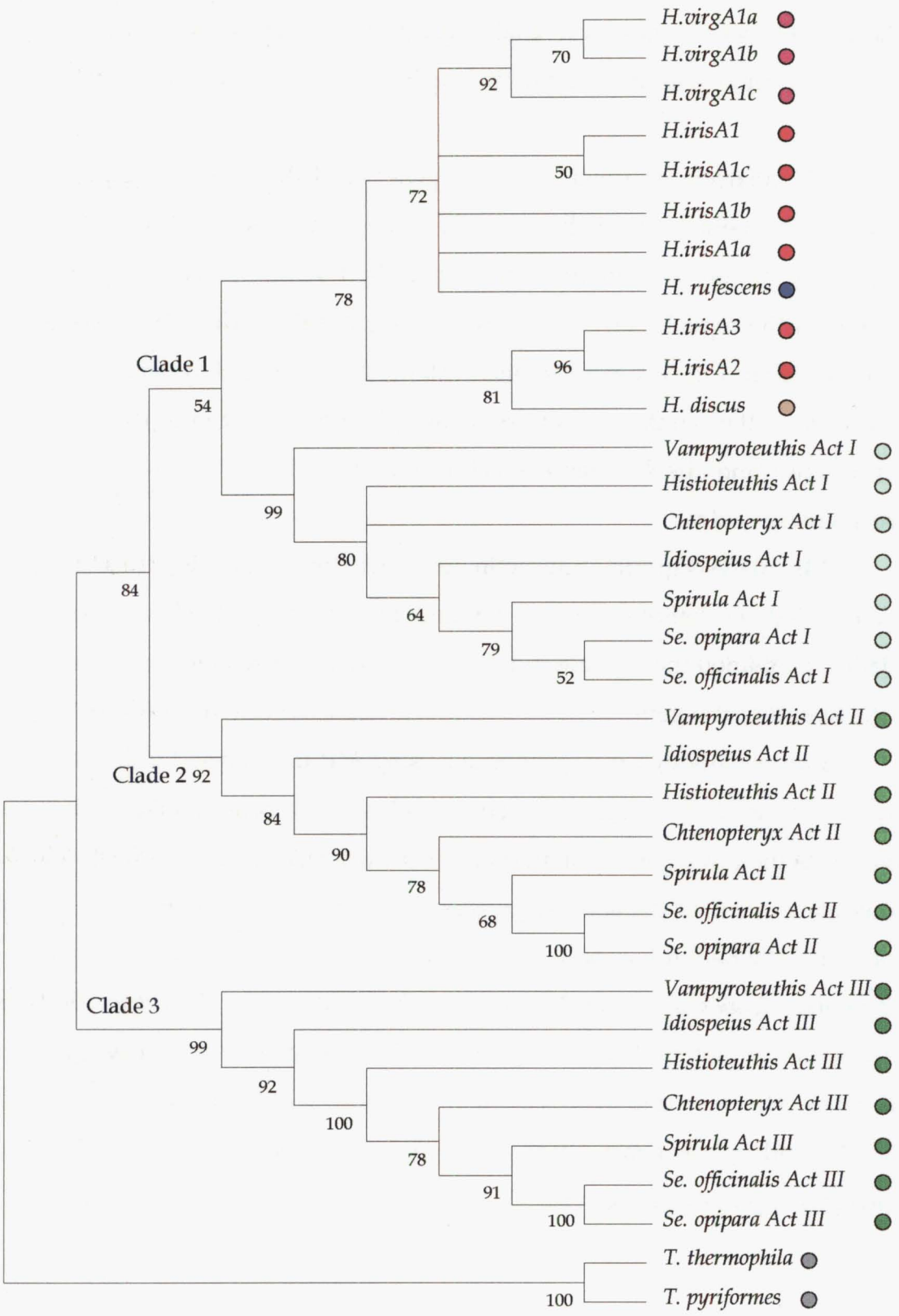
The haliotid actin genes were analysed by MP (Figure 4.9A) and ML (Figure 4.9B) with representative cephalopod actin gene sequences from Carlini *et al.* (2000).

MP analysis grouped the actin genes into three clades (Figure 4.9A). Clade 1 contained the haliotid actin genes and the cephalopod *Actin I* sequences. Clade 2 contained the cephalopod *Actin II* sequences. Clade 3 contained the cephalopod *Actin III* sequences. Within clade 1 the haliotid actin genes divided into two groups, one containing *H.irisA2*, *H.irisA3* and *H. discus hannai*, the other containing the remaining *H. iris* actin genes, the *H. virginea* actin genes and the *H. rufescens* actin gene. *Crassostrea GIA* was not placed within these clades.

ML analysis grouped the actin genes into two clades (Figure 4.9B). Clade 1 contained the haliotid actin genes and the cephalopod *Actin II* sequences. Clade 2 contained the cephalopod *Actin I* and *Actin III* sequences. Within Clade 1 the haliotid actin genes divided into two groups, as seen in the MP tree.

Further analysis was performed using MP to analyse the haliotid actin genes, the cephalopod actin sequences and mollusc actin genes, an approach similar to the phylogenetic analysis performed by Carlini *et al.* (2000). The MP consensus tree split the actin sequences into three clades (Figure 4.10). Clade 1 contained the haliotid, *Aplysia*, cephalopod *Actin 1*, *Placopecten*, *Mytilus* and the *Crassostrea 2* genes. Clade 2 contained the *Biomphalaria/Helisoma*, *Dreissena*, *Patella* and cephalopod *Actin II* genes. Clade 3 contained the cephalopod *Actin III* genes. The bootstrap supports for these clades were low. Within clade 1 the haliotid and *Aplysia* genes separated from the *Placopecten*, *Mytilus*, *Crassostrea 2* and cephalopod *Actin II* genes.

A. MP Consensus Tree



B. ML Consensus Tree

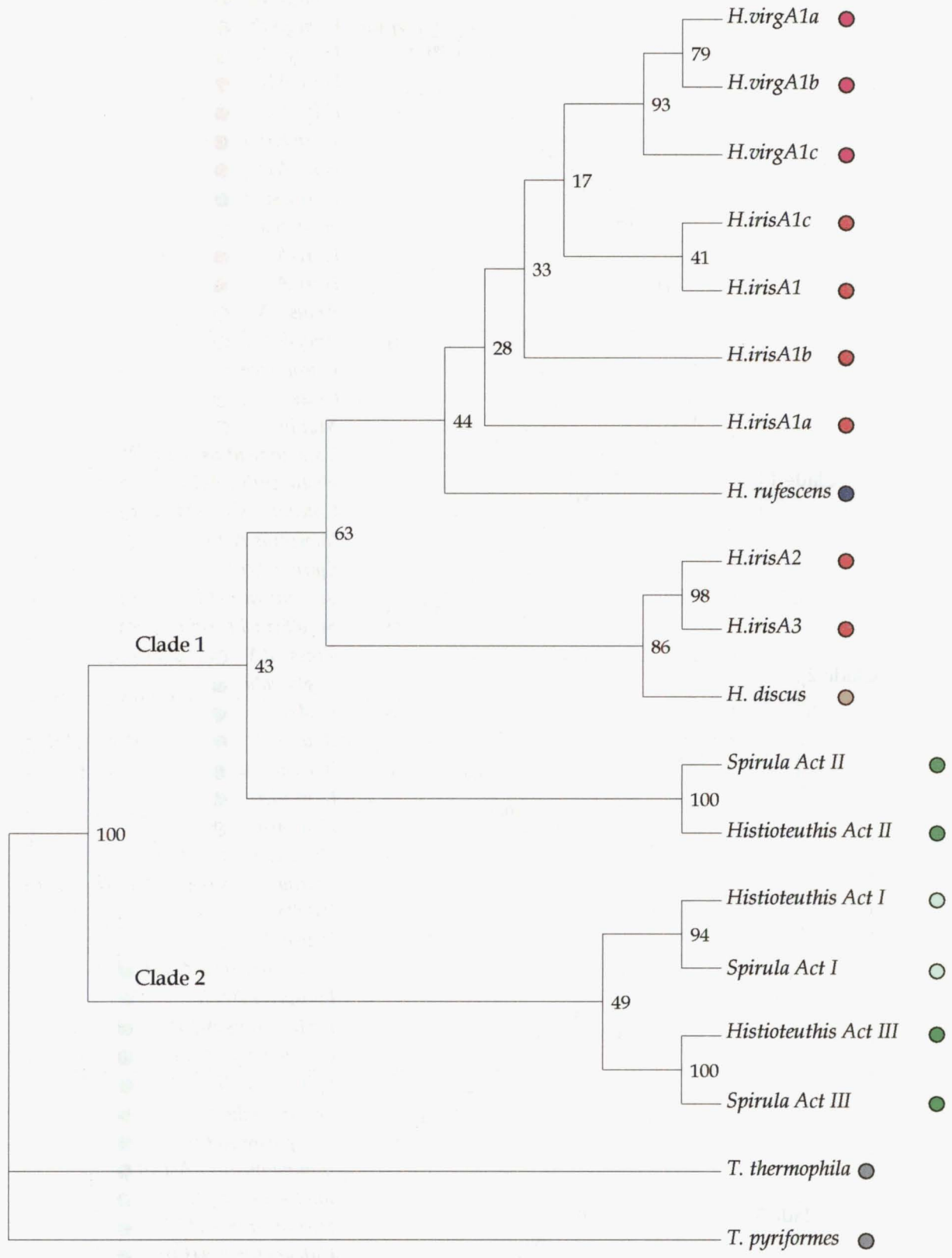


Figure 4.9. Phylogenetic reconstruction of *Haliotis* and cephalopod actin genes with *Tetrahymena* spp. as outgroups. Haliotid species are colour coded for ease of identification. Cephalopod actins are colour coded in green depending on Carlini *et al.*'s (2000) definition of actin gene type, rather than by species. **A. MP Consensus Tree.** Results of 500 bootstraps are shown. **B. ML Consensus Tree.** Results of 100 bootstraps are shown.

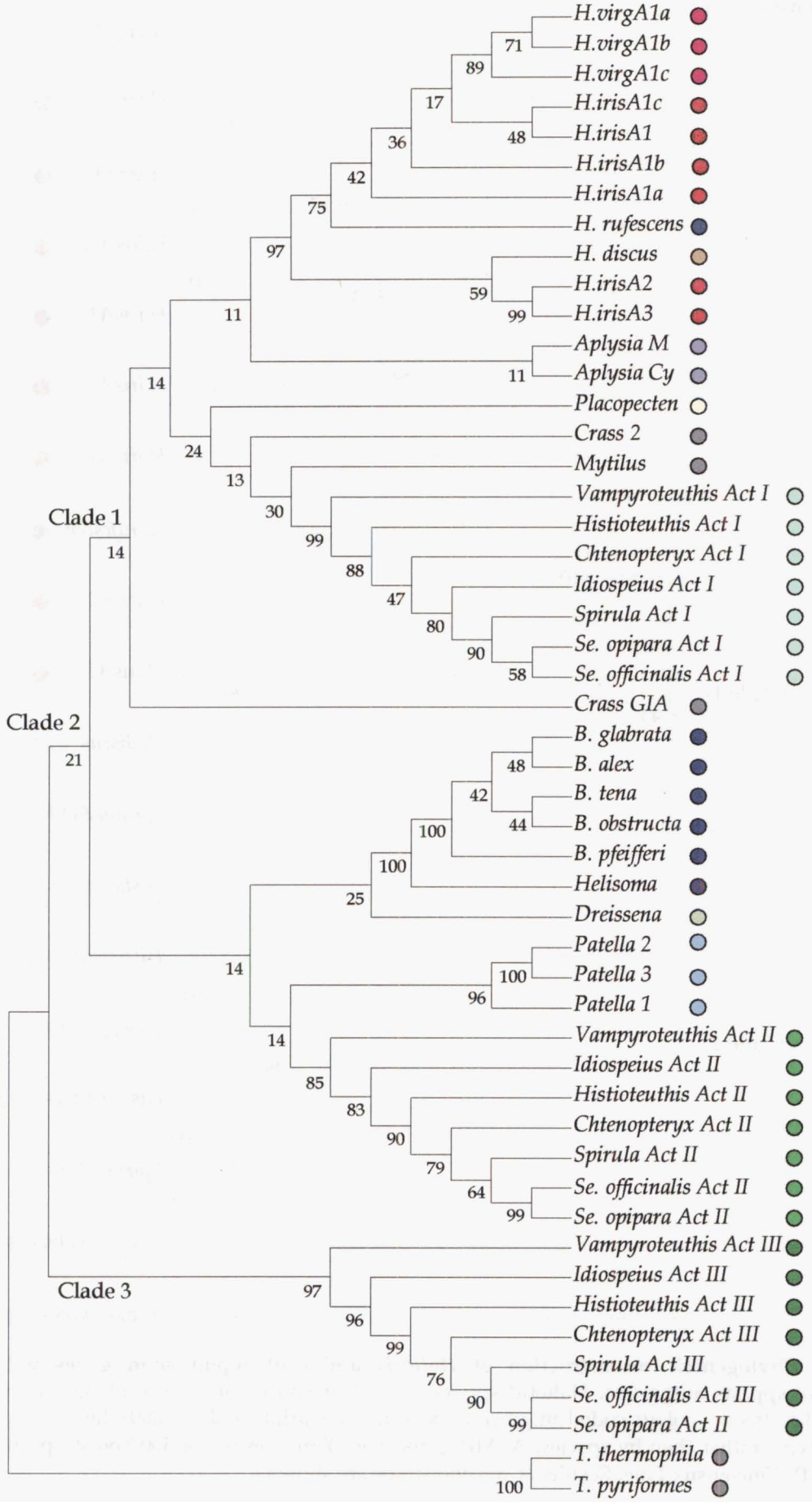

H.virgA1a
H.virgA1b
H.virgA1c
H.irisA1c
H.irisA1
H.irisA1b
H.irisA1a
H. rufescens
H. discus
H.irisA2
H.irisA3
Aplysia M
Aplysia Cy
Placopecten
Crass 2
Mytilus
Vampyroteuthis Act I
Histioteuthis Act I
Chtenopteryx Act I
Idiospeius Act I
Spirula Act I
Se. opipara Act I
Se. officinalis Act I
Crass GIA
B. glabrata
B. alex
B. tena
B. obstructa
B. pfeifferi
Helisoma
Dreissena
Patella 2
Patella 3
Patella 1
Vampyroteuthis Act II
Idiospeius Act II
Histioteuthis Act II
Chtenopteryx Act II
Spirula Act II
Se. officinalis Act II
Se. opipara Act II
Vampyroteuthis Act III
Idiospeius Act III
Histioteuthis Act III
Chtenopteryx Act III
Spirula Act III
Se. officinalis Act III
Se. opipara Act II
T. thermophila
T. pyriformes
Clade 1
Clade 2
Clade 3

Figure 4.10. MP phylogenetic reconstruction of haliotid, cephalopod and mollusc actin genes with *Tetrahymena* spp. as outgroups. The mollusc species and the cephalopod actin gene types (as defined by Carlini *et al.* (2000)) are colour coded for ease of identification. The result of 500 bootstraps is shown.

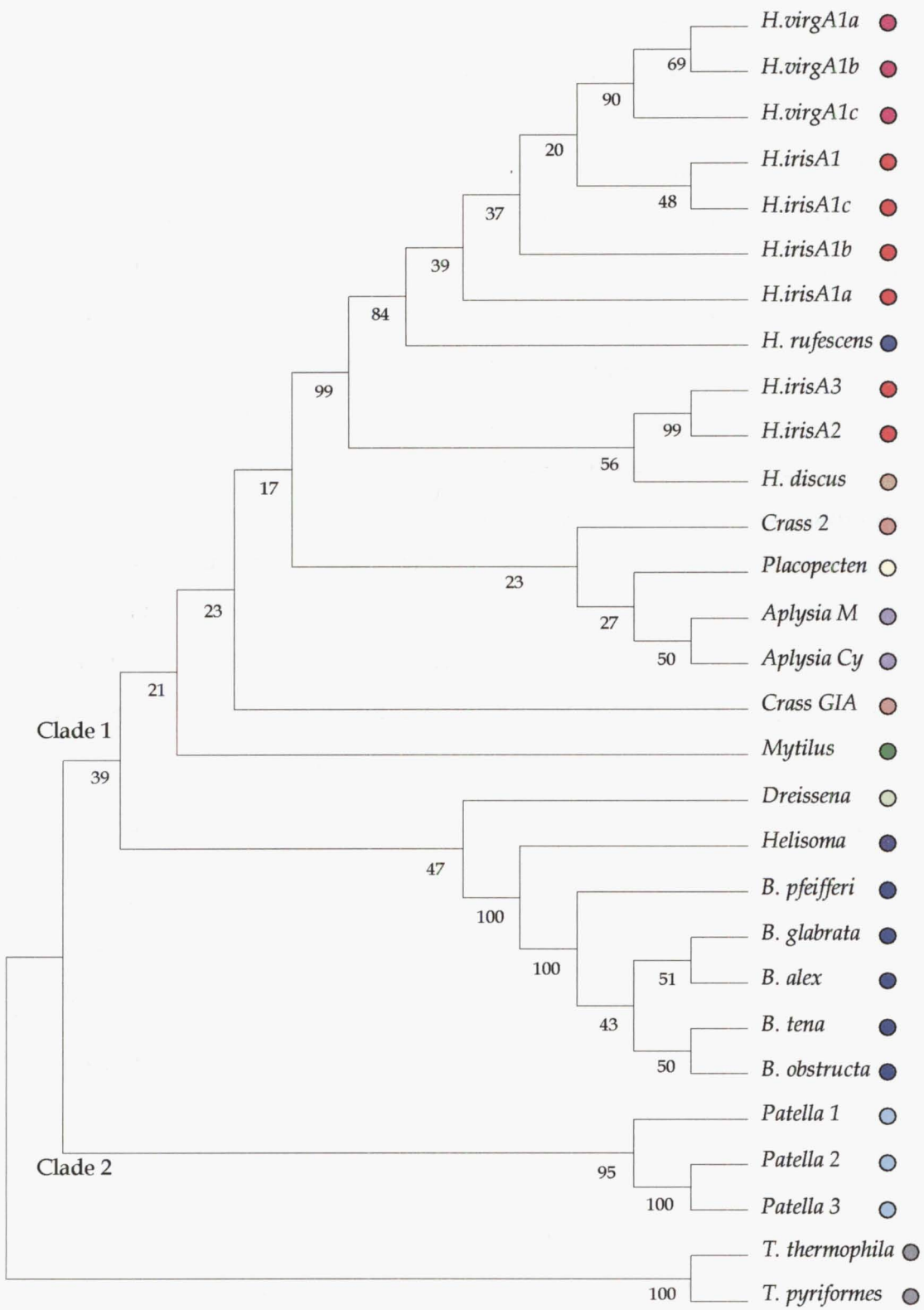
HALIOTIDS AND MOLLUSCS

The relationships between haliotid and mollusc actin genes were investigated by MP and ML. Cephalopods were excluded from this comparison as they were previously analysed (Figures 4.9 and 4.10) and their inclusion would have made ML phylogenetic analysis too computationally intensive.

The MP consensus tree (Figure 4.11A) divided haliotid actin genes into two clades. Clade 1 was poorly resolved and contained all mollusc actin genes, except for the *Patella* actin genes. Within clade 1 the haliotid actin genes were grouped together with high support (99% bootstrap value). Clade 2 contained the three *Patella* actin genes.

The ML consensus tree (Figure 4.11B) also grouped the haliotid actin genes into two clades. Clade 1 contained all mollusc actin genes except for the *Patella* and *B. glabrata* actin genes. Within clade 1 the haliotid actin genes grouped together, the *Aplysia* actin genes grouped together and the bivalve actin genes grouped together. The bivalve genes further divided into two groups, *Crassostrea 2, Mytilus* and *Placopecten,* and *Crassostrea GIA* and *Dreissena.*

A. MP consensus tree



B. ML consensus tree

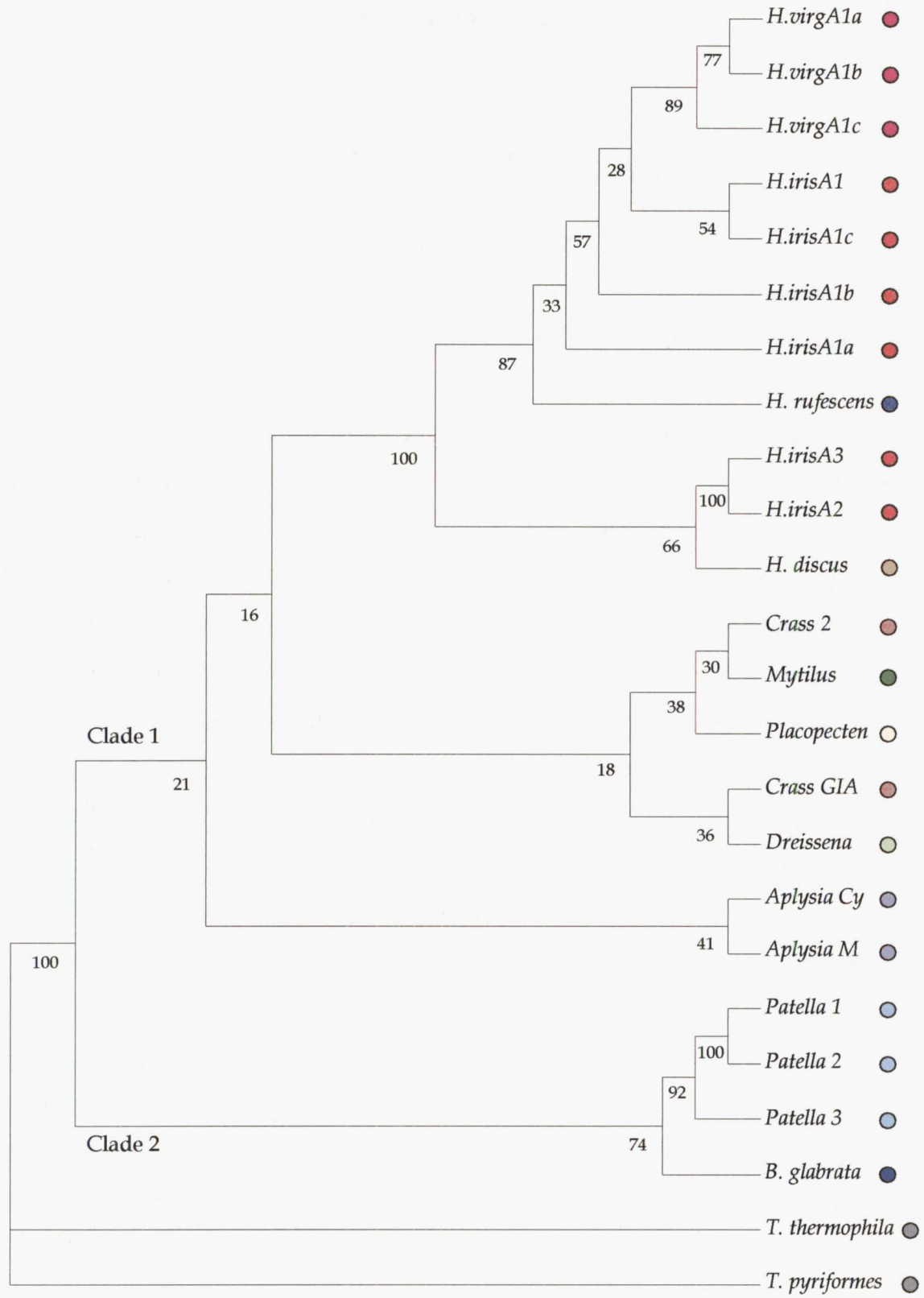


Figure 4.11. Phylogenetic reconstruction of *Haliotis* and molluscs with *Tetrahymena* spp. as outgroups. Haliotid species are colour coded for ease of identification. **A. MP consensus tree.** Results of 500 bootstraps are shown. **B. ML consensus tree.** Results of 100 bootstraps are shown.

DISCUSSION

NUCLEOTIDE AND CODON BIAS

NUCLEOTIDE BIAS

The nucleotide frequencies were very similar across haliotid species, there was less than 4% difference in the G+C content across the 11 actin sequences isolated from four species.

When nucleotide frequencies deviate from the expected values of 50% G+C and 50% A+T nucleotide bias is indicated. The nucleotide frequencies within haliotids indicated that nucleotide bias is present, although the nucleotide frequencies for overall codons were close to neutrality (51-55% G+C), the nucleotide frequencies at each codon position showed biases. At the first codon position there was a small G+C bias, with G+C content reaching 60% in *H. discus hannai*. At the second codon position there was a small A+T bias, with A+T nucleotide content reaching 62% in *H. rufescens* and *H. discus hannai*. At the third codon position there was a strong G+C bias, with G+C content reaching 68% in *H. iris*, *H. virginea* and *H. rufescens*. At the first position the increased G+C content was due to a higher G composition, whereas at the third position the increased G+C content was due to a higher C content. The A+T bias at the second position was contributed to by higher A and T contents.

The nucleotide compositions for 13 actin genes from nine mollusc species were compared (the *Haliotis* and *Biomphalaria* nucleotide compositions were averages for each genus). The nucleotide composition in other mollusc species showed greater variation than that in haliotids, with up to 16% difference between the G+C contents at the third codon position. Not only did molluscs vary in the degree of their nucleotide bias, they also differed in the nucleotides they favoured. At the third codon position *Patella*, *Biomphalaria* and *Helisoma* showed an A+T bias, whereas all other mollusc species showed a G+C bias.

In molluscs there was a small G+C bias at the first codon position and a small A+T bias at the second position. The nucleotide bias at the third codon varied between an A+T preference (five genes) or a G+C preference (eight genes) in the molluscs studied. The third codon position bias was very large in some species, the G+C content reached 80% in the *Aplysia* muscle actin gene. The G+C bias was also very high in the *Aplysia Cy*, *Placopecten* and haliotid actin genes (G+C contents of 77, 77 and 66%, respectively).

The phenomenon of nucleotide bias has been well documented in prokaryotes and eukaryotes, and is thought to be caused by bias mutation pressure due to differences in the forward and backward mutation rates between nucleotides (Sueoka, 1962). In *Drosophila* there is a bias toward an increasing G+C content (Shields *et al.*, 1989). In *Bacillus subtilis* there is a bias toward an increasing A+T (Shields and Sharp, 1987).

The degree of nucleotide bias is often strongest in the third codon position, as observed in the mollusc actin genes. This is due to the functional constraint in the protein influencing what nucleotide substitutions may occur, at the third codon position substitutions are often synonymous and will occur at high frequencies as the amino acid encoded will not be altered. If substitutions do not affect the protein, the substitutions that occur may be influenced by differences in mutation rates, causing nucleotide biases to result. Only a small number of substitutions at the first codon position, and no substitutions at the second codon position, will be synonymous, therefore there is little opportunity for bias to accumulate at these positions. However, in haliotids and molluscs a significant A+T bias was observed at the second codon position. Causes of A+T bias at the second codon position cannot be explained as they can in non-coding DNA such as introns and pseudogenes.

The nucleotide frequencies in the haliotid actin gene sequences were calculated. Introns showed a trend toward an A+T bias. This trend has also been observed in pseudogenes. Pseudogenes, which are also non-coding regions of DNA like introns, have a transition rate almost 2.5 times higher than the rate expected for random substitution patterns (Li, 1997). Of the possible

transitions that can occur, changes from C to T and changes from G to A occur more frequently, creating an A+T bias. The increased A+T bias also may be due to nucleotide chemistry, thymine can form from the deamination of cytosine (Coulondre *et al.*, 1978), therefore substitutions to thymine may result from a mechanism other than transcriptional error. In non-functional DNAs, such as pseudogenes and introns, the frequency of C deamination may increase, due to lack of functional constraint, leading to the higher A+T frequencies (Li *et al.*, 1985).

The nucleotide bias was shown to have a drastic effect on the phylogenetic reconstruction of dipteran actin genes using distance methods; when synonymous substitutions were used for analyses the orthologous relationships between dipterans were lost, due to the high G+C bias in *Drosophila* causing the paralogous *Drosophila* actin genes to group together (He and Haymer, 1995). In this study, phylogenetic analyses within haliotids should not suffer from this nucleotide bias, as the haliotid group showed similar nucleotide compositions. Nucleotide frequency biases may affect the appropriateness of the use of some mollusc species as outgroups, due to greater nucleotide composition variability within molluscs. Phylogenetic reconstruction by maximum likelihood (ML) should account for this effect.

CODON BIAS

The RSCU values indicated that codon usage bias existed in haliotid actin genes. The RSCU value is a proportional number which allows comparisons of codon usage across different genes. The RSCU value is expected to be one when all codons for an amino acid are used equally.

The patterns of codon usage among haliotids were similar, but not identical. For amino acids where codon bias varied, some differences were due to species changing their preferred codon, for example *H. rufescens* and *H. discus hannai* change from GUC to CUU (valine). Some differences were due to loss of amino acid bias, for example *H. virginea* does not show a bias for tyrosine UAC (although the biases for tyrosine in the other haliotids are small), *H. rufescens*

and *H. discus hannai* do not show a bias for alanine GCC and *H. iris* does not show a bias for glutamic acid GAG. Some differences were due to species gaining a bias where biases were not found in other species, for example, *H. discus hannai* shows a bias for the histidine codon CAC and *H. virginea* shows a bias for the cysteine codon UGC. Other differences were due to multiple codons being preferred, where only one codon is preferred in other species, for example *H. rufescens* and *H. discus hannai* showed a bias for the leucine codons CUC (as in other haliotids) and CUG, *H. iris* and *H. virginea* showed a bias for alanine GCU (as in other haliotids) and GCC.

For some amino acids, such as glycine and isoleucine, the codon preference was similar across molluscs, but some species show changes in preferences. In some cases the preferred codon changed in some species, such as isoleucine (change from AUC in molluscs to AUU in *Biomphalaria*). In other cases the number of preferred codons changed in some species. The number of preferred codons for threonine increased in *Biomphalaria* and *Helisoma* (molluscs favour ACC, *Biomphalaria* and *Helisoma* favour ACC and ACA). The number of preferred codons for arginine decreased in *Aplysia* and *Mytilus* (molluscs favour two codons, CGU/CGA and AGA/AGG, whereas *Aplysia* and *Mytilus* respectively favour CGU and AGA only).

For other amino acids the codon bias was variable across species. For leucine there are five patterns of codon usage across the 12 molluscs examined. For other sites the presence of codon usage bias was variable, in tyrosine seven species show bias whereas five do not.

Codon bias has been well documented in other species, including *E. coli* and *Drosophila* (Sharp and Li, 1986; Sharp *et al.*, 1988). In *E. coli*, codon usage bias increases in genes that are expressed at higher levels. This pattern of codon usage is consistent with selection for translational efficiency, as synonymous substitutions lead to a bias for codons that bind abundant tRNA molecules. In *Drosophila* codon bias also appears to be driven by translational efficiency, codon bias is higher in genes which are highly or rapidly expressed and favours

codons that bind the most abundant tRNA molecules (Powell and Moriyama, 1997).

In haliotid actin genes the degree of codon usage bias for an amino acid is correlated with the frequency of the amino acid within the actin protein. This may reflect selection of codons in relation to tRNA abundance, to increase translational efficiency (Ikemura and Ozeki, 1982). The frequencies of haliotid tRNA molecules are unknown, so a correlation between the codons favoured and the abundance of their tRNAs cannot be established. A possible cause for selection toward increased translational efficiency could be the high expression levels of actin (Kanaya *et al.*, 1999). If translational efficiency is present in mollusc actin genes, the tRNA frequencies within molluscs must vary, as molluscs show variability in the codons favoured.

TRANSITIONS/TRANSVERSIONS

Analysis of the transition/transversion ratio (R) shows that these types of nucleotide substitution do not occur equally. In haliotids and molluscs there is a large range in R. Within molluscs R ranged from 0.84-2.94. R was more variable in haliotids, with values ranging from 0.92-5.25. Analysis of R values in nuclear genes of other species shows that R is usually between 0.5-2, although R values as high as 15 have been found in mitochondrial DNA (Nei and Kumar, 2000). Like differences in nucleotide composition, the variability in R is caused probably due to mutational biases and the degeneracy of the genetic code (a greater proportion of transitions, than transversions, are synonymous) (Li *et al.*, 1985).

Saturation plots showed that transitions and transversions had not reached saturation in haliotids. Saturation can be a problem when performing phylogenetic analysis, as some evolutionary changes will be obscured by the presence of multiple mutations at nucleotide sites (Li, 1997). The variability in R is a problem for phylogenetic analysis by MP, as differences within species may erroneously affect phylogenetic reconstruction (Gaultier and Gouy, 1995).

Variability in R would be less of a problem in ML analyses, as a matrix of substitution rates and invariant corrections are incorporated into evolutionary models that infer phylogeny.

STATIONARITY

The disparity index test showed that there is little non-stationarity within haliotids, but there is non-stationarity within molluscs. As for transition/transversion ratio variability, the effect of non-stationarity would be more problematic for phylogenetic reconstruction of molluscs than haliotids. The non-stationarity would be best accounted for in ML analyses, where invariant corrections can be applied to evolution models. The implication for MP phylogenetic analyses is that the choice of mollusc outgroup for haliotid comparisons, and the species included in mollusc analyses, may influence the results due to non-stationarity effects. Conclusions drawn from analyses where non-stationarity occurs should be considered accordingly.

GENE CONVERSION

Gene conversion has been implicated in the evolution of *S. purpuratus* and *Pisaster* actin genes (Drouin, 2002). In this study the GENECONV program found evidence of gene conversion when nucleotides were compared, but not when the codons were compared. Codon analysis is more conservative, as nucleotide similarity may be the result of mutational biases, therefore the use of codon analysis are more valid. The codon analysis suggests that gene conversion has not been involved in the evolution of the *H. iris* actin gene family.

GENECONV detected high similarity between nt 250-325 of *H.irisA2* and *H.irisA3*. This similarity maybe the result of functional constraint, nucleotide biases or shared evolutionary history, rather than homogenisation by gene conversion. In *Drosophila* actin genes the possibility of gene conversion was investigated as a cause of the high similarity between paralogous genes. In this

study the gene similarity was attributed to nucleotide and codon biases rather than gene conversion (He and Haymer, 1995). As nucleotide and codon biases have been observed in haliotid actins, and as *H.irisA2* and *H.irisA3* consistently group together in MP and ML phylogenetic analyses, biases and shared evolutionary history may be the cause of the high similarity between *H.irisA2* and *H.irisA3* actin genes.

POSITIVE SELECTION

Comparison of the nonsynonymous and synonymous substitution rates shows that there is no evidence of positive selection in the *H. iris* actin gene family. Although evidence of positive selection has not been implicated in actin genes, the possibility of its presence was tested as the multiplicity of actin genes within the *H. iris* gene family could be undergoing functional specialisation by positive selection. The d_N and d_S values suggest there is no positive selection acting on actin proteins in the haliotid actin gene family.

Although an excess of transitions to transversions increases the synonymous substitution rate (Li *et al.*, 1985), the large number of transitions observed in haliotids are unlikely to have caused a bias significant enough to hide a positive selection signal, as the difference between the synonymous and nonsynonymous substitution rates per site were very large. Changes in mutational biases and nucleotide frequencies over time in different lineages may cause sequences to look like they have evolved by positive selection, however, comparison of synonymous and nonsynonymous substitution rates within a lineage will be subject to the same mutational and nucleotide biases and should not be prone to this error (Akashi, 1997).

Nielsen and Yang (1998) have used ML to test for the presence of positive selection on amino acids within HIV-I envelope proteins, rather than testing for positive selection for the whole protein, as performed in this study. This approach may be more appropriate, as positive selection may be driving the evolution of selected functionally important amino acids within a protein,

not driving positive selection at all amino acid sites within a protein. This method may also be a more sensitive detector of positive selection, as a positive selection signal at a small number of amino acids may be obscured by the absence of a positive selection signal in the remaining amino acids of the protein. Yang *et al.* (2000) used ML methods to investigate positive selection pressures within the amino acid sites of abalone lysin genes and showed that there is variability in the selective pressures at different sites. It is possible that analysis of individual amino acid sites in actin may detect positive selective forces within functionally important domains of the actin protein.

Comparison of synonymous and nonsynonymous substitutions show that synonymous substitutions are positively correlated with nonsynonymous substitutions in haliotids. Similar correlations have also been found in *Drosophila,* bacteria and mammals (Sharp and Li, 1987; Wolfe *et al.*, 1989; Akashi, 1994). In *Drosophila* the degree of codon bias increased as amino acids diverged, suggesting an increase in synonymous substitutions may act as a way to increase translational efficiency (Akashi, 1994).

PHYLOGENETICS

ACTIN GENE ORTHOLOGY

MP and ML phylogenetic reconstruction of haliotids each produced three topologies. The six MP and ML topologies showed a closely related group of *H. virginea* actin genes, a closely related group of *H.irisA2* and *H.irisA3* genes and a group of *H.irisA1* and *H.irisA1c* genes, but the support for the *H.irisA1* and *H.irisA1c* group was low (40-51% bootstrap value). Four of the topologies grouped *H. discus hannai* with *H.irisA2* and *H.irisA3*.

The groupings produced by haliotid phylogenetic analyses identified two sets of potentially orthologous genes; one set of orthologous genes contained *H.irisA2, H.irisA3* and the *H. discus hannai* actin gene, the other set of orthologous genes contained the remaining *H. iris* actin genes, the *H. virginea*

actin genes and the *H. rufescens* actin gene. These groups may be divided into actin cytoplasmic and muscle-types.

Within the *H.iris/H.virginea/H.rufescens* actin gene clade the relationships between actin genes were poorly resolved. Phylogenetic grouping within this clade showed paralogy may exist between *H.irisA1* and *H.irisA1c* and within the *H. virginea* actin genes, although only the *H. virginea* paralogies were consistently formed and well supported by bootstrapping.

Phylogenetic analyses have been used to identify orthologous relationships within dipterans (He and Haymer, 1995) and cephalopods (Carlini *et al.*, 2000), and in this study, haliotids. The reliability of the haliotid orthologies should be supported by other evidence, such as similar functions of actin genes within orthologous groups, or conserved sequences, as in dipterans. In dipterans, orthologous actin genes display similar expression patterns and share intron patterns. The ability of introns to support orthologous genes in haliotids is limited by the lack of intron variability in molluscs.

The haliotid actin genes were analysed with gastropod actin genes to examine whether orthologous gene relationships existed between genera. MP and ML methods suggest there are two groups of actin genes in gastropods. The orthology of these groupings is difficult to infer from the gastropod phylogeny, as the *Aplysia Cy* and *M* actin genes group together within clade 1. If the clades represented cytoplasmic and muscle-type actin genes the genes of *Aplysia* would have been expected to group in different clades.

The clades defined by gastropod phylogenies may reflect phylogenetic relatedness within mollusc species, rather than orthologous relationships between actin genes. This may be the result of large divergences between the species compared causing species to appear more related. This possibility should have been accounted for by ML analysis, but ML also divided gastropods into apparent species lineages rather than actin gene orthology. Alternatively, the gastropod phylogeny may not have been able to resolve orthologous relationships due to the sample set used - only one known muscle

actin gene was used and this may have lost its phylogenetic content due to the large number of other, possibly cytoplasmic, actin genes.

To investigate whether orthologous groups form when the data set used contains a wider range of actin gene types, the haliotid actin genes were compared to cephalopod actin genes, where orthologous genes have been well described (Carlini *et al.*, 2000). MP and ML analyses grouped mollusc and haliotid actins into ortholgous groups, however, the MP tree grouped the haliotids with the cephalopod *Actin I* genes, whereas the ML tree grouped the haliotids with the cephalopod *Actin II* genes. This difference is due to the phylogenetic approach used (see below).

To further investigate the haliotid and cephalopod actin gene relationships, MP was used to analyse sequences from haliotids, cephalopods and other molluscs. A similar approach was used by Carlini *et al.* (2000), who analysed cephalopod, mollusc and metazoan actin genes by MP of protein sequences. Carlini *et al.* (2000) showed that mollusc muscle actins used in the analysis (*Aplysia M* and *Placopecten actin*) grouped with cephalopod *Actin I* genes and concluded that the cephalopod *Actin I* genes encoded muscle-type actins. In this study, MP analysis grouped three bivalve mollusc actin genes, two *Aplysia* genes and all the haliotid actin genes with the cephalopod *Actin I* genes. The actin genes that grouped with cephalopod *Actin I* were muscle and cytoplasmic types. A ML analysis was not performed as the analysis of 29+ sequences would have been too computationally prohibitive. Considering the discrepancy between the MP and ML results of haliotid/cephalopod analysis, the ML tree of haliotid/cephalopod/molluscs would be expected to disagree with the haliotid/cephalopod/mollusc MP placement of haliotids actin genes with cephalopod Actin I, but rather show a topology similar to the ML haliotid/cephalopod tree.

As the ML method is more robust, and the cephalopod *Actin I* genes are thought to be muscle-type actins, the ML result is more conservative. It is unlikely that the majority of the haliotid actin genes isolated are more like muscle-type actins than cytoplasmic type actins, as few muscle-type actins have

been previously characterised in molluscs. Alternatively, the definition of cephalopod *Actin I* sequences as muscle-type actins based on phylogeny may have been premature, and further expression analysis is required to confirm cephalopod actin type.

When the cephalopod actin sequences were removed and the sequences were analysed by MP and ML, two clades formed. Clade 1 was large and contained the majority of actin genes analysed. Clade 2 was small and contained the *Patella* sequences, which grouped with *B. glabrata* in the ML tree. Within these clades, relationships between mollusc actin genes were seen. In ML clade 1 the haliotid genes grouped together, the Bivalve genes grouped together and the *Aplysia* genes grouped together. The haliotid genes divided into the relationships identified in the haliotid phylogenies, and may contain cytoplasmic and muscle-type divisions. The Bivalve group genes divided into two groups, one group contained *Placopecten actin*, *Placopecten actin* and *Crassostrea 2*. This group could represent bivalve muscle actin genes as *Placopecten* actin is considered a muscle-type actin and *Crassostrea 2* could be a muscle-type actin as it was isolated from a mantle cDNA library, however the actin type of *Mytilus* is unknown. The second group contained *Dreissena* actin and *Crassostrea GIA*. These genes may represent bivalve cytoplasmic actin genes, *Dreissena* has tentatively been assigned a cytoplasmic-actin function, but the function of *Crassostrea GIA* has not been established.

When orthologous actin genes have been detectable by phylogenetic analyses, the actin genes being compared have been within closely related taxa. He and Haymer (1995) used distance methods to show orthology between dipteran actin genes and Carlini *et al.* (2000) used MP methods to show orthology between cephalopod actin genes. In this study putative orthologous actin genes were consistently found within haliotids and putative orthologous actin genes were found within bivalves.

The support for orthologous gene groups between mollusc genera are low. The large divergence times between mollusc species may make the detection of orthology between mollusc genera difficult. Orthology within

closely related groups like dipterans, cephalopods, haliotids and bivalves show much better support, this may be due to the divergence between species being small enough for ancestral orthology to be detected.

HALIOTID PHYLOGENY

The inferred phylogenetic relationships between haliotid actins reflects the evolutionary history of actin genes, not the evolutionary history of haliotids. The haliotid actin genes form two clades, one containing genes from *H. iris*, *H. rufescens* and *H. virginea*, the other containing genes from *H. iris* and *H. discus hannai*.

Phylogenetic analyses of haliotids with rRNA ITS sequences, mitochondrial DNA and lysin gene sequences show that haliotid species are divided into groups based on geography (Lee and Vacquier, 1995; Lee *et al.*, 1995; Coleman and Vacquier, 2002). Of the haliotid species used in this analysis, *H. iris* and *H. virginea* would group in the New Zealand clade and *H. rufescens* and *H. discus hannai* would group within the Californian/Japanese clade. These clades were not evident in the actin phylogenies.

The phylogenetic relationship between actin genes in distinct haliotid species clades supports the conclusion that haliotids contain orthologous actin genes. An ancestral *H.irisA2/H.irisA3/H. discus hannai* actin gene and an ancestral *H.irisA1*-type/*H. virginea/H. rufescens* actin gene must have been present in the haliotid common ancestor.

METHODOLOGY

Phylogenetic analysis of haliotid actins gave similar results when MP and ML were used, suggesting that both methods were detecting the same phylogenetic signals in closely related species. The MP method probably worked as well as the ML method as the assumptions of MP were not violated significantly within haliotids. The choice of outgroup in MP analyses was important as MP assumption violation may occur depending on the species used. Tests of nucleotide bias and stationarity violation showed that *Crassostrea* and *Aplysia*

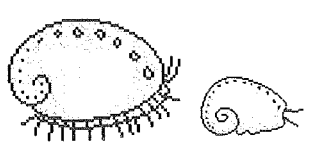
had a similar G+C bias at the third codon position, whereas *Mytilus* had an A+T bias at the third codon position. *Mytilus* showed the highest degree of non-stationarity compared to the *Crassostrea* and *Aplysia* outgroups. These possible violations of MP assumptions make the *Mytilus* topology less reliable than the *Crassostrea* and *Aplysia* topologies.

When MP and ML were used to analyse more divergent actin genes, such as those from molluscs, the results showed less agreement. In this situation the ML results were favoured, as the MP results were probably erroneous due to the large divergence times between actin genes and the violation of MP assumptions.

The results of phylogenetic analyses suggest that there are orthologous genes within the haliotids. A phylogenetic relationship between the actin genes of haliotids and other mollusc genera was evident, but confident conclusions cannot be made due to the effect of species divergence on the robustness of phylogenetic inference and the paucity of actin gene family data for a range of representative genera. As the species analysed became more divergent the ML method gave the most reliable results.

Chapter 5:

CONCLUSIONS



H. iris actin genes were amplified using PCR, with *H. iris* actin-subtype specific PCR primers and *H. rufescens* (Californian red abalone) or *Cyprinus carpio* (common carp) derived PCR primers. This approach detected three full length actin coding sequences and three partial actin coding sequences.

The *H. iris* actin genes showed high sequence similarity to previously isolated genes deposited in GenBank (82-95% similarity to molluscs, 82-91% similarity to echinoderms, 81-92% similarity to insects). *H. iris* actin genes contained one intron located between codons 41 and 42. The intron length ranged between genes, the intron sizes were: 174 nt, *H.irisA1*; 1,078 nt, *H.irisA2*; 581 nt, *H.irisA3*; 301 nt, *H.irisA1a*; 282 nt, *H.irisA1* and 229 nt, *H.irisA1c*.

Southern hybridisation detected at least eight actin-related DNA fragments, indicating the presence of a large gene family in *H. iris*. The number of actin genes identified in *H. iris* agrees with estimates of gene family size in other molluscs (Van Loon *et al.*, 1993; Patwary, 1996)

Differences between the full *H.irisA1* sequence determined during this research, and the partial *H.irisA1* sequence determined by Flint (2000), suggested these sequences may be from different genes. The ambiguity between the two *H.irisA1* sequences may represent the presence of a seventh *H. iris* actin gene.

The predicted *H. iris* actin proteins lacked the second amino acid usually found in invertebrate actin proteins. *H. rufescens* actin and *Patella* actins also lack this amino acid (Van Loon *et al.*, 1993; Gomez-Chiarri *et al.*, 1999). The second residue occurs within the acidic amino-terminus of the actin protein, a region involved in the binding of myosin. The absence of the second amino acid in *H. iris* may reflect co-evolution with *H. iris* myosin. Analysis of *H. iris* myosin may clarify the functional significance of this missing amino acid.

Analysis of gene expression by RT-PCR showed the three full length actin genes were functional. These genes were differentially expressed during development and in adult tissues. Actin gene expression in fertilised egg, blastula, trochophore larval and veliger larval stages, and in adult gill, gonad, hepatopancreas, foot and mantle tissues was tested. *H.irisA1* and *H.irisA2* were

expressed at low levels in fertilised eggs and blastula, with expression increasing in trochophore and veliger larvae. *H.irisA3* was not expressed in eggs, but was faintly detected in blastula and highly expressed in trochophore and veliger larvae. *H.irisA1* was ubiquitously expressed in adult gill, gonad, hepatopancreas, foot and mantle tissue, suggesting it may be a cytoplasmic-type actin. *H.irisA2* was expressed in all tissues except the hepatopancreas, although low expression may not have been detectable by electrophoresis of RT-PCR products, suggesting it may be a cytoplasmic-type actin. *H.irisA3* was expressed in gonad, foot and mantle, suggesting it may be a muscle-type actin.

The expression patterns of *H.irisA1*, *H.irisA2* and *H.irisA3* suggest that cytoplasmic and muscle-type actin genes exist in molluscs. The presence of a conserved amino acid, leucine 177, in *H.irisA3* as well as in previously described mollusc muscle actins, may indicate mollusc muscle and cytoplasmic-type actins display characteristic sequence differences, as found in insects and vertebrates. The potential existence of muscle-characteristic residues in mollusc muscle-type actins may have been previously overlooked due to the paucity of gene sequences for analysis, and the small number of proposed muscle-characteristic amino acids compared to the number found in insects (7-9 amino acids) and vertebrates (17-26 amino acids). If muscle-characteristic amino acids are found to be present in molluscs, this would support the functional diversification of mollusc muscle-actins.

Co-expression of *H.irisA1* and *H.irisA2* during development and in adult tissues may demonstrate the need for high levels of the actin protein in *H. iris*. The presence of differential gene expression of *H. iris* actin genes may suggest the functional specialisation of actin in *H. iris*. These findings may explain the presence of large actin gene families in molluscs.

Investigation of haliotid actin nucleotide composition and codon biases found evidence of selection for synonymous codons in haliotid actin genes. This selection could act to increase translational efficiency of highly expressed genes, such as actin. There was no evidence for positive selection or gene conversion in the evolution of *H. iris* actin genes.

Maximum parsimony and maximum likelihood phylogenetic analyses of the *H. iris* actin genes described in this study were performed with molluscan actin genes available on GenBank. Analyses suggested that haliotid actins can be divided into two orthologous clades, the first clade containing *H.irisA1, H.irisA1a, H.irisA1b, H.irisA1c, H.virgA1a, H.virgA1b, H.virgA1c* and *H. rufescens actin,* the second clade containing *H.irisA2, H.irisA3* and *H. discus hannai actin.* The presence of orthology indicates that the last common ancestor of haliotids had at least two actin genes. The clustering of actin genes within individual haliotid species suggests paralogy, resulting from duplication of actin genes within species. Evidence for gene orthology within mollusc actin genes was also found; however, further characterisation of actin genes from more mollusc species is required to infer the evolutionary significance of this orthology.

Maximum parsimony and maximum likelihood methods gave similar results for haliotid actin genes. The agreement of these methods may be due to the assumptions of maximum parsimony being met in the haliotid data set. Comparison of more divergent molluscs groups within this study may have been hindered by the presence of non-stationarity and among site rate variation between taxa.

The gene structure and expression studies cannot be used to support the results of the haliotid phylogenetic analyses because the conserved haliotid actin gene structure described in this study is phylogenetically uninformative; additionally, the lack of expression data for haliotid actin genes available in GenBank prevents valid conclusions from being drawn.

Gene isolation and identification by PCR and Southern hybridisation revealed at least eight actin genes in *H. iris,* of which three full length sequences and three partial sequences were determined. RT-PCR analysis of the full length genes, and codon usage analysis of the partial genes, suggested a large number of functional *H. iris* actin genes. Expression patterns suggested cytoplasmic and muscle-type actins exist in *H. iris.* Phylogenetic analyses of *H. iris* and molluscan actin genes detected the presence of orthologous relationships between haliotid actins, indicating that multiple actin genes were present in the

haliotid common ancestor and paralogous relationships within haliotid actins, indicating gene family expansion by duplication. Further analysis is required to elucidate the significance of orthology; however, expansion of the actin gene family may reflect the need for high levels of actin protein in *H. iris* suggested by the presence of a large muscular foot.

ACKNOWLEDGEMENTS

I thank my supervisors, Associate Professor Frank Sin and Dr Neil Gemmell, for their guidance and support. Frank, I thank you for your trust and patience in allowing me to pursue my own research interests. I also thank you for not just caring about the science, but also caring about me. Neil, I thank you for your wide-ranging enthusiasm for science (and willingness to share it), your belief in my abilities and for being a friend. I also thank Dr Jenny Khoo whose laboratory practice and standards rubbed off on me, for which I'm very grateful for.

I acknowledge Sharyn Goldstien, who collected *H. virginea* sperm for me and assisted and reassured me when it came to tackling phylogenetics. I would also like to acknowledge Alice Johnstone who provided me with total RNA from adult *H. iris* tissues. I thank the departmental secretaries and technicians for their assistance. I would particularly like to thank John Scott, Linda Morris, Nick Etheridge, Bruce Lingard and Jack van Berkel for technical support and for being pleasant and helpful people to work with. I thank Pacifica Fisheries for providing larval samples.

I acknowledge the Foundation for Research Science and Technology and the University of Canterbury for financial support during parts of this research.

I would like to mention my various lab colleagues over the years, who made a research lab a not so bad place to be. I would particularly like to mention Heather, Kath, Anna, Mel, Margee, Andrew, Tans, Vic, Ermin and Julia for their friendship, support and fun-ness.

And lastly, but obviously not least, I would like to thank my family and friends. To my Mum and Dad, I know you would've been proud and I'm sorry you couldn't see the end with me. To Karl, Bean and Jess, thank you for making me a home and something to look forward to at the end of the day.

REFERENCES

Abouheif, E., Akam, M., Dickinson, W.J., Holland, P.W., Meyer, A., Patel, N.H., Raff, R.A., Roth, V.L. and Wray, G.A. (1997). Homology and developmental genes. *Trends in Genetics* 13: 432-433.

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft. C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M, Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C. (2000). The genome sequence of *Drosophila melanogaster*. *Nature* 287: 2185-2195.

Adema, C.M. (2002). Comparative study of cytoplasmic actin DNA sequences from six species of Planorbidae (Gastropoda: Basommatophora). *Journal of Molluscan Studies* 68: 17-23.

Akashi, H. (1994). Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136: 927-935.

Akashi, H. (1997). Distinguishing the effects of mutational biases and natural selection on DNA sequence variation. *Genetics* 147: 1989-1991.

Akhurst, R.J., Calzone, F.J., Lee, J.J., Britten, R.J. and Davidson, E.H. (1987). Structure and organization of the *CyIII* actin gene subfamily of the sea urchin, *Strongylocentrotus purpuratus*. *Journal of Molecular Biology* 194: 193-203.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.

Arnone, M.I., Martin, E.L. and Davidson, E.H. (1998). *Cis*-regulation downstream of cell type specification: a single compact element controls the complex expression of the *CyIIa* gene in sea urchin embryos. *Development* 125: 1381-1395.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., G., S.J., Smith, J.A. and Struhl, K. (1997). *Short Protocols in Molecular Biology*. John Wiley & Sons, Inc. New York.

Bray, D. (1973). Cytoplasmic actin: A comparative study. *Cold Spring Harbor Symposia of Quantitative Biology* 37: 567-571.

Cadoret, J.-P., Debon, R., Cornudella, L., Lardans, V., Morvan, A., Roch, P. and Boulo, V. (1999). Transient expression assays with the proximal promoter of a newly characterized actin gene from the oyster *Crassostrea gigas*. *FEBS Letters* 460: 81-85.

Cao, L.G., Fishkind, D.J. and Wang, Y.L. (1993). Localization and dynamics of nonfilamentous actin in cultured cells. *Journal of Cell Biology* 123: 173-181.

Carlier, M.F. (1990). Actin polymerization and ATP hydrolysis. *Advances in Biophysics* 26: 51-73.

Carlier, M.F. and Pantaloni, D. (1997). Control of actin dynamics in cell motility. *Journal of Molecular Biology* 269: 459-467.

Carlini, D.B., Reece, K.S. and Graves, J.E. (2000). Actin gene family evolution and the phylogeny of coleoid cephalopods (*Mollusca: Cephalopoda*). *Molecular Biology and Evolution* 17: 1353-1370.

Coleman, A.W. and Vacquier, V.D. (2002). Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (*Haliotis*). *Journal of Molecular Evolution* 54: 246-257.

Condeelis, J. (1993). Life at the leading edge: the formation of cell protrusions. *Annual Review of Cellular Biology* 9: 411-444.

Cooper, A.D. and Crain, W.R., Jr. (1982). Complete nucleotide sequence of a sea urchin actin gene. *Nucleic Acids Research* 10: 4081-4092.

Cooper, J.A. (1991). The role of actin polymerization in cell motility. *Annual Review of Physiology* 53: 585-605.

Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 247: 775-780.

Cox, K.H., Angerer, L.M., Lee, J.J., Davidson, E.H. and Angerer, R.C. (1986). Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *Journal of Molecular Biology* 188: 159-172.

Crain, W.R., Jr., Boshar, M.F., Cooper, A.D., Durica, D.S., Nagy, A. and Steffen, D. (1987). The sequence of a sea urchin muscle actin gene suggests a gene conversion with a cytoskeletal actin gene. *Journal of Molecular Evolution* 25: 37-45.

Cupples, C.G. and Pearlman, R.E. (1986). Isolation and characterization of the actin gene from *Tetrahymena thermophila*. *Proceedings of the National Academy of Sciences of the United States of America* 83: 5160-5164.

Desgroseillers, L., Auclair, D. and Wickham, L. (1990). Nucleotide sequence of an actin cDNA gene from *Aplysia californica*. *Nucleic Acids Research* 18: 3654.

Desgroseillers, L., Auclair, D., Wickham, L. and Maalouf, M. (1994). A novel actin cDNA is expressed in the neurons of *Aplysia californica*. *Biochimica et Biophysica Acta* 1217: 322-324.

Dibb, N.J. and Newman, A.J. (1989). Evidence that introns arose at proto-splice sites. *The EMBO Journal* 8(7): 2015-2021.

Doolittle, R.F. (1989). Similar amino acid sequences revisited. *Trends in Biochemical Science* 14: 244-245.

Drouin, G. (2002). Testing claims of gene conversion between multigene family members: examples from echinoderm actin genes. *Journal of Molecular Evolution* 54: 138-139.

Durica, D.S., Garza, D., Restrepo, M.A. and Hryniewicz, M.M. (1988). DNA sequence analysis and structural relationships among the cytoskeletal actin genes of the sea urchin *Strongylocentrotus purpuratus*. *Journal of Molecular Evolution* 28: 72-86.

Durica, D.S., Schloss, J.A. and Crain, W.R.J. (1980). Organization of actin gene sequences in the sea urchin: Molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. *Proceedings of the National Academy of Sciences of the United States of America* 77: 5683-5687.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'connell, C., Spritz, R.A., Deriel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980). The structure and evolution of the human beta-globin gene family. *Cell* 21: 653-668.

Elder, J.F.J. and Turner, B.J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology* 70: 297-320.

Fleming, A.E. and Hone, P.W. (1996). Abalone aquaculture - Introduction. *Aquaculture* 140: 1-4.

Flescher, E., Madden, K. and Snyder, M. (1993). Components required for cytokinesis are important for bud site selection in yeast. *Journal of Cell Biology* 122: 373-386.

Flint, H.J. (2000). *A Molecular Investigation of Actin Genes in New Zealand's Black-Footed Abalone (Haliotis iris)*. MSc Thesis. Department of Zoology, University of Canterbury.

Foran, D.R., Johnson, P.J. and Moore, G.P. (1985). Evolution of two actin genes in the sea urchin *Strongylocentrotus franciscanus*. *Journal of Molecular Evolution* 22: 108-116.

Fryxell, K.J. (1996). The coevolution of gene family trees. *Trends in Genetics* 12: 364-369.

Fukui, Y. (1993). Toward a new concept of cell motility: cytoskeletal dynamics in amoeboid movement and cell division. *International Review of Cytology* 144: 85-127.

Fyrberg, C., Mahaffey, J.W., Bond, B.J. and Davidson, N. (1983). Transcripts of the six *Drosophila* actin genes accumulate in a stage- and tissue-specific manner. *Cell* 33: 115-123.

Fyrberg, E.A. (1984). Structural and functional analyses of *Drosophila melanogaster* actin genes. *Oxford Surveys on Eukaryotic Genes* 1: 61-86.

Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981). The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. *Cell* 24: 107-116.

Fyrberg, E.A., Kindle, K.L. and Davidson, N. (1980). The actin genes of *Drosophila*: a dispersed multigene family. *Cell* 19: 365-378.

Geiger, D.L. (1998). Recent genera and species of the family Haliotidae Rafinesque, 1815 (Gastropoda : Vetigastropoda). *Nautilus* 111: 85-116.

Gerdes, K., Moller-Jensen, J. and Bugge Jensen, R. (2000). Plasmid and chromosome partitioning: surprises from phylogeny. *Molecular Microbiology* 37: 455-466.

Giribet, G. and Wheeler, W.C. (2002). On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. *Invertebrate Biology* 121: 271-324.

Gomez-Chiarri, M., Kirby, V.L. and Powers, D.A. (1999). Isolation and characterization of an actin promoter from the red abalone (*Haliotis rufescens*). *Marine Biotechnology* 1: 269-278.

Gonnet, G.H., Cohen, M.A., Benner, S.A., Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1992). Exhaustive matching of the entire protein sequence database. *Science* 256: 1443-1445.

Gordon, D.J., Boyer, J.L. and Korn, E.D. (1977). Comparative biochemistry of non-muscle actins. *Journal of Biological Chemistry* 252: 8300-8309.

Gu, X. and Li, W.H. (1998). Estimation of evolutionary distances under stationary and nonstationary models of nucleotide substitution. *Proceedings of the National Academy of Sciences of the United States of America* 95: 5899-5905.

Hadden, T.J. and Sodja, A. (1994). An oligogene family encodes actins in the housefly, *Musca domestica*. *Biochemistry and Biophysics Research Communications* 203: 523-531.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Ser.* 41: 95-98.

Haymer, D.S., Anleitner, J.E., He, M., Thanaphum, S., Saul, S.H., Ivy, J., Houtchens, K. and Arcangeli, L. (1990). Actin genes in the Mediterranean fruit fly, *Ceratitis capitata*. *Genetics* 125: 155-160.

He, M. and Haymer, D.S. (1991). Isolation and characterization of actin homologous sequences from *Dacus dorsalis* (Hendel) (Diptera: Tephritidae). *Annals of the Entomological Society of America* 84: 601-607.

He, M. and Haymer, D.S. (1994). The actin gene family in the oriental fruit fly *Bactrocera dorsalis* muscle specific actins. *Insect Biochemistry and Molecular Biology* 24: 891-906.

He, M. and Haymer, D.S. (1995). Codon bias in actin multigene families and effects on the reconstruction of phylogenetic relationships. *Journal of Molecular Evolution* 41: 141-149.

Hightower, R.C. and Meagher, R.B. (1986). The molecular evolution of actin. *Genetics* 114: 315-332.

Hirono, M., Endoh, H., Okada, N., Numata, O. and Watanabe, Y. (1987). Tetrahymena actin. Cloning and sequencing of the Tetrahymena actin gene and identification of its gene product. *Journal of Molecular Biology* 194: 181-192.

Holder, M. and Lewis, P.O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *National Reviews Genetics* 4: 275-284.

Ikemura, T. and Ozeki, H. (1982). Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. *Cold Spring Harbor Symposium Quantitative Biology* 47: 1087-1096.

Johnson, P.J., Foran, D.R. and Moore, G.P. (1983). Organization and evolution of the actin gene family in sea urchins. *Molecular Cell Biology* 3: 1824-1833.

Kabsch, W. and Vandekerckhove, J. (1992). Structure and function of actin. *Annual Review of Biophyics and Biomolecular Structure* 21: 49-76.

Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 238(1):143-155.

Kirchhamer, C.V. and Davidson, E.H. (1996). Spatial and temporal information processing in the sea urchin embryo: modular and intramodular organization of the *CyIIIa* gene *cis*-regulatory system. *Development* 122: 333-348.

Kirchhamer, C.V., Yuh, C.H. and Davidson, E.H. (1996). Modular *cis*-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proceedings of the National Academy of Sciences of the United States of America* 93: 9322-9328.

Kissinger, J.C., Hahn, J.H. and Raff, R.A. (1997). Rapid evolution in a conserved gene family. Evolution of the actin gene family in the sea urchin genus *Heliocidaris* and related genera. *Molecular Biology and Evolution* 14: 654-665.

Korn, E.D. (1982). Actin polymerization and its regulation by proteins from nonmuscle cells. *Physiology Reviews* 62: 672-737.

Kornegay, J., Schilling, J. and Wilson, A. (1994). Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. *Molecular Biology and Evolution* 11: 921-928.

Kowbel, D.J. and Smith, M.J. (1989). The genomic nucleotide sequences of two differentially expressed actin-coding genes from the sea star *Pisaster ochraceus*. *Gene* 77: 297-308.

Kumar, S. and Gadagkar, S.R. (2001). Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics* 158: 1321-1327.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Lake, J.A. (1994). Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proceedings of the National Academy of Sciences of the United States of America* 91: 1455-1459.

Lardans, V., Ringaut, V., Duclermortier, P., Cadoret, J.P. and Dissous, C. (1997). Nucleotide and deduced amino acid sequences of *Biomphalaria glabrata* actin cDNA. *DNA Sequence* 7: 353-356.

Lee, J.J., Shott, R.J., Rose, S.J.I., Thomas, T.L., Britten, R.J. and Davidson, E.H. (1984). Sea urchin actin gene subtypes. Gene number, linkage and evolution. *Journal of Molecular Biology* 172: 149-176.

Lee, Y.-H., Ota, T. and Vacquier, V.D. (1995). Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Molecular Biology and Evolution* 12: 231-238.

Lee, Y.-H. and Vacquier, V.D. (1992). The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. *Biological Bulletin* 182: 97-104.

Lee, Y.-H. and Vacquier, V.D. (1995). Evolution and systematics in Haliotidae (Mollusca: Gastropoda): inferences from DNA sequences of sperm lysin. *Marine Biology* 124: 267-278.

Lespinet, O., Nederbragt, A.J., Cassan, M., Dictus, W.J., Van Loon, A.E. and Adoutte, A. (2002). Characterisation of two snail genes in the gastropod mollusc *Patella vulgata*. Implications for understanding the ancestral function of the snail-related genes in Bilateria. *Developmental Genetics and Evolution* 212: 186-195.

Li, W.-H. (1997). *Molecular Evolution*. Sinauer Associates Ltd. Sunderland, Massachusetts.

Li, W.-H., Luo, C.-C. and Wu, C.-I. (1985). Evolution of DNA Sequences. In: *Molecular Evolutionary Genetics*. Mac Intyre, R.J. (Ed.). Plenum Press. New York.

Lindberg, D.R. (1992). Evolution, distribution and systematics of Haliotidae. In: *Abalone of the World. Biology, Fisheries and Culture*. Shepherd, S.A., Tegner, M.J. and Guzman Del Proo, S.A (Eds.). Fishing News Books, Blackwll Scienfic Publications Ltd. Cambridge.

Lockhart, P.J., Steel, M., Hendy, M.D. and Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution* 11: 605-612.

Mange, A., Couble, P. and Prudhomme, J.C. (1996). Two alternative promoters drive the expression of the cytoplasmic actin A4 gene of *Bombyx mori*. *Gene* 183: 191-199.

Metz, E.C., Robles-Sikisaka, R. and Vacquier, V.D. (1998). Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 95: 10676-10681.

Minor, J.E., Lee, J.J., Akhurst, R.J., Leahy, P.S., Britten, R.J. and Davidson, E.H. (1987). Sea urchin actin gene linkages determined by genetic segregation. *Developmental Biology* 122: 291-295.

Miyamoto, H., Hamaguchi, M. and Okoshi, K. (2002). Analysis of genes expressed in the mantle of oyster *Crassostrea gigas*. *Fisheries Science* 68: 651-658.

Morse, D.E. (1984). Biochemical and genetic engineering for improved production of abalones and other valuable molluscs. *Aquaculture* 39: 263-282.

Mounier, N., Coulon, M. and Prudhomme, J.C. (1991). Expression of a cytoplasmic actin gene in relation to the silk production cycle in the silk glands of *Bombyx mori*. *Insect Biochemistry* 21: 293-301.

Mounier, N., Gouy, M., Mouchiroud, D. and Prudhomme, J.C. (1992). Insect muscle actins differ distinctly from invertebrate and vertebrate cytoplasmic actins. *Journal of Molecular Evolution* 34: 406-415.

Mounier, N. and Prudhomme, J.C. (1986). Isolation of actin genes in *Bombyx mori*: the coding sequence of a cytoplasmic actin gene expressed in the silk gland is interrupted by a single intron in an unusual position. *Biochimie* 68: 1053-1061.

Mounier, N. and Prudhomme, J.C. (1991). Differential expression of muscle and cytoplasmic actin genes during development of *Bombyx mori*. *Insect Biochemistry* 21: 523-533.

Mounier, N. and Sparrow, J.C. (1997). Structural comparisons of muscle and nonmuscle actins give insights into the evolution of their functional differences. *Journal of Molecular Evolution* 44: 89-97.

Nederbragt, A.J., Lespinet, O., Van Wageningen, S., Van Loon, A.E., Adoutte, A. and Dictus, W.J. (2002a). A lophotrochozoan twist gene is expressed in the ectomesoderm of the gastropod mollusk Patella vulgata. *Evolutionary Development* 4: 334-343.

Nederbragt, A.J., Van Loon, A.E. and Dictus, W.J. (2002b). Expression of Patella vulgata orthologs of engrailed and dpp-BMP2/4 in adjacent domains

during molluscan shell development suggests a conserved compartment boundary mechanism. *Developmental Biology* 246: 341-355.

Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annual Review of Genetics* 30: 371-403.

Nei, M. and Kumar, C. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, Inc. New York, New York.

Nielsen, R. and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929-936.

Nyberg, A.M. and Cronhjort, M.B. (1992). Intron evolution: A statistical comparison of two models. *Journal of Theoretical Biology* 157: 175-190.

O'Brien, E.K., Degnan, B. M. (2000). Expression of *POU*, *Sox*, and *Pax* genes in the brain ganglia of the tropical abalone *Haliotis asinina*. *Marine Biotechnology* 2: 545-557.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer-Verlag. New York.

Owen, B., Mclean, J.H. and Meyer, R.J. (1971). Hybridization in the eastern Pacific abalones (*Haliotis*). *Bulletin of the Los Angeles County Museum of Natural History and Science* 9: 1-37.

Patwary, M.U., Reith, M. and Kenchington, E.L. (1996). Isolation and characterization of a cDNA encoding an actin gene from sea scallop (*Placopecten magellanicus*). *Journal of Shellfish Research* 15: 265-270.

Pollard, T.D. (1976). The role of actin in the temperature-dependent gelation and contraction of extracts of Acanthamoeba. *Journal of Cell Biology* 68: 579-601.

Pollard, T.D. (1990). Actin. *Current Opinion in Cell Biology* 2: 33-40.

Pollard, T.D., Almo, S., Quirk, S., Vinson, V. and Lattman, E.E. (1994). Structure of actin binding proteins: insights about function at atomic resolution. *Annual Review of Cellular Biology* 10: 207-249.

Pollard, T.D. and Cooper, J.A. (1986). Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. *Annual Review of Biochemistry* 55: 987-1035.

Pollard, T.D. and Weihing, R.R. (1974). Actin and myosin and cell movement. *CRC Critical Reviews in Biochemistry* 2: 1-65.

Poore, G.C.B. (1969). The Ecology of the New Zealand *Haliotis* Species (Mollusca). PhD Thesis. Department of Zoology, University of Canterbury.

Poore, G.C.B. (1972a). Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda). 1. Feeding. *New Zealand Journal of Marine and Freshwater Research* 6: 11-22.

Poore, G.C.B. (1972b). Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda). 2. Seasonal and Diurnal Movement. *New Zealand Journal of Marine and Freshwater Research* 6: 245-258.

Poore, G.C.B. (1972c). Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda). 3. Growth. *New Zealand Journal of Marine and Freshwater Research* 6: 534-559.

Poore, G.C.B. (1973). Ecology of New Zealand abalones, *Haliotis* species (Mollusca: Gastropoda). 4. Reproduction. *New Zealand Journal of Marine and Freshwater Research* 7: 67-84.

Posada, D. and Crandall, K. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Powell, J.R. and Moriyama, E.N. (1997). Evolution of codon usage bias in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 94: 7784-7790.

Rayment, I., Holden, H.M., Whittaker, M., Yohn, C.B., Lorenz, M., Holmes, K.C. and Milligan, R.A. (1993). Structure of the actin-myosin complex and its implications for muscle contraction. *Science* 261: 58-65.

Read, E.B., Okamura, H.H. and Drubin, D.G. (1992). Actin- and tubulin-dependent functions during *Saccharomyces cerevisiae* mating projection formation. *Molecular Biology of the Cell* 3: 429-444.

Reece, K.S. and Stokes, N.A. (2003). Molecular analysis of a haplosporidian parasite from cultured New Zealand abalone *Haliotis iris*. *Diseases in Aquatic Organisms* 53: 61-66.

Reisler, E. (1993). Actin molecular structure and function. *Current Opinion in Cell Biology* 5: 41-47.

Rubenstein, P.A. (1990). The functional importance of multiple actin isoforms. *Bioessays* 12: 309-315.

Ruppert, E.E. and Barnes, R.D. (1994). Invertebrate Zoology (6th Edition). Saunders College Publishing. USA.

Rzhetsky, A. and Nei, M. (1995). Tests of applicability of several substitution models for DNA sequence data. *Molecular Biology and Evolution* 12: 131-151.

Sainsbury, K.J. (1982). Population dynamics and fishery management of the paua, *Haliotis iris*. I. Population structure, growth, reproduction, and mortality. *New Zealand Journal of Marine and Freshwater Research* 16: 147-161.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press.

Sawyer, S. (1989). Statistical tests for detecting gene conversion. *Molecular Biologoy and Evolution* 6: 526-538.

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

Scheller, R.H., Mcallister, L.B., Crain, W.R.J., Durica, D.S., Posakony, J.W., Thomas, T.L., Britten, R.J. and Davidson, E.H. (1981). Organization and expression of multiple actin genes in the sea urchin. *Molecular and Cellular Biology* 1: 609-628.

Schiel, D.R. (1992). The paua (abalone) fishery of New Zealand. In: *Abalone of the World. Biology, Fisheries and Culture.* Shepherd, S.A., Tegner, M.J. and Guzman Del Proo, S.A (Eds.). Fishing News Books, Blackwll Scienfic Publications Ltd. Cambridge.

Schiel, D.R. (1993). Experimental evaluation of commercial-scale enhancement of abalone *Haliotis iris* populations in New Zealand. *Marine Ecology Progress Series* 97: 167-181.

Schiel, D.R. (1997). Review of abalone culture and research in New Zealand. *Molluscan Research* 18: 289-298.

Schiel, D.R. and Breen, P.A. (1991). Population structure, ageing, and fishing mortality of the New Zealand abalone *Haliotis iris*. *Fishery Bulletin, U. S.* 89: 681-691.

Sharp, P.M. and Li, W.-H. (1987). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Molecular Biology and Evolution* 4:222-230.

Sharp, P.M. and Li, W.-H. (1988). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Research* 14: 7737-7749.

Sheterline, P. and Sparrow, J.C. (1994). Actin. *Protein Profile* 1: 1-121.

Shields, D.C. and Sharp, P.M. (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Research* 15: 8023-8040.

Shields, D.C., Sharp, P.M., Higgins, D.G. and Wright, F. (1988). "Silent" sites in *Drosophila* genes are not neutral: Evidence of selection among synonymous codons. *Molecular Biology and Evolution* 5: 704-716.

Shott, R.J., Lee, J.J., Britten, R.J. and Davidson, E.H. (1984). Differential expression of the actin gene family of *Strongylocentrotus purpuratus*. *Developmental Biology* 101: 295-306.

Steel, M., Huson, D. and Lockhart, P.J. (2000). Invariable Sites Models and Their use in Phylogeny Reconstruction. *Systematic Biology* 49: 225-232.

Stiller, J. and Hall, B. (1999). Long-Branch Attraction and the rDNA Model of Early Eukaryotic Evolution. *Molecular Biology and Evolution* 16: 1270-1279.

Swanson, W.J., Aquadro, C.F. and Vacquier, V.D. (2001). Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive darwinian selection of sperm lysin. *Molecular Biology and Evolution* 18: 376-383.

Swanson, W.J. and Vacquier, V.D. (1998). Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* 281: 710-712.

Swofford, D. L. (1999). *PAUP 4.65*. Sinaur. Sunderland, Massuchusetts.

Sueoka, N. (1962). On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the National Academy of Sciences of the United States of America* 48: 582-592.

Tessmar-Raible, K. and Arendt, D. (2003). Emerging systems: between vertebrates and arthropods, the Lophotrochozoa. *Current Opinion in Genetics & Development* 13: 331-340.

Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F. and Higgins, D. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

Thornton, J.W. and Desalle, R. (2000). Gene family evolution and homology: Genomics meets phylogenetics. *Annual Review of Genomics and Human Genetics* 1: 41-73.

Tong, L.J., Moss, G.A., Redfearn, P. and Illingworth, J. (1992). *A manual of techniques for culturing paua, Haliotis iris, through to the early juvenile stage.* New Zealand Fisheries Technical Report: 31.

Van Loon, A., Goedemans, H., Daemen, A., Van De Kamp, A. and Van Den Biggelaar, J.A.M. (1993). Actin genes expressed during early development of *Patella vulgata. Roux's Archives of Developmental Biology* 202: 77-84.

Vandekerckhove, J. and Weber, K. (1978). Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins. *Proceedings of the National Academy of Sciences of the United States of America* 75: 1106-1110.

Weber, K. and Kabsch, W. (1994). Intron positions in actin genes seem unrelated to the secondary structure of the protein. *The EMBO Journal* 13(6): 1280-1286.

White, M. and Crother, B. (1999). Diagnostic amino acids in actin genes: an idea whose time has gone. *Molecular Biology and Evolution* 16: 876-879.

White, M.E. and Crother, B.I. (2000). Gene conversions may obscure actin gene family relationships. *Journal of Molecular Evolution* 50: 170-174.

Wolfe, K.H., Sharp, P.M. and Li, W.-H. (1989). Mutation rates differ among regions of the mammalian genome. *Nature* 337: 283-285.

Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution* 11: 367-372.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 13: 555-556. (http://abacus.gene.ucl.ac.uk/software/paml.html).

Yang, Z. and Kumar, S. (1996). Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Molecular Biologoy and Evolution* 13: 650-659.

Yang, Z. and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biologoy and Evolution* 17: 32-43.

Yang, Z., Swanson, W.J. and Vacquier, V.D. (2000). Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution* 17: 1446-1455.

Yokoyama, S. and Radlwimmer, F.B. (1999). The molecular genetics of red and green color vision in mammals. *Genetics* 153: 919-932.

Yoon, S.H. and Kim, W. (2000). Phylogeny of some gastropod mollusks derived from 18S rDNA sequences with emphasis on the Euthyneura. *Nautilus* 114: 84-92.

Zulauf, E., Sanchez, F., Tobin, S.L., Rdest, U. and Mccarthy, B.J. (1981). Developmental expression of a *Drosophila* actin gene encoding actin I. *Nature* 292: 556-558.

APPENDIX A - REAGENTS, BUFFERS & GELS

General Solution	Recipe
Extraction Buffer	0.1 M Tris-HCl pH 8.0 0.1 M NaCl 20 mM Na_2EDTA
Proteinase K 10% SDS	20mg/ml in ddH_2O Dissolved in ddH_2O and adjusted To pH 7.2 with HCl
3 M Sodium Chloride pH 5.2	Dissolved in ddH_2O and adjusted To pH 5.2 with HCl
Sodium Chloride/Sodium Citrate Buffer (SSC)	150 mM NaCl 15 mM sodium citrate dissolved in ddH_2O and adjusted to pH 7.0 with NaOH
3 M Sodium Acetate pH 5.2	Dissolved in ddH_2O and adjusted to pH 5.2 with glacial acetic acid
6 M Sodium Iodide	6 M NaI 0.1 M Na_2SO_3 Filtered through Whatman paper
Tris-HCl/EDTA Buffer pH8 (TE8)	10 mM Tris-HCl pH 8 1 mM EDTA pH 8
Tris-Borate Electrophoresis Buffer (TBE)	0.09 M Tris Base 0.09 M Boric Acid 0.01 M EDTA

Southern Hybridisation Solution	Recipe
Denaturation Solution	1.5 M NaCl 0.5 M NaOH
100 × Denhardt's Solution	2 g Ficoll (Type 400, Pharmacia) 2 g polyvinylpyrrolidone 2 g BSA (Fraction V, Sigma)
Depurination Solution	0.125 M HCl
Prehybridisation Solution	6 × SSC 5 × Denhardt's solution 0.5% SDS 0.1% blocking agent (Roche)

Autoradiography Solution	Recipe
Developer	Agfa G150 $^1/_6$ dilution
Stop Solution	Water with a splash of glacial acetic acid
Fixer	Ilford Hypam Fixer $^1/_5$ dilution

Gels	Recipe
1% Agarose	0.35g agarose powder 35 ml 0.5 × TBE
4 % PAGE Gels	1.325 ml 30% acrylamide: bisacrylamide (29:1) 1.0 ml 10 × TBE 7.625 ml ddH_2O Degas for 10-15 min Add: 100 µl 10% ammonium persulfate (APS) 3.5 µl TEMED
6% Sequencing PAGE Gels	35.0 g urea 10.5 ml 40% acrylamide: bisacrylamide (19:1) 7.0 ml 10 × TBE Heat and stir to dissolve. Make up to volume of 70 ml with ddH_2O. Degas for 10-15 min. Add: 700 µl 10% APS 14 µl TEMED

APPENDIX B -SUPPLIERS

Consumable	Supplier
Agarose	Pure Sciences
Agfa G150	Agfa
BioMax MR film (Kodak)	Radiographic Supplies
Chloroform	BDH
Cla I	Amersham Biosciences (USB)
dNTPs (Eppendorf)	Eppendorf (Global Science)
*Eco*R I	Amersham Biosciences (USB)
Ethanol	BDH
Ethidium Bromide	Sigma
Na_2EDTA	Pure Sciences
Glacial Acetic Acid	BDH
Hae III	Amersham Biosciences (USB)
Hind III	Amersham Biosciences (USB)
Hybond XL	Amersham Biosciences
Hydrochloric Acid	BDH
Ilford Hypam Fixer	H E Perry
Ilford Hypam Hardener	H E Perry
Isoamyl Alcohol	BDH
Isopropanol	BDH
Orthoboric Acid	Scientific Supplies Ltd
PCR Primers	Invitrogen
Phenol	BDH
Polynucleotide Kinase (USB)	Amersham Biosciences
Proteinase K	Roche Applied Sciences
Pst I	Amersham Biosciences (USB)
Pvu II	Amersham Biosciences (USB)
RNase A	Roche Applied Sciences
RNase H	Amersham Biosciences (USB)
NaOAc	BDH
NaCl	Pure Sciences
Sodium Citrate	BDH
SDS	Scientific Supplies Ltd
NaOH	Scientific Supplies Ltd
NaI	Scientific Supplies Ltd
Taq Polymerase	Roche Applied Sciences
Thermo Sequenase Cycle Sequencing Kit	USB (Amersham Biosciences)
Tris(hydroxymethyl)aminomethane (Tris)	Pure Sciences
α^{32}P-dCTP (NEN)	Biolab Scientific
γ^{33}P-dATP (NEN)	Biolab Scientific

Equipment	Supplier
Barnstead Nanopure Ultrapurification system	Medic
Eppendorf Centrifuge 5417R	Lab Supply Pierce
Eppendorf Mastercycler® Gradient	Global Science
Gene Power Supply 200/400 (Pharmacia)	Amersham Biosciences
HE33 Mini Horizontal Submarine tank (Hoeffer)	Amersham Biosciences
Hybaid Hybridisation Oven	SciTech
LiCor 4000	John Morris Scientific
LKB 2197 Power Supply	Amersham Biosciences
Model S2 Sequencing Gel Apparatus	LKB
UltraLum Gel Documentation	SciTech
Ultrospec II	LKB Biochrom Ltd
Wide Mini Sub Cell GT tank	BioRad

APPENDIX C – 18S rRNA ALIGNMENT

```
S. conf       ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CACAGTGAAACTGCGAATG  [71]
H. discus     TATCTGGTTGATCCTGCCAGTAGCCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CACAGTGAAACTGCGAATG  [94]
E. adan       ---------------------AGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACACGCCCTCGGCACGGCGAAACTGCGAATG  [74]
M. dilat      -----GGTTGATCCTGCCAGTAGT-ATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTCG-CCCAGTGAAACTGCGAATG  [88]
C. cruc       ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CACAGTGAAACTGCGAATG  [71]
P. midas      ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACACGCCCTCAGCACGGCGAAACTGCGAATG  [72]
H. iris       ---------------------AATCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CACAGTGAAACTGCGAATG  [73]
B. corn       TACCTGGTTGATCCTGCCAGTAATCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CACAGTGAAACTGCGAATG  [94]
T. gigas      -----GGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CCCAGTGAAACTGCGAATG  [89]
H. tuber      ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTAG-CACAGTGAAACTGCGAATG  [71]
E. adan2      ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACACGCCCTCGGCACGGCGAAACTGCGAATG  [72]
C. argy       -----GGTTGATCCTGCCAGTAATCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACTTACTCTAG-CACAGTGAAACTGCGAATG  [89]
D. graeca     ----------------------TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTACAAACTCTCG-CCCAGTGAAACTGCGAATG  [71]
                                     ****************************************** *** * * ******************

S. conf       GCTCATTAGATCAGTTATGGTTCCTTAGACGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTATAGCTCCGACC  [165]
H. discus     GCTCATTAGATCAGTTATGGTTCCTTAGATGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTAAAGCTCCGACC  [188]
E. adan       GCTCATTAGATCAGTTATGGTTCCTTAGACGGTACGATTCCTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCGACCCAGCTCCGACC  [169]
M. dilat      GCTCATTACATCAGTTATGGTTCCTTGGACGATACCAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTTCGGCTCCGACC  [182]
C. cruc       GCTCATTAGATCAGTTATGGTTCCTTAGATGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTATAGCTCCGACC  [165]
P. midas      GCTCATTAGATCAGTTATGGTTCCTTAGACGGTACGATTCCTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCGACCCAGCTCCGACC  [167]
H. iris       GCTCATTAGATCAGTTATGGTTCCTTAGATGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTAAAGCTCCGACC  [167]
B. corn       GCTCATTAGATCAGTTATGGTTCCTTAGATGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTATAGCTCCGACC  [188]
T. gigas      GCTCAGTACATCAGTTATGGTTTATTGGACGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACATCAGCTCCGACC  [183]
H. tuber      GCTCATTAGATCAGTTATGGTTCCTTAGATGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTAAAGCTCCGACC  [165]
E. adan2      GCTCATTAGATCAGTTATGGTTCCTTAGACGGTACGATTCCTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGCGACCCAGCTCCGACC  [167]
C. argy       GCTCATTAGATCAGTTATGGTTCCTTAGATGATACAAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTATAGCTCCGACC  [183]
D. graeca     GCTCATTACATCAGTTATGGTTCCTTGGACGATACCAT-CCTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACTTCGGCTCCGACC  [165]
              ***** ** *************  ** ** * *** ** ****************** ***********************************
```

```
S. conf      CTTTCGCGAGGGGAAGA-GCGCGTTTATCAG-TTCGAA----------------------------GCCAGCCGGG------------------  [211]
H. discus    CTTTCT-GAGGGGAAGAAGCGCGTTTATCAG-TTCGAA----------------------------GCCAGC-GAG------------------  [233]
E. adan      -TCTCCCGCAGGGAAGA-GCGCTTTTATTAG-TTTCAAAAAACCCGCACGCGGATCGGGCTTCACGGGCGGCGGGGGTCTTGGCTCGTCTGGGCT  [260]
M. dilat     CTTTGT-TTGGGGAAGA-GCGCATTTATCAGCTTCGAA----------------------------GCCAGCCGGA------------------  [228]
C. cruc      CTTCTG---GGGGAAGA-GCGCGTTTATCAG-TTCGAA----------------------------GCCAGCCGGG------------------  [208]
P. midas     -TCTCCCGCAGGGAAGA-GCGCTTTTATTAGTTTCAAAAAA-CCCGCACGCGGGCCGGGCTGCCGCGGCGGCGGGGGSTTCAGTGCGACTTGCCT  [259]
H. iris      CTCTCT-GAGGGGAAGA-GCGCGTTTATCAG-TTCGAA----------------------------GCCAGCCGAG------------------  [212]
B. corn      CTTTCGCGAGGGGAAGA-GCGCGTTTATCAG-TTCGAA----------------------------GCCAGCCGGG------------------  [234]
T. gigas     CTTTCCCGAGGGGAAGA-GCGCGTTTATCAGCTTCGAA----------------------------GCCAGCCGAG------------------  [230]
H. tuber     CTTTCT-GAGGGGAAGA-GCGCGTTTATCAG-TTCGAA----------------------------GCCAGCCGAG------------------  [210]
E. adan2     -TCTCCCGCAGGGAAGA-GCGCTTTTATTAGTTTCAAAAAA-CCCGCACGCGGATCGGGCTTCACGGGCGGCGGGGGTCTTGGCTCGTCTGGGCT  [259]
C. argy      CTTTCGCGAGGGGAAGA-GCGCGTTTATCAGCTTCGAA----------------------------GCCAGCCGGG------------------  [230]
D. graeca    CTTTCCCAAGGGGAAGA-GCGCATTTATCAGCCTCGAA----------------------------GCCAGCCGGG------------------  [212]
              *        ******* **** ***** **  *  **                             * * ** *

S. conf      ------------------------------------TTGCAAAACTCGACTCTTT-------GGTGACTCTGGTAACTTT-TGCGGATCGCAAC  [261]
H. discus    ------------------------------------CCGCAAGGCTCGTCTCTT--------GGTGACTCTGGTAACTCT-AGCGGATCGCAAC  [282]
E. adan      CCCACCACCGCCCGGGT------------------TTCGCTCGCCGCGTCTCTCCAACCGATGGTGAATCTGGTAACTTGGATCGGATCGCAAC  [336]
M. dilat     ------------------------------------TCGCAAGGTCCGTCCACCTT-----TGGTGACTCTGGTAACTGT-TGCGGATCGCAAT  [280]
C. cruc      ------------------------------------TCGCAAGGCCCGTCTCTTT-------GGTGACTCTGGTAACTCT-AGCGGATCGCAAT  [258]
P. midas     GGCTCGTCGGACTTGGTGCCCCCACCACCGTCGCGGTACGCTCTCCGCGTTTTCTAAACCGATGGTGAATCTGGTAACTTCGATCGGATCGCAAC  [354]
H. iris      ------------------------------------CCGCAAGGCTCTNCTCTTT-------GGTGACTCTGGTAACTCT-AGCGGATCGCAAC  [262]
B. corn      ------------------------------------TCGCAAGGCCCGTCTCTTT-------GGTGACTCTGGTAACTCT-AGCGGATCGCAAT  [284]
T. gigas     ------------------------------------GTGCAAACCTCGTTGATTCTA-----GGTGACTCTGGTAACCGT-TGCGGATCGCAAT  [282]
H. tuber     ------------------------------------CCGCAAGGCTCGTCTCTTT-------GGTGACTCTGGTAACTCT-AGCGGATCGCAAC  [260]
E. adan2     CCCACCACCGCCCGGGT------------------TTCGCTCGCCGCGTCTCTCCAACCGATGGTGAATCTGGTAACTTGGATCGGATCGCAAC  [335]
C. argy      ------------------------------------TCGCAAGGCCCGTCTCTTT-------GGTGACTCTGGTAACTCT-AGCGGATCGCAAC  [280]
D. graeca    ------------------------------------TCGCAAGGCCCGTCCACCTT-----TGGTGACTCTGGTAACTGT-TGCGGATCGCAAC  [264]
                                                  **      *               ***** **********     **********
```

```
S. conf       GG--TCTCGAACCGGCGACGCGTCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [354]
H. discus     GG--CCTCGAGCCGGCGACGCATCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [375]
E. adan       GGGCCCTCGAGCCGGCGACGCGTCTTTCAAATGTCTGCCCTATCA-ACTTTCGACGGTACGTGCCCTGCCCACCGTGGTCGCAACGGGTGACGGG  [430]
M. dilat      GG--CCCTGAGCCGGCGACGCGTCCATCAAATGTCTGCCCTATCAGACTTTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [373]
C. cruc       GG--CCCTGAGCCGGCGACGCATCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [351]
P. midas      GGGCCTTCGAGCCGGCGACGCGTCTTTCAAATGTCTGCCCTATCA-ACTTTCGACGGTACGTGCCCTGCCCACCGTGGTCGCAACGGGTGACGGG  [448]
H. iris       GG--CCTCGAGCCGGCGACGCATCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [355]
B. corn       GG--CCTTGAGCCGGCGACGCATCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [377]
T. gigas      GA--CCCCGAGTCGGCGACGCGTCCATCAAGTGTCTGCCCTATCAGACTTTCGATGGTAAGTGCTATGCTTACCATGGTTATAACGGGTAACGGA  [375]
H. tuber      GG--CCTCGAGCCGGCGACGCATCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [353]
E. adan2      GGGCCCTCGAGCCGGCGACGCGTCTTTCAAATGTCTGCCCTATCA-ACTTTCGACGGTACGTGCCCTGCCCACCGTGGTCGCAACGGGTGACGGG  [429]
C. argy       GG--CCTTGAGCCGGCGACGCATCTATCAAGTGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [373]
D. graeca     GG--CCCCGAGCCGGCGACGCGTCCATCAAATGTCTGCCCTATCAGACTGTCGATGGTAAGTGCTATGCTTACCATGGTGATAACGGGTAACGGG  [357]
              *       **  ********* **  **** ************** *** **** **** ****  ***  *** ****   ******* ****

S. conf       GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGACACGAG  [448]
H. discus     GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGATACGAG  [469]
E. adan       GAATCAGGGTTCGGTTCCGGAGAGGGAGCATGCGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAACTTACCCACTCCTGGCACGGG  [524]
M. dilat      GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGATACGAG  [467]
C. cruc       GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGACACGAG  [445]
P. midas      GAATCAGGGTTCGATTCCGGAGAGGGAGCATGCGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAACTTACCCACTCCTGGCACGGG  [542]
H. iris       GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGATACGAG  [449]
B. corn       GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGACACGAG  [471]
T. gigas      GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGGGGAAGGCAGCAGGCGCGCAAATTACCCAATCTCGATACGAG  [470]
H. tuber      GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGATACGAG  [447]
E. adan2      GAATCAGGGTTCGGTTCCGGAGAGGGAGCATGCGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAACTTACCCACTCCTGGCACGGG  [523]
C. argy       GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGACACGAG  [467]
D. graeca     GAATCAGGGTTCGATTCCGGAGAGGGAGCATGAGAAACGGCTACCACATCCAAGG-AAGGCAGCAGGCGCGCAAATTACCCAATCTCGATACGAG  [451]
              ************** ****************** ********************** ****************** ******* **  *  *** *
```

```
S. conf      GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGCAATCTAAACGTTTGCACGAGGATCTATTGG  [540]
H. discus    GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGTACTCTAAACGTGTGCACGAGGATCTATTGG  [561]
E. adan      GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCTCGTGATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCCATTGG  [616]
M. dilat     GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGTACTCTAAACGTGTGCACGAGGATCCATTGG  [559]
C. cruc      GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGCACTCTAAACGTGTGCACGAGGATCTATTGG  [537]
P. midas     GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCTCGTGATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCCATTGG  [634]
H. iris      GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGTACTCTAAACGTGTGCACGAGGATCTATTGG  [541]
B. corn      GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGCACTCTAAACGTGTGCACGAGGATCTATTGG  [563]
T. gigas     GAGGTAGTAGACGAAAAATAACAATACTGGGACTCTTTCGAGAGCCTCGTAATTGGAATGAGTGTACTCTAAACGTGTACACGAGGATCCATTGG  [565]
H. tuber     GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGTACTCTAAACGTGTGCACGAGGATCTATTGG  [539]
E. adan2     GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCTCGTGATTGGAATGAGTACACTTTAAACCCTTTAACGAGGATCCATTGG  [615]
C. argy      GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGCACTCTAAACCTGTGCACGAGGATCTATTGG  [559]
D. graeca    GAGGTAGT-GACGAAAAATAACAATAC-GGGACTCTTTCGAG-GCCCCGTAATTGGAATGAGTGTACTCTAAACGTGTGCACGAGGATCCATTGG  [543]
             ******** ****************** ************** *** *** ************  * * ******  ********* *****

S. conf      AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [634]
H. discus    AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGCTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [655]
E. adan      AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGCATTTAAAAAGCTCGTAGTTGGATCTCGGG  [710]
M. dilat     AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCTGTATATTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [654]
C. cruc      AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [631]
P. midas     AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGCATTTAAAAAGCTCGTAGTTGGATCTCGGG  [728]
H. iris      AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGCTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [635]
B. corn      AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [657]
T. gigas     AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATACTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAAG  [659]
H. tuber     AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGCTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [633]
E. adan2     AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGCATTTAAAAAGCTCGTAGTTGGATCTCGGG  [709]
C. argy      AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [653]
D. graeca    AGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC-GTATATTAAAGTTGTTGTGGTTAAAAAGCTCGTAGTTGGATCTCAGG  [637]
             ********************************************** ***** ***** ****************************  *
```

```
S. conf      GT---------CGTTCGCGTCGGTCCC---TTTCCTCA--ATGGTTCGGGT---ACTGC-GCGGTGTCGACCCC---CTTGCCGGACCTCTCAAA  [708]
H. discus    GT---------CGTTCGCGTCGGTCCC---TTGCTGCA--ATGGCTCGGGT---ACTGC-GCGGTGTCGTCCCC---CTTGCCAGACCTCTCAAA  [729]
E. adan      TTTCCGACCGGCGGTCTCGCCACCCCC---TTTCACGAG-AGGGAGCGGGTCTCACCGCCGACGCTTCGGACCCAG-CCGGCCGGCTTTCTCCTC  [800]
M. dilat     GT---------CGTTCGCGTCGATCCTCGGTGTCCTTCA-CAGGTTCCGGTGGCACCAC-GCGTTGTCGTCCCCTGTCCTTCCAGTTCTCTCAGC  [738]
C. cruc      GT---------CGTTCGCGTCGGTCCC---TTGCTTCA--ATGGCTCGGGT---ACTGC-GCGGTGTCGTCCCC---CTTGCCAGACCTCTCAAA  [705]
P. midas     TTTCCGACCGGCGGTCTCGCCAACCCCC--TTTCACGAG-AGGGCGCGGGTCTCACCGCCGCAGCTTCGGACCCAG-CTGGCCGG-CGTCTCCTT  [818]
H. iris      GT---------CGTTCGCGTCGGTCCC---TTGCTGCA--ATGGCTCGGGT---ACTGC-GCGGTGTCGTCCCC---CTTGCCAGACCTCTCAAA  [709]
B. corn      GT---------CGTTCGCGTCGGTCCC---TTGCTGCA--ATGGCTCGGGT---ACTGC-GCGGTGTCGTCCCC---CTTGCCAGACCTCTCAAA  [731]
T. gigas     GC---------CCGACGTGACGGGACGCGGACTCAATTG-CTTGGGACCGTGGAAACGT--CCGTCAGGTGCCC--CCGCTCCAGGCCTCTCAAC  [740]
H. tuber     GT---------CGTTCGCGTCGGTCCC---TTGCTGCA--ATGGCTCGGGT---ACTGC-GCGGYGTCSTCCCC---CTTGCCAGACCTCTCAAA  [707]
E. adan2     TTTCCGACCGGCGGTCTCGCCACCCCC---TTTCACGAG-AGGGAGCGGGTCTCACCGCCGACGCTTCGGACCCAG-CCGGCCGGCTTTCTCCTC  [799]
C. argy      GT---------CGTTCGCGTCGGTCCC---TTGCTTCA--ATGGCTCGGGT---ACTGC-GTGATGTCGTCCCC---CTTGCCAGTCCTCTCAAA  [727]
D. graeca    GT---------CGTTCGCGTCGGTCCTCGGTGTCCACTCGTTGGCTCCGGCGGCACCAC-GCGTTGTCGTCCCCTGACCGTCCAGTCCCCTCAAC  [722]
               **  * * * **  * * ****** * ***

S. conf      AAGGTTGCTCTTGACTGAGTGGCCTGATCGAGTGGCC-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGTA  [800]
H. discus    AAGGTTGCTCTTGACTGAGTGGCCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCC-GCCTGTA  [820]
E. adan      ---GGTGCTCTTGACTGCA--GCGTCGGGGAG-GTCC-GGAACGTTTACTTTGAAAAAATTAGAGTGCTCAAAGCAGGCCGTCGACTCGCCTGAA  [888]
M. dilat     AGGGCTGCTCTTGACTGAGTGGTCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGTCCGCCTGAA  [830]
C. cruc      AAGGTTGCTCTTGACTGAGTGGCCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGTA  [797]
P. midas     ---GGTGCTCTTGACTGCA--GTGT-CTAGGGGGTCC-GGAACGTTTACTTTGAAAAAATTAGAGTGCTCAAAGCAGGCCGTCGACTTGCCGGAA  [906]
H. iris      AAGGTTGCTCTTGACTGAGTGGCCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGTA  [801]
B. corn      AAGGTTGCTCTTGACTGAGTGGCCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGTA  [823]
T. gigas     CTGACTGCTCTTCGATGCGCGGTCTGCCCGAGTGGCTTGGAGAGATTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGTCCGCCTGAA  [833]
H. tuber     AAGGTTGCTCTTGACTGAGTGGCCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGTA  [799]
E. adan2     ---GGTGCTCTTGACTGCA--GCGTCGGGGAG-GTCC-GGAACGTTTACTTTGAAAAAATTAGAGTGCTCAAAGCAGGCCGTCGACTCGCCTGAA  [887]
C. argy      AAGGTTGCTCTTGACTGAGTGGCCTGCTCGAGTGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGTA  [819]
D. graeca    AGGGCTGCTCTTCACTGAGTGGTCTGCTCGAGAGGCT-GGAGAGTTTACTTTGAAAAAATTAGAGTGTTCAAAGCAGGCA--CAGCCCGCCTGAA  [814]
               ********* *  * * * * *  ***  * ********************** ************  *** * *
```

```
S. conf     TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTTG-CT----GGTTTTT----------------------------------  [854]
H. discus   TAATGGTGCATGGAATAATGGAATAGGAACCCCGGATC-CTATTTTGGCT----GGTTTTT----------------------------------  [875]
E. adan     TAATGGTGCATGGAATAATAGAACAGGA-CCTCGGTTCTCTTTTGTGGGTTTTCGCTTTTTCCTCCCAGGCGTCTCTCTCTCTCCGCGGGCTTCG  [982]
M. dilat    TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTCG-CT----GGTTTTACCT-----G--------------------------  [888]
C. cruc     TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTTG-CT----GGTTTTT----------------------------------  [851]
P. midas    TAATGGTGCATGGAATAATAGAACAGGA-CCTCGGTTCTCTTTTGTGGGTTTTCGCTTTCTCTTCTCTCTCTTCGGTGTCTCTTCTCGTCGTTCC  [1000]
H. iris     TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTTG-CT----GGTTTTT----------------------------------  [855]
B. corn     TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTTG-CT----GGTTTTT----------------------------------  [877]
T. gigas    TAATGATGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTCG-CT----GGTTTTACCT-------------------------------  [890]
H. tuber    TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTTG-CT----GGTTTTT----------------------------------  [853]
E. adan2    TAATGGTGCATGGAATAATAGAACAGGA-CCTCGGTTCTCTTTTGTGGGTTTTCGCTTTTTCCTCCCAGGCGTCTCTCTCTCTCCGCGGGCTTCG  [981]
C. argy     TAATGGTGCATGGAATAATAGAATAGGA-CCCCGGATC-CTATTTTG-CT----GGTTTTT----------------------------------  [873]
D. graeca   TAATGGTGCATGGAATAATGGAATAGGA-CCCCGGATC-CTATTTCG-CT----GGTTTTACCTG------------------------------  [872]
            ***** ************ *** **** ** *** ** ** **  *  * * ***

S. conf     ------------------------------------GGAG--CTCAGGG---------------GTAATGATTAAGAGGAACAGGGCGGGGGC
H. discus   ------------------------------------GGAG--CTCAGGG---------------GTAATGATTA-GAGGAACAGGGCGGGGGC
E. adan     GAACCCTTGGCGGGGCGACGGGGTCCCGGGGGGAATCGGAG--CGCAGGGAGGGACGGAACCGGAGGTAATGATTAAGAGAGACAGA-CGGGGGC
M. dilat    ------------------------------------GGATTTCTCAGGG---------------GTAATGATTAAGAGGAACGGCGCGGGGGC
C. cruc     ------------------------------------GGAG--CTCAGGG---------------GTAATGATTAAGAGGAACAGGGCGGGGGC
P. midas    G---CGTC-----------------GGGGAGTTCCCGGAGAAGGCGGAGA---ACGGAACCGGAGGTAATGATTAAGAGAGACAGA-CGGGGGC
H. iris     ------------------------------------GGAG--CTCAGGG---------------GTAATGATTAAGAGGAACAGGGCGGGGGC
B. corn     ------------------------------------GGAG--CTCAGGG---------------GTAATGATTAAGAGGAACAGGGCGGGGGC
T. gigas    -------------G----------------------GGATC-TCCAGGG---------------GTAATGATCAAGAGGGACAGT-CGGGGGC
H. tuber    ------------------------------------GGAG--CTCAGGG---------------GTAATGATTAAGAGGAACAGGGCGGGGGC
E. adan2    GAACCCTTGGCGGGGCGACGGGGTCCCGGGGGGAATCGGAG--CGCAGGGAGGGACGGAACCGGAGGTAATGATTAAGAGAGACAGA-CGGGGGC
C. argy     ------------------------------------GGAG--CTCAGGG---------------GTAATGATTAAGAGGAACAGGGCGGGGGC
D. graeca   ------------------------------------GGATCTCTCAGGG---------------GTAATGATTAAGAGGAACGGCGCGGGGGC
             ***  * * * ******** * ***  ** *  *******
```

```
S. conf      ATCCGTATTGCGGCGTTAGAGGTGAAATTCTTGGATCGTCGCAAGACGCACTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
H. discus    ATCCGTATTACGGCGTTAGAGGTGAAATTCTTGGATCGTCGTAAGGCAC-CTACAGCGAAAGCATTTGCCAAGTAATGTTTTCAT-TAATCAAGA
E. adan      ATTCGTATTGCGGTGTTAGAGGTGAAATTCTTGGATCGCCGCAGGACGGACCACTGCGAAAGCATTTGCCAAGCA-TGTTTTCGTCTAATCAAGA
M. dilat     ATTCGTATTGCGACGCTAGAGGTGAAATTCTTGGATCGTCGCAAGACGCACTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
C. cruc      ATCCGTATTACGGCGTTAGAGGTGAAATTCTTGGATCGTCGTAAGACGCCCTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
P. midas     ATTCGTATTGCGGTGTTAGAGGTGAAATTCTTAGATCGCCGCAGGACGGGCCACTGCGAAAGCATTTGCCAAGCA-TGTTTTCGTCTAATCAAGA
H. iris      ATCCGTATTACGGCGTTAGAGGTGAAATTCTTGGATCGTCGTAAGACGCCCTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
B. corn      ATCCGTATTACGGCGTTAGAGGTGAAATTCTTGGATCGTCGTAAGACGCCCTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
T. gigas     ATCCGTATTGTGGCGTTAGAGGTGAAATTCTTGGATCGCCGCAAGACGCTCGACAGCGAAAGCATTTGCCTAGGA-TGTTTTCAT-TAATCAAGA
H. tuber     ATCCGTATTACGGCGTTAGAGGTGAAATTCTTGGATCGTCGTAAGACGCCCTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
E. adan2     ATTCGTATTGCGGTGTTAGAGGTGAAATTCTTGGATCGCCGCAGGACGGACCACTGCGAAAGCATTTGCCAAGCA-TGTTTTCGTCTAATCAAGA
C. argy      ATCCGTATTACGGCGTTAGAGGTGAAATTCTTGGATCGTCGTAAGACGCCCTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
D. graeca    ATCCGTATTGCGGCGCTAGAGGTGAAATTCTTGGATCGTCGCAAGACGCACTACAGCGAAAGCATTTGCCAAGTA-TGTTTTCAT-TAATCAAGA
             ** ******  *  * ***************** ***** ** * * ** ** *************** ** * ******* * *********

S. conf      ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
H. discus    ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
E. adan      ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAGCCGGCGATCCGCCGGAGTTGCT-TCGATGACT
M. dilat     ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCTATTGGCGGCCCGTCGGCGTTGCTCTATTCGACT
C. cruc      ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
P. midas     ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAGCCGGCCATCCGCCGGAGTTGCT-TCGATGACT
H. iris      ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
B. corn      ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
T. gigas     ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCGGTCCGTTGGCGTTGCCGAATTCGACT
H. tuber     ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTSGTAGTTCTGMCCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
E. adan2     ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCATAAACGATGCCAGCCGGCGATCCGCCGGAGTTGCT-TCGATGACT
C. argy      ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAACTGGCAGCCCGTCGGCGTTGCTCTATTCGACT
D. graeca    ACGAAAGTCAGAGGTTCGAAGACGATCAGATACCGTCGTAGTTCTGACCGTAAACTATGCCAATTGGCGGCCCGTCGGCGTTGCTCTATTCGACT
             ************************************ ********* ** ***** ***** ******  ** ***** ****
```

```
S. conf       CGGCGGGAGGCCTACCTGAAAAGA--AAGTCTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
H. discus     CGGCGGGAGGCCTACCCGAAAGGA--AAGTCTACGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
E. adan       CGGCGGGCAGCCTCCGGGAAACCA--AAGCTTTTGGGTTCTGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
M. dilat      CGGCGGGAAGCCTACGAGAAATTATTAAATTTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
C. cruc       CGGCGGGAGGCCTACCGAAAACGA--AAGTCTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
P. midas      CGGCGGGCAGCCTCCGGGAAACCA--AAGCTTTTGGGTTCTGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
H. iris       CGGCGGGAGGCCTACCCGAAAGGA--AAGTCTACGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
B. corn       CGGCGGGAGGCCTACCGAAAACGA--AAGTCTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
T. gigas      CGATGGGCCGCCCCCGAGAAATCT--AAGTTTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
H. tuber      CGGCGGGAGGCCTACCGGAAAGGA--AAGTCTWCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
E. adan2      CGGCGGGCAGCCTCCGGGAAACCA--AAGCTTTTGGGTTCTGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
C. argy       CGGCGGGAGGCCTACCGAAAACGA--AAGTCTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
D. graeca     CGGCGGGAAGCCTACGAGAAATTATTAAATTTTCGGGTTCCAGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCA
              **  ***  ***  ****  ***  ******  **************************************************

S. conf       CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGGAAGGATTGACAGATTAAAAGCTCTTTCTTG
H. discus     CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCTGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
E. adan       CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGCAAGGATTGACAGATCGATAGCTCTTTCTTG
M. dilat      CCAGGAGTGGGACCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACTCGGTCCGGACACTGGAAGGATTGACAGATTGACAGCTCTTTCTTG
C. cruc       CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
P. midas      CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGCAAGGATTGACAGATCGATAGCTCTTTCTTG
H. iris       CCAGGAGTGG-GACTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCTGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
B. corn       CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
T. gigas      CCAGGAGTGGGACCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACTCGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
H. tuber      CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGRAAACTCACCTGGTCCGGACACTGGAAGGATTGACAGATTGAGAGCTCTTTCTTG
E. adan2      CCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGCAAGGATTGACAGATCGATAGCTCTTTCTTG
C. argy       CCAGGAGTGGAGCCCGCGGCTTAATTTGACTCAACACGGGAAAACTCACCCGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
D. graeca     CCAGGAGTGGGACCTGCGGCTTAATTTGACTCAACACGGGAAAACTCACTCGGTCCGGACACTGGAAGGATTGACAGATTGAAAGCTCTTTCTTG
              ********** ************************ ********  ************ **************  * ************
```

```
S. conf      ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
H. discus    ATTCAGTGGGTGGTGGTGCATGGCCGTTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
E. adan      ATTCGGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTAC
M. dilat     ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
C. cruc      ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
P. midas     ATTCGGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTAC
H. iris      ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
B. corn      ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
T. gigas     ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
H. tuber     ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
E. adan2     ATTCGGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTAC
C. argy      ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
D. graeca    ATTCAGTGGGTGGTGGTGCATGGCCGTT-CTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCTGCTAACTA-
             **** ********************** ************************************************************

S. conf      GTTCGCC----------------GACAGCTCGTTGTC-----------------------------------GCGCGGCGC-----------CAT
H. discus    GTTCGCC----------------GACAGCTCGTTGTC-----------------------------------GCGCGGCGC-----------CAT
E. adan      GTTCGCC----------------GATCCGACGAGCTC-----------------------------------GAG--------------------
M. dilat     GTTCGCC----------------GATTCCTTTGAATC-----------------------------------GCGCGTCGCATCGTTGGGTTTCGG
C. cruc      GTTCGCC----------------GACAACTC-TTGTC-----------------------------------GCGCGGCGT-----------CAC
P. midas     GTTCGCC----------------GATCCGTCGTAG-C-----------------------------------TCGTG--------------------
H. iris      GTTCGCC----------------GACAGCTCGTTGTC-----------------------------------GCGCGGCGC-----------CAT
B. corn      GTTCGCC----------------GACAGCTCGTTGTC-----------------------------------GCGCGGCGT-----------CAC
T. gigas     GTTCGCCCGTTCCCATCGTGTTTGTTGGGTCGTTGTCCTCGGATGGCGGCTCGGCGGGCTCGGGGAACGGTGCGTGGCGCTGCTGGGGGGCAACC
H. tuber     GTTCGCC----------------GACAGCTCGTTGTC-----------------------------------GCGCGGCGC-----------CAT
E. adan2     GTTCGCC----------------GATCCGACGAGCTC-----------------------------------GAG--------------------
C. argy      GTTCGCC----------------GACAGCTCGTTGTC-----------------------------------GCGCGGCGT-----------CAC
D. graeca    GTTCGCC----------------GATTCAACTTCGAATC---------------------------------GCGCGTCGCTCCGTTGGGTTTCGG
             ******* * *
```

```
S. conf      TTCGGTGGCA--AACTTCTTAGAGGGACAGGTGGCGTTTAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
H. discus    TTCGGTGGCA--AACTTCTTAGAGGGACAG-TGGCGTATAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCAGGGCCG
E. adan      -TCGGCGGCCC-AACTTCTTAGAGGGACAAGTGGCGCACAGCCACGCGAAATTGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
M. dilat     CTCGGCGGTGCTAACTTCTTAGAGGGACAGGTGGCGTTTAGTCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGAGGCCG
C. cruc      TTCGGTGGCA--AACTTCTTAGAGGGACAGGTGGCGTTTAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
P. midas     -TCGGCGGCC--AACTTCTTAGAGGGACAAGTGGCGTTTAGCCACGCGAAATTGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
H. iris      TTCGGTGGCA--AACTTCTTAGAGGGACAGGTGGCGTATAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCAGGGCCG
B. corn      TTCGGTGGCA--AACTTCTTAGAGGGACAGGTGGCGTTTAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
T. gigas     TTCGGTGGCC--AACTTCTTAGAGGGACAGGTGGCGTTTAGTCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGAGGCCG
H. tuber     TTCGGTGGCA--AACTTCTTAGAGGGACAGGTGGCGTATAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCAGGGCCG
E. adan2     -TCGGCGGCCC-AACTTCTTAGAGGGACAAGTGGCGCACAGCCACGCGAAATTGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
C. argy      TTCGGTGGCA--AACTTCTTAGAGGGACAGGTGGCGTTTAGCCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCCG
D. graeca    CTCGGCGGCGCTAACTTCTTAGAGGGACAGGTGGCGTTTAGTCACACGAAGTAGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGAGGCCG
              **** ** ****************  ******* *** **** * **********************************  *****

S. conf      CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCTACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
H. discus    CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
E. adan      CACGCGCGCTACAATGGAAGAAGCAGCGTGGATACCACCCTGCTCCGAGAGGAGTGGGAAACCCCTTGAATATCTTTCTCGATGGGGATCGGGGC
M. dilat     CACGCGCGTCACACTGAATGAATCAACGTGGATGCTCACCTGCTCCGAGAGGAGTAGGAAACCCGGTGAATCTCATTCGTGATGGGGATTGGGGC
C. cruc      CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
P. midas     CACGCGCGCTACAATGGAAGAAGCAGCGTGGATACC-ACCTGCTCCGAGAGGAGTGGGAAACCCGTTGAATATCTTTCTCGATGGGGATCGGGGC
H. iris      CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
B. corn      CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
T. gigas     CACGCGCGTCACACTGAATGAATCAACGTGGATGCTCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCATTCGTGATGGGGATTGGGGC
H. tuber     CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
E. adan2     CACGCGCGCTACAATGGAAGAAGCAGCGTGGATACCACCCTGCTCCGAGAGGAGTGGGAAACCCCTTGAATATCTTTCTCGATGGGGATCGGGGC
C. argy      CACGCGCGCTACACTGAAAGAATCAACGTGGATGCCCACCTGCTCCGAGAGGAGTGGGAAACCCGATGAATCTCTTTCGTGATGGGGATTGGGGC
D. graeca    CACGCGCGTCACACTGAATGAATCAACGTGGATGCTCACCTGCTCCGAGAGGAGTAGGAAACCCGATGAATCTCATTCGTGATGGGGATTGGGGC
             ********  *** ** * *** ** ******* ****************** ********  ***** ** ***  ********* *****
```

```
S. conf      TTGTAATTTTTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATCAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
H. discus    TTGTTAATTATCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATCAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
E. adan      TTGTAATTGTTCCCCGTGAACGAGGAATTCCCAGTAAGCGCCAGTCACAAGCTGGCGCTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
M. dilat     TTGTAATTGTTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATAAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
C. cruc      TTGTAATTATTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATAAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
P. midas     TTGTAATTGTTCCCCGTGAACGAGGAATTCCGAGTAAGCGCCAGTCATAAGCTGGCGCTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
H. iris      TTGTAATTATTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATCAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
B. corn      TTGTAATTATTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATAAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
T. gigas     T-GTAGTTGTTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATAAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
H. tuber     TTGTAATTATTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATCAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
E. adan2     TTGTAATTGTTCCCCGTGAACGAGGAATTCCCAGTAAGCGCCAGTCACAAGCTGGCGCTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
C. argy      TTGTAATTATTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATAAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
D. graeca    TTGTAATTGTTCCCCATGAACGAGGAATTCCCAGTAAGCGTCAGTCATAAGCTGGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCG
             * ***  ***** ********************** ******** ******  ******** ************************************

S. conf      CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCTGTTTG-----------CTTCTTTCGGGAGGCGA--ACAG-GGACGGAAAG
H. discus    CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAATCCGTTTG-----------CTGCTTTCGGGTGGCGG--ACGG-GGACGGAAAG
E. adan      CTACTACCGATTGAACGGTTTAGTGAGGTCTTCGGATTGGTCCCGGAACGGGCGGGAGTCACCGCTCTCGTTCGGC----ACGGCGGCCGAAAAT
M. dilat     CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCTGCTGGG--TG------GTTCTTTCGGGGATCACCTGCGG-GGACGGAAAG
C. cruc      CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCCGTCTG-----------CCGCTTTC--------------------------
P. midas     CTACTACCGATTGAACGGTTTAGTGAGGTCTTCGGATTGGGCCCGGGACGGCCGGG-GGCTCCGCCCTCGTTCGGC----ACGGCGGCCGAAAAT
H. iris      CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAATCCGTTTG-----------CTGCTTTCGGGTGGCGG--ACGG-GGACGGAAAG
B. corn      CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCCGTCTG-----------CCGCTTTCGGGTGGCGG--ACGG-GGACGGAAAG
T. gigas     CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCCGCGGGCC-----GCCGACTCGTTCGGCCGGCCT--TCGG-GGACGGAAAG
H. tuber     CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAATCCGTTTG-----------CTGCTTTCGGGTGGCGG--ACGG-GGACGGAAAG
E. adan2     CTACTACCGATTGAACGGTTTAGTGAGGTCTTCGGATTGGTCCCGGAACGGGCGGGAGTCACCGCTCTCGTTCGGC----ACGGCGGCCGAAAAT
C. argy      CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCCGACTG-----------CCGCTTTCGGGTGGCGG--TCGG-GGACGGAAAG
D. graeca    CTACTACCGATTGAATGGTTTAGTGAGACCCTCGGACTGGAGTCCGCTGGGTGG-------CCTCTTTCGGGGGGTCGCCTGCGGGGACGGAAAG
             *************** ***********  * ***** **** **  *  **
```

```
S. conf     ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
H. discus   ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTA
E. adan     TTGATCAAACTTGATCGTTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
M. dilat    ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTA
C. cruc     --------------------------------------------------------------------------
P. midas    TTGATCAAACTTGATCGTTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
H. iris     ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
B. corn     ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATT-
T. gigas    ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTA
H. tuber    ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
E. adan2    TTGATCAAACTTGATCGTTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
C. argy     ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGA------
D. graeca   ACGGTCGAACTTGATCATTTAGAGGAAGTAAAAGTCGTAACAAGGTTTCC--------------------------
```

Species Abbreviation:	Species Name:	Accession Number:
S. conf	*Sinezona confusa*	AF120512.1
H. discus	*Nordotis discus*	AF082177.1
E. adan	*Entemnotrochus adansonianus*	AF120509.1
M. dilat	*Macroschisma dilatatum*	AF335560.1
C. cruc	*Clanculus cruciatus*	AF120514.1
P. midas	*Perotrochus midas*	AF120510.1
H. iris	*Haliotis iris*	AF492441.1
B. corn	*Batillus cornutus*	AF165311.1
T. gigas	*Tugali gigas*	AF335561.1
H. tuber	*Haliotis tuberculata*	AF120511.1
E. adan2	*Entemnotrochus adansonianus*	AY090809.1
C. argy	*Chlorostoma argyrostoma*	AF335562.1

APPENDIX D -AA CODE AND CONSERVED AA GROUPS (GONNET PAM250)

Standard Genetic Code

Alanine	A/Ala	GCU, GCC, GCA, GCG
Arginine	R/Arg	CGU, CGC, CGA, CGG, AGA, AGG
Asparagine	N/Asn	AAU, AAC
Aspartate	D/Asp	GAU, GAC
Cysteine	C/Cys	UGU, UGC
Glutamate	E/Glu	GAA, GAG
Glutamine	Q/Gln	CAA, CAG
Glycine	G/Gly	GGU, GGC, GGA, GGG
Histidine	H/His	CAU, CAC
Isoleucine	I/Ile	AUU, AUC, AUA
Leucine	L/Leu	UUA, UUG, CUU, CUC, CUA, CUG
Lysine	K/Lys	AAA, AAG
Methionine	M/Met	AUG
Phenylalanine	F/Phe	UUU, UUC
Proline	P/Pro	CCU, CCC, CCA, CCG
Serine	S/Ser	UCU, UCC, UCA, UCG, AGU, AGC
Threonine	T/Thr	ACU, ACC, ACA, ACG
Tryptophan	W/Trp	UGG
Tyrosine	Y/Tyr	UAU, UAC
Valine	V/Val	GUU, GUC, GUA, GUG
Stop		UAA, UAG, UGA

Strong Group (:)

STA
NEQK
NGQK
QHRK
MILV
MILF
HY
FYW

Weaker Group (.)

CSA
ATV
SAG
STNK
STPA
SGND
SNDEQK
NDEQHK
FVLIM
HFY

APPENDIX E -*H. IRIS* ACTIN RESTRICTION SITES

Restriction Enzyme	Recognition Site	*H.iris* A1	*H.iris* A1a	*H.iris* A1b	*H.iris* A1c	*H.iris* A2	*H.iris* A3
Aat II	gacgt↓c	0	0	0	0	0	1
Acc I	gt↓mkac	2	0	0	0	0	0
*Acc*65 I	g↓gtacc	1	0	0	0	0	1
Aci I	ccgc($^{-3}$/$_{-1}$)↓	4	1	1	1	3	3
Acl I	aa↓cgtt	0	0	0	0	1	1
Afe I	agc↓gct	1	0	0	0	0	0
Afl II	c↓ttaag	0	1	0	1	0	0
Afl III	a↓crygt	1	2	0	0	1	2
Age I	a↓ccggt	0	0	0	0	0	0
Ahd I	gacnnn↓nngtc	0	0	0	0	0	0
Alu I	ag↓ct	7	5	5	6	9	8
Alw I	ggatc(n)^{4}/$_{5}$↓	4	2	2	2	4	2
*Alw*N I	cagnnn↓ctg	0	0	0	0	0	0
Apa I	gggcc↓c	0	0	0	1	0	0
*Apa*L I	g↓tgcac	0	0	0	0	2	0
Apo I	r↓aatty	1	1	1	0	2	1
Asc I	gg↓cgcgcc	0	0	0	0	0	0
Ase I	at↓taat	0	0	0	0	2	1
Ava I	c↓ycgrg	0	0	0	0	1	1
Ava II	g↓gwcc	1	0	0	0	1	1
Avr II	c↓ctagg	0	0	0	0	0	0
Bae I	↓^{10}/$_{15}$(n)acnnnngtayc(n)^{12}/$_{7}$↓	0	0	0	0	0	0
*Bam*H I	g↓gatcc	0	0	0	0	0	0
Ban I	g↓gyrcc	1	0	0	0	0	1
Ban II	grgcy↓c	1	1	1	2	1	1
Bbs I	gaagac(n)^{2}/$_{6}$↓	4	0	0	0	2	2
Bbv I	gcagc(n)^{8}/$_{12}$↓	2	0	0	0	2	2
*Bbv*C I	cctcagc^{-5}/$_{-2}$↓	0	0	0	0	0	0
Bcg I	↓^{10}/$_{12}$(n)cgannnnnntgc(n)^{12}/$_{10}$↓	0	0	0	0	1	1
*Bci*V I	gtatcc(n)^{6}/$_{5}$↓	1	1	0	1	2	2
Bcl I	t↓gatca	0	0	0	0	1	0
Bfa I	c↓tag	1	0	1	0	0	0
Bgl I	gccnnnn↓nggc	1	0	1	1	1	2
Bgl II	a↓gatct	1	1	1	1	1	1
Blp I	gc↓tnagc	0	0	0	0	0	0
Bmr I	actggg(n)^{5}/$_{4}$↓	1	1	1	1	1	1
Bpm I	ctggag(n)^{16}/$_{14}$↓	2	2	0	1	4	3
*Bpu*10 I	cctnagc^{-5}/$_{-2}$↓	0	0	0	0	0	0
Bsa I	yag↓ctr	2	1	1	0	2	1
*Bsa*A I	yac↓gtr	1	1	1	0	0	1
*Bsa*B I	gatnn↓nnatc	3	0	0	0	3	3

*Bsa*H I	gr↓cgyc	0	0	0	0	0	2
*Bsa*J I	c↓cnngg	3	1	2	2	4	5
*Bsa*W I	w↓ccggw	1	0	0	0	1	0
*Bsa*X I	↓$^{9}/_{12}$(n)acnnnnnctcc(n)$^{10}/_{7}$↓	2	2	1	1	2	2
*Bse*R I	gaggag(n)$^{10}/_{8}$↓	0	2	1	2	3	0
Bsg I	gtgcag(n)$^{16}/_{14}$↓	0	0	0	0	1	0
*Bsi*E I	cgry↓cg	2	1	2	2	1	1
*Bsi*HKA I	gwgcw↓c	2	4	2	1	5	2
*Bsi*W I	c↓gtacg	0	0	0	0	1	0
Bsl I	ccnnnnn↓nngg	4	2	1	2	0	3
Bsm I	gaatgc(n)$^{1}/_{-1}$ ↓	3	0	0	0	1	1
*Bsm*A I	gtctc(n)$^{1}/_{5}$↓	4	3	4	2	6	2
*Bsm*B I	cgtctc(n)$^{10}/_{8}$↓	0	0	1	0	1	0
*Bsm*F I	gggac(n)$^{10}/_{14}$↓	4	2	2	2	4	5
*Bsp*1286 I	gdgch↓c	2	4	2	2	5	2
*Bsp*E I	t↓ccgga	1	0	0	0	1	0
*Bsp*H I	t↓catga	3	1	2	2	3	2
*Bsp*M I	acctgc(n)$^{4}/_{8}$↓	0	0	0	0	0	0
Bsr I	actgg(n)$^{1}/_{-1}$↓	1	1	1	2	2	3
*Bsr*B I	ccgctc$^{-3}/_{-3}$↓	0	0	0	0	0	0
*Bsr*D I	gcaatg(n)$^{2}/_{0}$↓	0	0	0	1	1	1
*Bsr*F I	r↓ccggy	3	0	1	1	0	1
*Bsr*G I	t↓gtaca	1	0	0	0	1	1
*Bss*H II	g↓cgcgc	0	0	0	0	0	0
*Bss*K I	↓ccngg	4	1	1	2	5	5
*Bss*S I	cacgag$^{-5}/_{-1}$ ↓	1	0	0	0	0	0
*Bst*AP I	gcannnn↓ntgc	1	0	0	0	0	0
*Bst*B I	tt↓cgaa	0	0	0	1	0	1
*Bst*E II	g↓gtnacc	0	0	0	0	0	0
*Bst*F5 I	ggatg(n)$^{2}/_{0}$↓	8	0	2	2	4	5
*Bst*N I	cc↓wgg	4	1	1	2	3	4
*Bst*U I	cg↓cg	1	0	1	0	0	0
*Bst*X I	ccannnnn↓ntgg	1	0	0	0	0	0
*Bst*Y I	r↓gatcy	3	3	3	3	2	2
*Bst*Z17 I	gta↓tac	1	0	0	0	0	0
*Bsu*36 I	cc↓tnagg	1	0	0	0	0	0
Btg I	c↓crygg	1	1	1	1	0	1
Bts I	gcagtg(n)$^{2}/_{0}$↓	1	0	1	0	1	1
*Cac*8 I	gcn↓ngc	3	1	1	1	2	1
Cla I	at↓cgat	0	0	0	0	0	0
Dde I	c↓tnag	6	5	3	5	4	4
Dpn I	ga↓tc	9	5	4	5	10	9
Dra I	ttt↓aaa	0	1	1	1	1	2
Dra III	cacnnn↓gtg	0	0	0	0	0	0
Drd I	gacnnnn↓nngtc	1	1	1	1	2	3
Eae I	y↓ggccr	3	2	2	3	2	1
Eag I	c↓ggccg	2	1	1	2	1	0
Ear I	ctcttc(n)$^{1}/_{4}$↓	3	0	0	0	3	4

Eci I	ggcgga(n)^{11}/$_{9}$↓	0	0	0	0	0	0
*Eco*N I	cctnn↓nnnagg	0	1	1	1	0	1
*Eco*O109 I	rg↓gnccy	0	0	0	2	0	0
*Eco*R I	g↓aattc	1	0	0	0	0	0
*Eco*R V	gat↓atc	1	0	0	0	0	1
Fau I	cccgc(n)^{4}/$_{6}$↓	0	0	0	0	0	0
*Fnu*4H I	gc↓ngc	2	0	0	0	4	3
Fok I	ggatg(n)^{9}/$_{13}$↓	8	0	2	2	4	5
Fse I	ggccgg↓cc	0	0	0	0	0	0
Fsp I	tgc↓gca	0	0	0	0	1	0
Hae II	rgcgc↓y	1	0	0	0	0	0
Hae III	gg↓cc	4	2	3	5	5	5
Hga I	gacgc(n)^{5}/$_{10}$↓	3	0	0	0	0	2
Hha I	gcg↓c	1	0	1	0	1	0
Hinc II	gty↓rac	2	0	0	0	0	0
Hind III	a↓agctt	0	0	0	1	0	0
Hinf I	g↓antc	9	1	4	1	4	3
*Hin*P1 I	g↓cgc	1	0	1	0	1	0
Hpa I	gtt↓aac	0	0	0	0	0	0
Hph I	ggtga(n)^{8}/$_{7}$↓	9	6	6	6	5	10
*Hpy*188 I	tcn↓ga	5	5	6	4	4	1
*Hpy*99 I	cgwcg↓	5	2	4	2	7	4
*Hpy*CH4 III	can↓gt	7	5	4	3	9	5
*Hpy*CH4 IV	a↓cgt	4	4	5	2	3	4
*Hpy*CH4 V	tg↓ca	14	0	2	2	9	6
Kas I	g↓gcgcc	0	0	0	0	0	0
Kpn I	ggtac↓c	1	0	0	0	0	1
Mbo II	gaaga(n)^{8}/$_{7}$↓	11	3	3	2	10	10
Mfe I	c↓aattg	0	0	0	0	0	1
Mlu I	a↓cgcgt	0	0	0	0	0	0
Mly I	gagtc(n)^{5}/$_{5}$↓	2	1	1	0	3	2
Mnl I	cctc(n)^{7}/$_{6}$↓	22	12	10	13	17	16
Msc I	tgg↓cca	1	0	1	1	1	1
Mse I	t↓taa	3	3	2	3	12	8
Msl I	caynn↓nnrtg	2	0	1	1	1	1
Msp I	c↓cgg	5	0	1	1	4	4
*Msp*A1 I	cmg↓ckg	2	1	0	1	0	0
Mwo I	gcnnnnn↓nngc	6	1	1	1	2	5
Nae I	gcc↓ggc	1	0	1	1	0	0
Nar I	gg↓cgcc	0	0	0	0	0	0
Nci I	cc↓sgg	0	0	0	0	2	1
Nco I	c↓catgg	0	0	0	0	0	0
Nde I	ca↓tatg	0	1	0	0	0	1
*Ngo*M IV	g↓ccggc	1	0	1	1	0	0
Nhe I	g↓ctagc	0	0	0	0	0	0
Nla III	catg↓	16	6	7	4	13	11
Nla IV	ggn↓ncc	5	2	1	3	4	5
Not I	gc↓ggccgc	0	0	0	0	0	0

Nru I	tcg↓cga	0	0	0	0	0	0
Nsi I	atgca↓t	4	0	0	0	1	2
Nsp I	rcatg↓y	4	2	0	0	2	4
Pac I	ttaat↓taa	0	0	0	0	0	0
Pci I	a↓catgt	1	1	0	0	1	2
*Pfl*M I	ccannnn↓ntgg	0	0	0	0	0	0
Ple I	gagtc(n)$^{4}/_{5}$↓	2	1	1	0	3	2
Pme I	gttt↓aaac	0	0	0	0	0	0
Pml I	cac↓gtg	0	0	0	0	0	0
*Ppu*M I	rg↓gwccy	0	0	0	0	0	0
*Psh*A I	gacnn↓nngtc	0	0	0	0	1	0
Psi I	tta↓taa	1	0	0	0	2	0
*Psp*G I	↓ccwgg	4	1	1	2	3	4
*Psp*OM I	g↓ggccc	0	0	0	1	0	0
Pst I	ctgca↓g	0	0	0	0	0	0
Pvu I	cgat↓cg	0	0	0	0	0	0
Pvu II	cag↓ctg	1	0	0	0	0	0
Rsa I	gt↓ac	7	4	3	3	7	6
Rsr II	cg↓gwccg	0	0	0	0	0	0
Sac I	gagct↓c	1	1	1	1	1	1
Sac II	ccgc↓gg	0	0	0	0	0	0
Sal I	g↓tcgac	0	0	0	0	0	0
Sap I	gcttcttc(n)$^{1}/_{4}$↓	1	0	0	0	0	1
*Sau*3A I	↓gatc	9	5	4	5	10	9
*Sau*96 I	g↓gncc	1	0	0	2	3	2
Sbf I	cctgca↓gg	0	0	0	0	0	0
Sca I	agt↓act	0	0	0	0	0	0
*Scr*F I	cc↓ngg	4	1	1	2	5	5
*Sex*A I	a↓ccwggt	1	0	0	1	0	0
*Sfa*N I	gcatc(n)$^{5}/_{9}$↓	5	1	2	2	4	2
Sfc I	c↓tryag	2	0	0	0	0	0
Sfi I	ggccnnnn↓nggcc	0	0	0	0	0	0
Sfo I	ggc↓gcc	0	0	0	0	0	0
*Sgr*A I	cr↓ccggyg	1	0	0	0	0	1
Sma I	ccc↓ggg	0	0	0	0	0	0
Sml I	c↓tyrag	3	2	2	4	1	1
*Sna*B I	tac↓gta	1	0	1	0	0	0
Spe I	a↓ctagt	1	0	0	0	0	0
Sph I	gcatg↓c	2	0	0	0	0	0
Ssp I	aat↓att	0	1	0	0	3	1
Stu I	agg↓cct	0	0	0	0	0	0
Sty I	c↓cwwgg	0	0	1	1	1	1
Swa I	attt↓aaat	0	0	1	0	0	1
Taq I	t↓cga	7	2	2	3	4	4
Tfi I	g↓awtc	7	0	3	1	1	1
Tse I	g↓cwgc	2	0	0	0	2	2
*Tsp*45 I	↓gtsac	7	4	4	4	4	8
*Tsp*509 I	↓aatt	5	2	2	1	13	7

*Tsp*R I	castgnn↓	10	4	8	4	12	6
*Tth*111 I	gacn↓nngtc	0	0	0	0	0	0
Xba I	t↓ctaga	0	0	0	0	0	0
Xcm I	ccannnnn↓nnnntgg	1	1	1	1	1	0
Xho I	c↓tcgag	0	0	0	0	0	0
Xma I	c↓ccggg	0	0	0	0	0	0
Xmn I	gaann↓nnttc	0	0	0	0	0	0

APPENDIX F - *H. VIRGINEA* ACTIN SEQUENCES

H.virgA1a:

```
AGAGCTGTCTTCCCCTGCATCGTCGGCCGACCCAGACATCAGgataatatagtttggatg  [60]
 R  A  V  F  P  C  I  V  G  R  P  R  H  Q                     [14]
ttgctatggaaacatagaggtacaggacattatttgagacagtagtatttgagcatacat  [120]
catggtcttagtaagaactagcgacgacgtggtaattcaattctcagagaggttgtagcc  [180]
agatcgcggtcgaatcacttgccagcctcagcagaaatattgcacataaatgattttgaa  [240]
gataacaaaatatattgaaacacaactatagtttacaataaaaaaaggcaatacagggt  [300]
attacacaaggcagcagctaaaaggcaacaatgccctataccatgtgtacaaggccactc  [360]
acggaatgtccagaagtgaattcttttcaaactgcattgtgccattatttgaatcataac  [420]
taataagattcgtaaaaattacataatgaatgaattttgtgtttcagGGCGTGATGGTCG  [480]
                                                 G  V  M  V    [18]
GTATGGGTCAGAAAGACAGCTACGTTGGAGACGAGGCTCAGTCCAAGAGAGGTATCCTAA  [540]
G  M  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K  R  G  I  L    [38]
CACTCAAGTACCCCATCGAGCACGGAATCGTCACCAACTGGGACGATATGGAGAAGATCT  [600]
T  L  K  Y  P  I  E  H  G  I  V  T  N  W  D  D  M  E  K  I    [58]
GGCATCACACCTTCTACAACGAGCTCCGTGTTGCACCCGAGGAGCACCCCGTCCTCCTGA  [660]
W  H  H  T  F  Y  N  E  L  R  V  A  P  E  E  H  P  V  L  L    [78]
CAGAGGCTCCCCTCAACCCCAAGGCCAACCGTGAAAAGATGACCCAGATCATGTTCGAGA  [720]
T  E  A  P  L  N  P  K  A  N  R  E  K  M  T  Q  I  M  F  E    [98]
CCTTCAACTCTCCAGCTATGTATGTGGCCATCCAGGCTGTTCTGTCTCTGTATGCTTCTG  [780]
T  F  N  S  P  A  M  Y  V  A  I  Q  A  V  L  S  L  Y  A  S    [118]
GTCGTACCACGGGTATTGTGTTGGACTCTGGTGATGGTGCCACCCACAcTGTGCCCATCT  [840]
G  R  T  T  G  I  V  L  D  S  G  D  G  A  T  H  T  V  P  I    [138]
ATGAAGGTTACGCCCTTCCCCACGCCATCATGAGGTTGGATCTTGCCGGTCGTGACCTGA  [900]
Y  E  G  Y  A  L  P  H  A  I  M  R  L  D  L  A  G  R  D  L    [158]
CTGATTACCTCATGAAGATCCTCACTGAGCGTGGTTACTCCTTCACCACCACCGCTGAGA  [960]
T  D  Y  L  M  K  I  L  T  E  R  G  Y  S  F  T  T  T  A  E    [178]
GAGAAATCGTCAGAGACATCAAAGAGAAGCTCTGCTACATGCCCCTTG              [1008]
R  E  I  V  R  D  I  K  E  K  L  C  Y  M  P  L                [194]
```

H.virgA1b

```
AGAGCTGTCTTCCCCTCCATCGTGGGTCGCCCAAGACATCAGgtaatatagtttagtggt  [60]
 R  A  V  F  P  S  I  V  G  R  P  R  H  Q                     [14]
gctgtggaaacatagaggttcaaagattaatttcatataaaratactaagtttcaccgatg  [120]
gtattcgagttaaaatcccctaaratratgtattattacaacgtggcaaattcagttctc  [180]
ataacgttgtcgccttatgtgaaagcagaaataacatacgtgaatgattttgaaaataag  [240]
aataaaatttgtttcagGGCGTAATGGTCGGTATGGGACAGAAAGACAGCTACGTTGGAG  [300]
                    G  V  M  V  G  M  G  Q  K  D  S  Y  V  G  [28]
ACGAGGCTCAGTCCAAGAGAGGTATCCTCACACTCAAGTACCCCATCGAGCACGGAATCG  [360]
D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P  I  E  H  G  I    [48]
TCACCAACTGGGACGACATGGAGAAGATCTGGCATCACACCTTCTACAACGAGCTCCGTG  [420]
V  T  N  W  D  D  M  E  K  I  W  H  H  T  F  Y  N  E  L  R    [68]
TTGCACCCGAGGAGCATCCCGTCCTCCTGACAGAGGCTCCCCTCAACCCCAAGGCTAACC  [480]
V  A  P  E  E  H  P  V  L  L  T  E  A  P  L  N  P  K  A  N    [88]
GTGAAAAGATGACCCAAATCATGTTCGAGACCTTCAACTCTCCAGCTATGTATGTGGCCA  [540]
R  E  K  M  T  Q  I  M  F  E  T  F  N  S  P  A  M  Y  V  A    [108]
TCCAGGCTGTTCTGTCTCTGTATGCTTCTGGCCGTACCACGGGTATTGTGTTGGACTCTG  [600]
I  Q  A  V  L  S  L  Y  A  S  G  R  T  T  G  I  V  L  D  S    [128]
GTGATGGTGTCACCCACACTGTGCCCATCTATGAAGGTTACGCCCTTCCCCACGCCATCA  [660]
G  D  G  V  T  H  T  V  P  I  Y  E  G  Y  A  L  P  H  A  I    [148]
```

```
TGAGGTTGGATCTTGCCGGTCGTGGTCTGACTGATTACCTCATGAAGATCCTCACTGAGC  [720]
M  R  L  D  L  A  G  R  G  L  T  D  Y  L  M  K  I  L  T  E    [168]
GTGGTTATTCCTTCACCACCACCGCTGAGAGAGAAATCGTCAGAGACATCAAGGAGAAGC  [780]
R  G  Y  S  F  T  T  T  A  E  R  E  I  V  R  D  I  K  E  K    [188]
TCTGCTACATGCCCCTTG                                            [798]
L  C  Y  M  P  L                                              [194]
```

H.virgA1c

```
AGAGCTGTCTTCCCCTCCATCGTCGGACGTCCAAGACATCAGgtactgtcgtgtattgtt  [60]
 R  A  V  F  P  S  I  V  G  R  P  R  H  Q                     [14]
actttggaaacaaggtttgatcgaacatgtttcccgggtgtatgaagggttcagttctca  [120]
aaaacgttgacaccagatggcattatgtaaaagcagtaataattcacgtgaataattctg  [180]
aaaataacaaatgacttcgtgtttcagGGCGTGATGGTCGGTATGGGTCAGAAAGACAGC  [240]
                            G  V  M  V  G  M  G  Q  K  D  S    [25]
TATGTTGGAGATGAGGCTCAGTCCAAGAGAGGTATCCTCACACTCAAGTACCCCATCGAG  [300]
 Y  V  G  D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P  I  E  [45]
CACGGTATCGTCACCAACTGGGACGACATGGAGAAAATCTGGCATCACACCTTCTACAAC  [360]
 H  G  I  V  T  N  W  D  D  M  E  K  I  W  H  H  T  F  Y  N  [65]
GAGCTCCGTGTTGCACCCGAGGAGCATCCCGTCCTCCTGACAGAGGCTCCCCTCAACCCC  [420]
 E  L  R  V  A  P  E  E  H  P  V  L  L  T  E  A  P  L  N  P  [85]
AAGGCCAACCGTGAAAAGATGACCCAGATCATGTTCGAGACCTTCAACTCTCCAGCTATG  [480]
 K  A  N  R  E  K  M  T  Q  I  M  F  E  T  F  N  S  P  A  M  [105]
TTTGTGGCCATCCAGGCTGTTCTGTCTCTCTACGCTTCTGGTCGTACCACAGGTATTGTG  [540]
 F  V  A  I  Q  A  V  L  S  L  Y  A  S  G  R  T  T  G  I  V  [125]
TTGGACTCTGGTGATGGTGTCACCCACACTGTCCCTATCTATGAAGGTTACGCCCTTCCC  [600]
 L  D  S  G  D  G  V  T  H  T  V  P  I  Y  E  G  Y  A  L  P  [145]
CACGCCATCATGAGGTTGGATCTTGCCGGTCGTGACCTGACTGATTACCTCATGAAGATC  [660]
 H  A  I  M  R  L  D  L  A  G  R  D  L  T  D  Y  L  M  K  I  [165]
CTTACTGAGCGTGGTTACTCCTTTACAACCACCGCTGAGAGAGAAATCGTCAGAGACATC  [720]
 L  T  E  R  G  Y  S  F  T  T  T  A  E  R  E  I  V  R  D  I  [185]
AAAGAGAAGCTCTGCTACATGCCCCTTG                                  [747]
 K  E  K  L  C  Y  M  P  L                                    [194]
```