Concerted Evolution of the rDNA ITS-1 in Two Caddisfly Species

James Bower

A thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Ecology at Massey University, Palmerston North, New Zealand.

February, 1999

Abstract

The ribosomal RNA (rRNA) gene family shows high levels of within-species homogeneity and between-species diversity. This pattern of mutation distribution among gene family members is termed concerted evolution. Molecular drive is a process by which mutations spread through a gene family (homogenization) and through a population (fixation) as a consequence of non-Mendelian DNA transfer within and between chromosomes. The fast evolving internal transcribed spacers (ITS) of the rDNA are considered to be homogenized within species and within genetically isolated populations. However, recent studies question the dogma of homogenization and caution the use of ITS in phylogenetic and population studies.

Two presumed sister species (*Triplectides obsoleta* and *T. dolichos*) were chosen for an investigation of ITS-1 variation. ITS-specific PCR primers were developed that amplified products homologous to a range of other insects. In a pilot study, variation in a single individual *T. obsoleta* was found among eight ITS-1 clones but not in ITS-2. Therefore, ITS-1 was chosen for a survey of the nature of the variation, testing the extent of heterogeneity in these species, the types of coevolutionary changes taking place in this region, and its usefulness as a population genetic marker.

This survey required an hierarchical approach, sampling multiple ITS-1 copies from within and among individuals. 203 ITS-1 clones were screened for conformational differences by single-strand conformation polymorphism (SSCP) analysis of restriction digested fragments. Some ITS-1 variants were present at a higher frequency than others in all 17 individuals tested. All variants derived from independent PCR reactions were sequenced. The type of DNA sequence variation was very similar in both species though *T. dolichos* was

more variable. A majority of the variation appears to have resulted from replication slippage and the remainder by point mutation. Repeat DNA and an A+T bias were found in the ITS-1; a characteristic of cryptically simple non-coding DNA. Extensive ITS-1 heterogeneity was present and parsimony analysis for both species showed the absence of population-specific variation. ITS-1 variants common in some individuals were rare in others, even within the same population. Neither complete homogenization nor fixation of the ITS-1 occurs in these species. This precludes its use as a population genetic marker in these species and cautions against the assumption of homogenization.

Acknowledgements

I am indebted to both my supervisors, Dr. Ian Henderson and Prof. Dave Lambert. I thank Ian for helping me collect and identify caddisflies, and for giving up some of his own. Through Ian I have formed an interest in aquatic insects and gained most of my knowledge of phylogenetics. I am forever thankful to Dave for opening my eyes to the many problems in biology today. Through Dave I have learnt much about the history of biology and much of what I think I know about evolution. I thank both for their patience, especially in light of the dramatic changes this thesis weathered. Most of all I thank Ian and Dave for their friendship. I will truly miss these two rationalists and I only regret being too busy to work with them closer than I did.

In a very practical sense, I am indebted to Peter Ritchie and Dr. Leon Huynen. I arrived in 'the lab' with very little beside a theoretical knowledge of the techniques required for my thesis. Although I was very determined to learn and very serious in manner, this thesis would have become a serious flop without their help. I progressed from listening to everything they said, to a stage of giving them the benefit of the doubt, but more usually disagreeing completely.

Although 'doing science' isolates a scientist from the rest of his species, a surprisingly large number of these people are very willing to help each other. I thank the following for helping me at some time during my thesis: John Ward of the Canterbury Museum, Christchurch, Tania Waghorn of Massey University, Dr. Alice Wells of the Australian Biological Resources Survey, CSIRO, Canberra, Alex Wilson of the Australian National University, Canberra, Dr Karl Kjer of Rutgers University, New Jersey, Dr. Michael Gillings of MacQuarie University, Sydney, Dr. Gejamo Orti of the University of Nebraska, Lincoln, Dr. Dianne Gleeson of Landcare Research, Auckland, Dr. Craig Millar

of the University of Auckland, and Assoc. Prof. Ted Steele of the University of Wollongong.

I single out Craig and Ted for special mention. It was through Craig that I 'discovered' concerted evolution, therefore setting the trajectory that my thesis was to take and indeed for the PhD that I am about to undertake. To Ted I am enormously indebted for making my thesis more rigorous. I was deeply suspicious about how people sample variation in gene families. His advice came at the right time and was well received.

I am also very thankful to Massey University, especially for an MSc scholarship, the J. P. Skipworth scholarship, and support from two rounds of the Ecology section's development fund. All this, despite the fact that New Zealand has the lowest potential of any OECD country to produce knowledge.

I thank Jarn and Young Rob for many discussions as to why we must leave Palmerston North forever and never ever come back. Of course I also thank the prurient boys: Pete, Leon and Alex for 'raising' the standards of conversation in the lab. Leon deserves a special mention because he became my third, albeit unofficial, supervisor (unbeknown to him). Good on ya Leon!

Although isolated from my family, I received a lot 'hang in there James' over the phone. I thank my kin and wish the best upon them. I am especially thankful to my big sister Karen and my little mum, Delwyn. I hope to see you all in Australia sometime.

Living in Palmerston North (the paddock) for the past three years has been very trying for this city boy. A final thankyou goes out to all those who made this stay bearable. I will miss you all and I hope you leave the paddock one day yourselves.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction: Concerted Evolution

1.1 Introduction

Concerted evolution

Eukaryotic (and some archaebacterial) genomes contain substantial numbers of multiple-copy families of genes and non-coding sequences. It has been apparent for some years that many such families exhibit unexpected sequence homogeneity within and between individuals of a species (Dover, 1982; 1986). These repeat DNA families have high levels of within-species homogeneity and between-species heterogeneity irrespective of their copy-number, function and chromosomal distribution. This pattern of sequence variation is known as concerted evolution (Polanco et al., 1998; Zimmer et al., 1980), although it has previously been called horizontal evolution (Brown et al., 1972) and coincidental evolution (Hood et al., 1975). The importance of concerted evolution lies primarily with our understanding of how the great many repeated DNA sequences of the eukaryotic genome evolve.

Concerted evolution was first demonstrated in the ribosomal RNA gene family of two species of African toad, *Xenopus*, by Brown et al. (1972). This, and subsequent studies (e.g., Coen et al., 1982b; Williams et al., 1985; Bowen & Dover, 1995) demonstrated that the individual repeat members of a multigene family evolve in concert rather than independently. On the surface this suggests a process of parallel evolution for all members of a gene family. However, Coen et al. (1982b) state that, "concerted evolution is an observation of family

homogeneity that is an outcome of processes in the genome that are able to progressively fix a variant member in the family." Through genetic interactions among its members, multiple-copy families of genes and non-coding sequences evolve in a concerted fashion, seemingly as a single unit or locus.

Concerted evolution challenges both conventional neutralist and selectionist evolutionary explanations. Neither the data nor their implications can be squeezed into the classical mould of a two-party (selection versus drift) state (Dover, 1986). Under genetic drift and selection, individual gene family members are expected to evolve independently by mutation, gradually diverging from one another over time. If each member of a family is free to accumulate independent mutations over time then the coefficient of identity between any two members within a species would equal that of any two members chosen at random from two different species (Dover, 1982; Ridley, 1993). However, mutations do not occur independently in the different members of a family in neither the gene-coding nor, more unexpectedly, non-coding regions of gene families and the coefficient of identity is higher within than between species.

Negative (purifying) selection is thought to impede the spread of new mutations in the functionally important (gene-coding) regions of a gene family, because of their potentially negative effect on an individual's fitness (Wheeler & Honeycutt, 1988). This argument though does not explain within-species homogeneity found in the non-coding regions of a gene family. Intraspecific homogeneity is maintained not by selective constraints but by processes which can spread mutations in a repeat unit 'horizontally' to other members of the family (Li, 1997).

DNA turnover processes and molecular drive

Molecular drive is the process where mutations spread through a repeat family within an individual (homogenization) and through a sexually reproducing population (fixation) as a consequence of non-Mendelian DNA transfer within and between chromosomes. Homogenization and fixation are inextricably linked because chromosomes are continually being shuffled and assorted into new combinations at each generation through sex (Dover, 1986). The generative processes involved include; gene conversion (Eldeman & Gally, 1970; Nagylaki & Petes, 1982), unequal crossing-over (UCO) (Smith, 1973; Ohta, 1976), replication slippage (slipped-strand mispairing), transposition and RNA-mediated exchange (Dover, 1982; 1986). Of these, gene conversion, UCO and replication slippage are considered to be the most important. For recent reviews see Elder and Turner (1995) and Li (1997).

Gene conversion is a non-reciprocal recombination process where one allele is converted to the other (Li, 1997). It is thought gene conversion results from DNA repair after invasion of a double helix of one allele by a single strand of the helix of another, immediately after chromosome replication. Gene conversion changes allele frequencies and any bias in the DNA repair leads to the progressive and accelerated fixation of an allele (variant) (Dover, 1982; Li & Graur, 1991). UCO is a reciprocal recombination process that creates a sequence duplication in one chromatid or chromosome and a corresponding deletion in the other (Li & Graur, 1991). It is an interhelical event that can only take place during chromosome alignment in cell division (Levinson & Gutman, 1987). Continual UCO has the effect of expanding and contracting the copy number of a family and can lead to the stochastic fixation of any one variant. Replication slippage involves local denaturation and displacement of DNA strands (of a single DNA duplex) followed by base mispairing at a site containing a tandem repeat sequence. The consequence of mispairing, when followed by replication or DNA repair, is an insertion or deletion of one or several of the tandem

repeats (Levinson & Gutman, 1987). Slipped-strand mispairing is perhaps a better term because it also occurs prior to cell division, in non-replicating DNA.

These processes of DNA turnover are analogous to the turnover of new for old banknotes (Dover, 1982). They cause a continual non-Mendelian gain or loss of genetic material during mitosis and meiosis, involving 'old' variation (nucleotide sequences already present in the genome) and/or 'new' variation, such as point mutation. Although it defies the paradigm that cell division is the necessary event for point mutation and DNA turnover, both have been demonstrated in resting cells (Bridges, 1997). DNA turnover is a larger mutational event but does not involve the creation of new variation, just changes in its frequency. DNA turnover spreads variation at a significantly higher rate than can be generated by point mutations alone (Hoelzel & Dover, 1991). The balance between the spreading consequences of DNA turnover and the accumulation of point mutations is manifest as a degree of homogeneity found among the multiple copies of repeat DNA sequences.

C-value paradox and selfish DNA

The lack of correspondence between the amount of DNA in the haploid genome (C value) and the amount of gene-coding DNA present has become known as the C-value paradox. Solving the C-value paradox and accounting for the repeated structure of the eukaryotic genome requires an evolutionary mechanism for the long-term maintenance of vast quantities of nongenic (non-coding), seemingly superfluous DNA (Li & Graur, 1991). This functionless non-coding DNA, often called 'junk DNA' (Ohno, 1972) or 'selfish DNA' (Orgel & Crick, 1980; Doolittle & Sapienza, 1980), has the property of forming additional copies of itself while making no contribution to the phenotype. This lack of phenotypic expression means that normal phenotypic (Darwinian) selection cannot explain its continued presence, as junk DNA contributes little or nothing to an organism's fitness.

Dover and Doolittle (1980) suggested there are constraints that impose a type of selection on these sequences, not necessarily as a result of their phenotypic effects but as a consequence of the molecular processes of replication and recombination. The generative processes of molecular drive, that is, DNA turnover, provide an explanation for the presence of repeated sequences (regardless of their function) and hence much of the structure of the eukaryotic genome.

1.2 The nuclear ribosomal RNA gene family

Structure and function

The ribosomal RNA (rRNA) genes, and their associated spacer regions, are collectively called ribosomal DNA (rDNA) (Hillis & Dixon, 1991). They exhibit the pattern of concerted evolution typical of other gene families. Ribosomal RNA has a primary functional role in most if not all stages of protein synthesis (Dahlberg, 1989). The rDNA is composed of multiple copies of a repeating unit. This includes the highly conserved gene-coding rRNA molecules which are synthesized by RNA polymerase I, before being processed to yield mature rRNA, and several fast evolving non-coding spacers (see figure 1.1). These rRNAs combine with ribosomal proteins to form ribosomes, the organelles that direct protein synthesis from messenger RNA. The entire rDNA intergenic spacer (IGS) acts as an enhancer for ribosomal transcription by polymerase I (Mougey et al., 1996), and the other spacers (ETS, ITS-1 and ITS-2) are necessary for pre-rRNA processing (Musters et al., 1990).

The rDNA of a eukaryote nuclear genome is typically clustered in one or a few tandem arrays per haploid chromosome set. The length of the transcribed unit for most eukaryotes is about 8 kb, while the length of the intervening IGS can vary considerably, e.g., from 3.5 to >20 kb in a hybrid *Neotyphodium* fungi (Ganley & Scott, 1998) and 7.2 to 20.5 kb in the greenbug, *Schizaphis graminum*

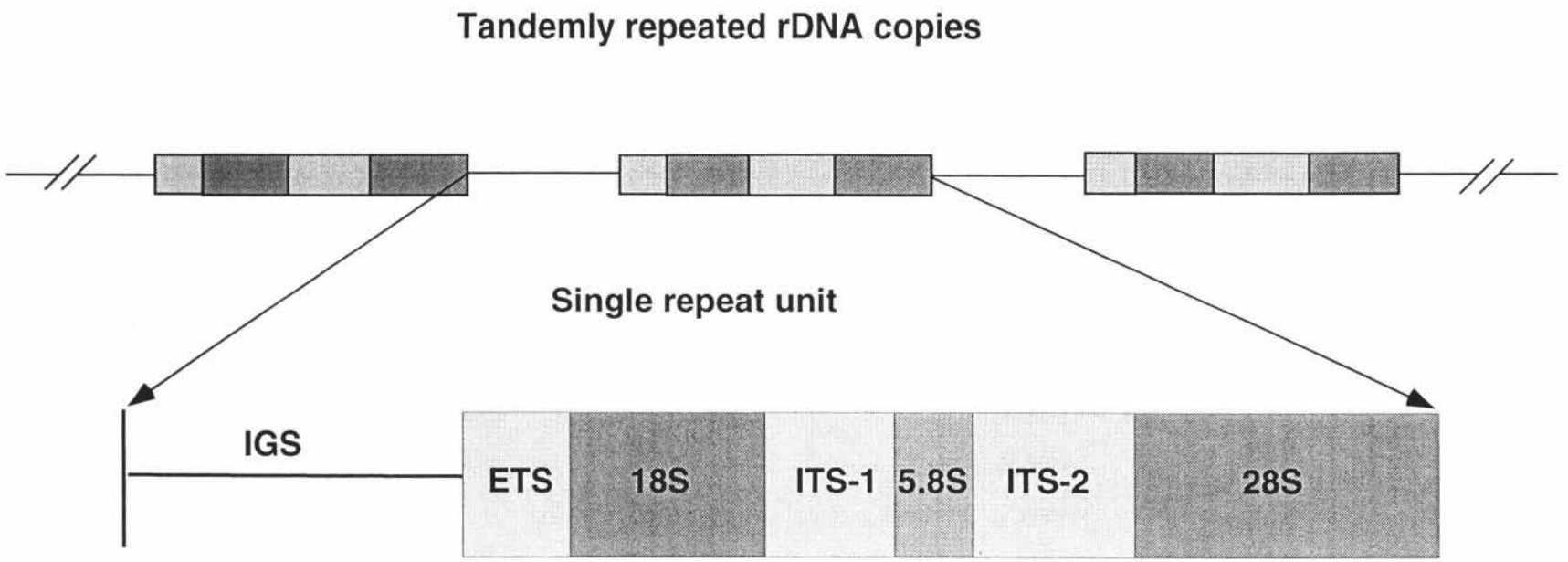

Figure 1.1 The ribosomal RNA gene family is clustered into a group of tandemly repeated units (copies) in a head-to-tail fashion. Three of these are pictured above and a single unit below it. A single unit contains both gene-coding and non-coding regions: IGS = intergenic spacer, ETS = external transcribed spacer, 18S, 5.8S and 28S = rRNA genes, and ITS-1 and ITS-2 = the first and second transcribed spacers, respectively.

(Shufran et al., 1997). The number of gene copies also varies considerably among taxa. Within insects it is common to find several hundred copies present per haploid genome and several thousand in plants (Long & Dawid, 1980).

Variable regions

The terms 'variable region' and 'expansion segment' are both used to describe regions varying between species in both length and content, which lie between regions of conserved sequences (Wesson et al., 1992). In general, the conserved regions of the rDNA are gene-coding and share greater amounts of sequence homology between species than do the variable regions. Variable regions include the non-coding spacers and the expansion segments of 18S and 28S rDNA. Expansion segments are separated by conserved 'core' regions, and may be functionless structures which are tolerated only because they do not interfere with ribosome function (Gerbi, 1985).

The redundancy of rDNA spacers

The rates of change within the different regions of the rDNA unit are thought to differ as a consequence of functional constraints. Non-coding spacers such as the internal transcribed spacer (ITS) of the rDNA are generally considered to be under low functional constraint (Schlotterer et al., 1994), and therefore free to accumulate mutations. These spacers could be viewed as 'junk DNA', perhaps redundant entities hitch-hiking their way in the genome aboard the gene-coding rDNA. However, studies involving DNA sequence comparisons and secondary structure analyses have shown that variation in these regions is non-random, perhaps indicating some functional constraint (see Schlotterer et al., 1994; Wesson et al., 1992; Linares et al., 1994). These variable regions provide an opportunity to study the types of coevolutionary changes taking place in gene families, and the dynamics of their establishment (Tautz et al., 1988).

1.3 "Variation is Evolution"

The idea of variation as random

> "Variation, whatever may be its cause, and however it may be limited, is the essential phenomenon of evolution. Variation, in fact, is evolution. The readiest way, then, of solving the problem of evolution is to study the facts of variation" (Bateson, 1894).

It is commonly thought that the action of an external force (natural selection) and the random fluctuation of gene frequencies (genetic drift) segregates variation in both space and time. The dynamics of natural selection and genetic drift are modelled on the premise that mutations are unitary and passive events (Dover, 1982). Without the sifting mechanism of selection acting on what is assumed to be random variation, we reason there would be nothing but incoherent disorder (Kauffman, 1995). This 'sifting mechanism' is thought to create order from what would otherwise be an unstructured world. Unfortunately, the neo-Darwinian concept of random variation carries with it the major fallacy that everything conceivable is possible (Ho & Saunders, 1984). It is important to understand just how entrenched the idea of chance and random variation is in biology.

> "We say that these events [mutations] are accidental, due to chance. And since they constitute the *only* possible source of modifications in the genetic text, itself the *sole* repository of the organism's hereditary structures, it necessarily follows that chance *alone* is at the source of every innovation, of all creation in the biosphere. Pure chance, absolutely free but blind, at the very root of the stupendous edifice of evolution: this central concept of modern biology is no longer one among other possible or even conceivable hypotheses. It is today the *sole* conceivable hypothesis, the only one compatible with observed and tested fact. And nothing warrants the supposition (or the hope) that conceptions about this should, or even could, be revised" (Monod, 1972). [emphasis in original.]

Chance; the happening of events *without any cause* that can be seen or understood (Hornby, 1970). Monod's dogmatic quote is the expression of a neo-Darwinian view that evolution, that is, the origin and maintenance of variation is solely a chance driven process, immune to any causal explanation. Chance and randomness are concepts regularly used in neo-Darwinism to cover ignorance (Lima-de-Faria, 1988). Although contingency plays an important role, our understanding of evolution need not be restricted to probabilistic models. A great deal of variation is non-randomly distributed in time and space, and can only be understood in the light of real demonstrable biological processes (generative processes) which order and segregate genetic variation.

Cairns et al. (1988) were able to generate gene mutations in cultures of bacteria in a seemingly directed fashion. The main purpose of their paper was to show how insecure our belief is in the spontaneity (randomness) of most mutations, it seems to be a doctrine that has never been properly put to the test (Cairns et al., 1988). The phenomenon of compensatory mutations that preserve base pairing in RNA molecules provides another example of the existence of a process that can generate non-random mutation (variation). Given that an RNA sequence can be copied back into DNA, a proportion of DNA mutations must also arise through natural error-prone copying involving RNA intermediates rather than in a haphazard, mysterious fashion (Steele et al., 1998).

1.4 Interpreting the variation

A problem of explanation

It is the task of science to discover which hypothetical or imagined processes are not imaginary but real; or, to put it the other way, to discover what the processes are, and to produce an adequate account of them (Bhaskar, 1978). From an idealist (e.g. neo-Darwinian) perspective, there is an inclination that building a plausible narrative explanation into a model, which is then

empirically tested, is a sufficient explanation. Lambert (1995) contends "by this we create a kind of 'taxonomy of knowledge,' that is, a system for the organization of what we know [and in] using this procedure we can potentially explain any example."

From a realist perspective a distinction is made between observable patterns and the generative processes responsible for their production (Webster, 1984). The failure of neo-Darwinism to make this distinction perhaps brought about the necessity for a new evolutionary theory to understand the intraspecific homogeneity of repeat DNA families. In the theory of molecular drive, as outlined by Dover (1982) and Ohta and Dover (1984), the distinction is made between the observed pattern of concerted evolution and the DNA turnover processes that generate repeat sequence homogeneity.

Processes and models

Genetics, and in turn, population genetics, forms the conceptual basis of neo-Darwinism. Through population genetics, neo-Darwinism claims a monopoly in explaining the processes of evolution (Lovtrup, 1976, 1977). Despite the sophistication of the mathematics of natural selection and genetic drift, there are relatively few experimental proofs of the genetic processes inherent in these models (Dover, 1982). Underlying this may be a confusion between the value of a model of evolution over generative processes. From an idealist perspective however, it is only necessary to establish that such processes might exist. For the idealist, a process is imaginary, for the realist it may be real, and come to be established as such (Bhaskar, 1978).

It is considered that population genetic models greatly sharpen our understanding of the population dynamics of concerted evolution (Li, 1997). The main aim of these models is to obtain probabilities of identity by descent (i.e., genetic identities) under a variety of assumptions and constraints.

However, the genealogies produced by such models restrict our interpretation of evolution to that of a historical narrative. This is further restricted because the evolutionary genetics of molecular drive, like that of Mendelian populations under selection and drift, can be modelled only in general terms (Ohta & Dover, 1984). The models themselves do not provide an adequate rationale for the use of laws to explain phenomena in open systems or the experimental establishment of that knowledge in the first place (Bhaskar, 1978). This can only be achieved by investigation of the underlying generative processes.

Homology

Phylogenies for systematics and population studies require a model of evolution that contains an inherent assumption of identity by descent. A common ancestor and independent character change are necessary conditions for reconstructing phylogenies. Concerted evolution is considered advantageous for phylogeny reconstruction as it produces almost completely homogenized gene families (Sanderson & Doyle, 1992). However, multigene families arise by gene duplication from a single 'ancestral gene', and this duplication creates a set of unique problems regarding the accuracy and reconstruction of organismal phylogenies.

Homology in gene families is complicated by gene duplication events. This necessitates the elaboration of the concept of homology to include orthology (derivation by speciation; shared by descent) and paralogy (derivation by duplication) (Patterson, 1988). If one is interested in reconstructing the phylogenetic history of taxa by inferring relationships among genes contained in those taxa, then it is usually necessary to examine orthologous genes (Hillis, 1994). Paralogy is not considered a problem with multigene families because the mechanisms of DNA turnover are thought to homogenize all members within a family, such that each member has a nearly identical sequence. The whole family is treated as a single character and the non-independence of the

individual members is nullified by concerted evolution.

The dogma of homogenization

Williams et al. (1988), and more recently, Rich et al. (1997) suggest that the dogma of rDNA gene family homogenization is overstated. The assumption of Hillis and Davis (1988) that a high level of gene family homogenization within species is reached in only short periods of evolutionary time, may have tempted workers to under-sample the amount of variation present. Recent studies however, (Wesson et al., 1992; Volger & DeSalle, 1994; Tang et al., 1996; Rich et al., 1997) have shown that rDNA homogeneity can no longer be assumed. This is especially important for population studies, where intra-individual variation can be as high as variation in the entire species (e.g., Volger & DeSalle, 1994).

1.5 A study of ITS-1 variation in two caddisflies

Caddisflies

The Trichoptera (caddisflies) are a holometabolous order of insects, with egg, larval, and pupal stages generally living in freshwater habitats and with adults in terrestrial habitats (Morse, 1997). Caddisflies are ubiquitous in both still and running freshwater environments. They are of considerable ecological importance and form a major part of freshwater communities. They are also used as bio-indicators in water quality monitoring because many caddisfly species tolerate only a narrow range of abiotic conditions.

Triplectides

The stick caddis; *T. obsoleta* (McLachan, 1862), *T. dolichos* (McFarlane & Cowie, 1981), and *T. cephalotes* (Walker, 1852) constitute the three endemic *Triplectides* species of New Zealand. They are common aquatic insects and are found throughout New Zealand. The common name refers to the habit of larvae to use a hollowed-out twig as a portable case. The larvae eat detritus (mainly leaves

and wood) and appear as slowly moving sticks in freshwater habitats. These species have distinct larval habitat preferences: *T. obsoleta* prefers slow flowing waters of lowland streams and rivers, *T. dolichos* prefers pools in small mountain streams (headwaters) and seepages, while *T. cephalotes* prefers lakes and backwaters of larger rivers.

T. obsoleta and *T. dolichos* are presumed sister species. Both species are morphologically very similar, a fact that only lead to the 'discovery' of *T. dolichos* in 1981. Both are large insects as adults (~2 cm in length) that can be easily recognised in the field among a mass of other nocturnal insects that are attracted to UV-light traps. Trap records of these species provide good data of their geographic distribution. Although common, there are several indications that their dispersal away from water is restricted and hence, they may be likely to have genetically structured populations (Collier & Smith, 1996; Jackson & Resh, 1992).

Panmixus

If the dispersal ability of a species is extreme, a state of reproductive panmixus may result and genetic variation would not be restricted geographically. *T. dolichos*, in particular, is not expected to be panmictic because its habitat (headwaters and seepages) is disconnected spatially, whereas the river habitat of *T. obsoleta* is connected, at least to the level of a drainage basin. However, the Cook Strait that separates the North and South Islands of New Zealand is likely to be a barrier to dispersal for both species, as it is for many presumptive sister species pairs of caddisflies that are isolated from each other by the strait (Henderson, pers. comm).

The results of several studies (Jackson & Resh, 1992; Hughes et al., 1995, 1996; Bunn & Hughes, 1997) indicate that many aquatic invertebrate species do not represent single panmictic populations throughout their entire range, but are

subdivided regionally and sometimes even locally into a number of genetically distinct populations. Analysis of the genetic structure of larval populations of some caddisfly species suggests that in-stream movement is also limited to a small spatial scale (Bunn & Hughes, 1997). Limited inland dispersal may be due, impart, to the reproductive behaviour of individuals. Freshly emerged female caddisflies frequently fly inland following mating where they may rest until eggs mature (Ross, 1944), though after egg maturation, females always return to the water to oviposit (Kovats et al., 1996).

Collier and Smith (1996) used several kinds of traps to investigate distances travelled by adult aquatic insects into riparian zones. In their study *T. obsoleta* were caught up to 30 m and *T.* dolichos up to 80 m from the edge of three different streams. The distance measure for *T. dolichos* may be confounded by the presence of other suitable habitat (small seepages) away from these streams. These direct estimates suggest both species do not disperse widely despite having large wings. The possession of large wings does not necessarily mean that their function is for long distance dispersal. For example, Liebherr (1988) found no correlation with the degree of flight-wing development and dispersal ability among five beetle species (*Platynus spp.*). One fully winged species exhibited genetic heterogeneity of the same order as a vestigally winged species (Liebherr, 1988).

ITS-1 variation in Triplectides

Many studies of ITS-1 variation in insects have been carried out on flies (Diptera) (e.g. fruit flies: Strachan et al., 1985; mosquitoes: Wesson et al., 1992; and black flies: Tang et al., 1996), which are close relatives to caddisflies. As a result of these studies, there are now many rDNA sequences published in GenBank which enable comparisons of conserved gene-coding regions for identifying PCR products and designing PCR primers. The position of the spacer region between the rRNA molecules can also be estimated by

comparison to the rRNA secondary structure analysis of the fruit fly, *Drosophila melanogaster* (Hancock et al., 1988).

In this study, *T. obsoleta* and *T. dolichos* were assayed for ITS-1 variation in the rRNA gene family at the intra-individual and intra-population level. The purpose of the assay was to determine the extent of homogeneity in the spacer molecule, the frequency of ITS-1 variants within individuals, and the presence of any population-specific variation. This study therefore, constitutes a further test of the assumption of gene family homogenization and a test of a prediction of molecular drive, that populations should be evolving in a cohesive manner.

Chapter 2

Development of ITS-Specific PCR Primers for Caddisflies (Trichoptera)

2.1 Introduction

Current use of ITS

The internal transcribed spacers (ITS-1 and ITS-2) of the ribosomal RNA (rRNA) gene family have become popular genetic markers for population and phylogenetic studies (e.g., Volger & DeSalle, 1994; Miller et al., 1996; Coleman & Mai, 1997). The presence of a set of non-Mendelian inheritance mechanisms in gene families, the action of these now collectively called molecular drive, challenges our understanding of how this variation is generated. The variable nature of these spacers makes them key elements for investigating coevolutionary changes in the rRNA gene family (see Schlotterer et al., 1994; Polanco et al., 1998).

Properties of ITS

ITS-1 and ITS-2 (collectively ITS) lie within the rRNA gene-coding subunits of the tandemly repeated rRNA gene cluster (see chapter 1; figure 1.1). These spacers contain repeated DNA sequences which change more rapidly than the adjacent gene-coding regions. Therefore, the ITS is not considered to be of great functional importance (Long & Dawid, 1980; Hillis & Dixon, 1991; Elder & Turner, 1995). The rates of nucleotide change vary widely across the ITS region, the size of the region varies widely among taxa, and there is no obvious sequence homology between more distantly related taxonomic groups (Hillis et

al., 1996). However, there is a degree of ITS secondary structure conservation within groups of closely related taxa and at the species level, the ITS is considered to be homogenized for sequence variation (Coen et al., 1982a, 1982b; Wesson et al., 1992; Fritz et al., 1994).

PCR amplification of the ITS

Amplification of the two internal transcribed spacers via the polymerase chain reaction can be routinely carried out using universal primers which bind to conserved flanking regions of the 18S, 5.8S and 28S genes (Hillis & Dixon, 1991). However, these primers may amplify contaminant rDNA from other species. Therefore it is desirable to design primers that are more specific to the taxa of interest, thereby reducing the problem of contamination.

ITS-specific PCR primers for Triplectides

In this study, ITS-specific PCR primers were designed for the caddisfly genus *Triplectides*, in particular, to the three New Zealand species: *T. obsoleta*, *T. dolichos* and *T. cephalotes* (see chapter 1). This required the use of existing primers located outside the region of interest. Primer design also required DNA sequence data of the 18S, 5.8S and 28S genes from related taxa within the Trichoptera and a range of outgroups. Sequence data from these highly conserved gene-coding regions was obtained by a combination of direct sequencing of PCR amplified products and cloning, and from published sequences in the GenBank database. Multiple clones were constructed from a single *T. obsoleta* sample (see chapter 3; figure 3.1, individual F1OPN) to obtain 5.8S sequence data. These clones were also used in a pilot study to test for intra-individual variation (a preliminary test for homogenization) in both spacer molecules. The ITS-specific primers were then tested on a wide taxonomic range of caddisflies.

2.2 Materials and Methods

Taxa used in this study

Samples were caught using a UV light trap. The light trap was an 8W Sylvania Blacklight Blue F8T5/BLB fluorescent tube, run from a 12V battery, and laid across a white plastic tray containing several centimetres of water and a few drops of household detergent into which the insect fell (Ward et al., 1996). All samples were preserved in 70% ethanol until the time of DNA extraction. Six species within the subfamily Triplectinae (*Triplectidina mosleyi, Triplectides cephalotes, T. obsoleta, T. dolichos, T. similis,* and *T. australis*) and a single species from the subfamily Leptocerinae (*Oecetis unicolor*) were used to represent the family Leptoceridae. Several species from a wide taxonomic range of Trichoptera and the cosmopolitan tobacco bud worm (*Heliothis armigera*: Noctuidae: Lepidoptera) were also used to test the versatility of the ITS primers. The trichopteran species are shown in table 2.1

Superfamily	Family	Species
Rhyacophiloidea	Hydrobiosidae	*Hydrobiosis spatulata*
	Hydroptilidae	*Oxyethira albiceps*
Hydropsychoidea	Philopotamidae	*Hydrobiosella mixta*
	Polycentropodidae	*Polyplectropus aurifusca*
	Hydropsychidae	*Orthopsyche fimbriata*
Limnephiloidea	Oeconesidae	*Oeconesus maori*
	Leptoceridae	7 species (see text)
	Chathamiidae	*Philanisus plebeius*
	Conoesucidae	*Olinga feredayi*

Table 2.1 Trichopteran species used to test the versatility of the ITS-specific primers. Superfamily and family level placement is according to Neboiss (1991).

DNA extraction

DNA was extracted from a minimum of two individuals per species. The protocol used for the extraction of genomic DNA follows that of Sambrook et al. (1989) with slight modifications. The insects were prepared by removing the wings, head and thorax, using tweezers and a scalpel that were flamed between samples with 95% ethanol. Samples were dried free of excess ethanol by blotting on lint free tissue. These were then homogenized with a pipette tip in an Eppendorf tube containing 500µl CTAB extraction buffer (100mM Tris pH 8.0, 20mM EDTA pH 8.0, 1.4M NaCl, 2% w/v CTAB), 4µl β-mercaptoethanol, and 10µl Proteinase K (20mg/ml). Digestion then followed by incubation in a rotisserie overnight at 55°C.

Proteins were removed from the solution by the addition of an equal volume of phenol (500µl) to each tube. These were rocked for 30 min then centrifuged in a microcentrifuge for 10 min at 15,500 g. The aqueous phase (containing the DNA) and an equal volume of phenol:chloroform:isoamyl alcohol (PCI, 25:24:1) were added to a new tube. The rocking and centrifugation step was repeated then the interface and organic phase were discarded. An equal volume of chloroform:isoamyl alcohol (CI, 24:1) was added then the rocking and centrifugation step was repeated. Finally, both the interface and organic phase were discarded.

The DNA was precipitated by adding 1 ml of 100% ethanol to each tube and leaving at -20°C overnight. These were then centrifuged for 10 min at 15,500 g. The supernatant was discarded and the pelleted DNA was washed with 0.5ml of 70% ethanol. This step was repeated three times. The ethanol was removed between steps with a fine pipette. The pellet was dried under vacuum for 5 min and resuspended in 30µl of dH_2O. One microlitre of RNase (0.5µg) was added to digest the RNA present in each sample.

Primers and PCR amplification

A range of PCR primers were utilised to design the more specific ITS primers for *Triplectides*. These included primers designed by Schlotterer et al. (1994); CS249 and CS250, those recommended by Karl Kjer of Rutgers University, New Jersey (pers. comm); F-18S-3′ and R-28S-5′, and the universal primers designed by Hillis and Dixon (1991); 18d and 28z. The DNA sequence of each of these primers is shown in table 2.2. Primers designed in this study are shown in table 2.3. The relative position and orientation of all primers are shown in relation to the spacers and adjacent gene-coding regions on a primer map (figure 2.1).

Name	Oligonucleotide sequence
F-18S-3′	5′-CAG TAA GTG CGG GTC ATA AGC-3′
R-28S-5′	5′-CCT TAG ATG GAG TTT ACC ACC-3′
CS249 (18S)	5′-TCG TAA CAA GGT TTC CG-3′
CS250 (28S)	5′-GTT GGT TTC TTT TCC TC-3′
18d	5′-CAC ACC GCC CGT CGC TAC TAC CGA TTG-3′
28z	5′-AGA CTC CTT GGT CCG TGT TTC AAG AC-3′

Table 2.2 PCR primers used in the preliminary design of ITS-specific primers.

Name	Oligonucleotide sequence
FITS-1_18S	5′-TCC GAC GCG TCG GAA AGT TTG ACC G-3′
RITS-1_5.8S	5′-CGC AAT GTG CGT TCG AAA TGT C-3′
FITS-2_5.8S	5′-GTG GGT CGA TGA AGA ACG CAG C-3′
RITS-2_28S	5′-CCT CAC GGT ACT TGT TCG CTA TCG G-3′
RITS-2_core	5′-CCG TTT CGC TCG CCG CTA CTC G-3′

Table 2.3 ITS-specific PCR primers designed for *Triplectides*.

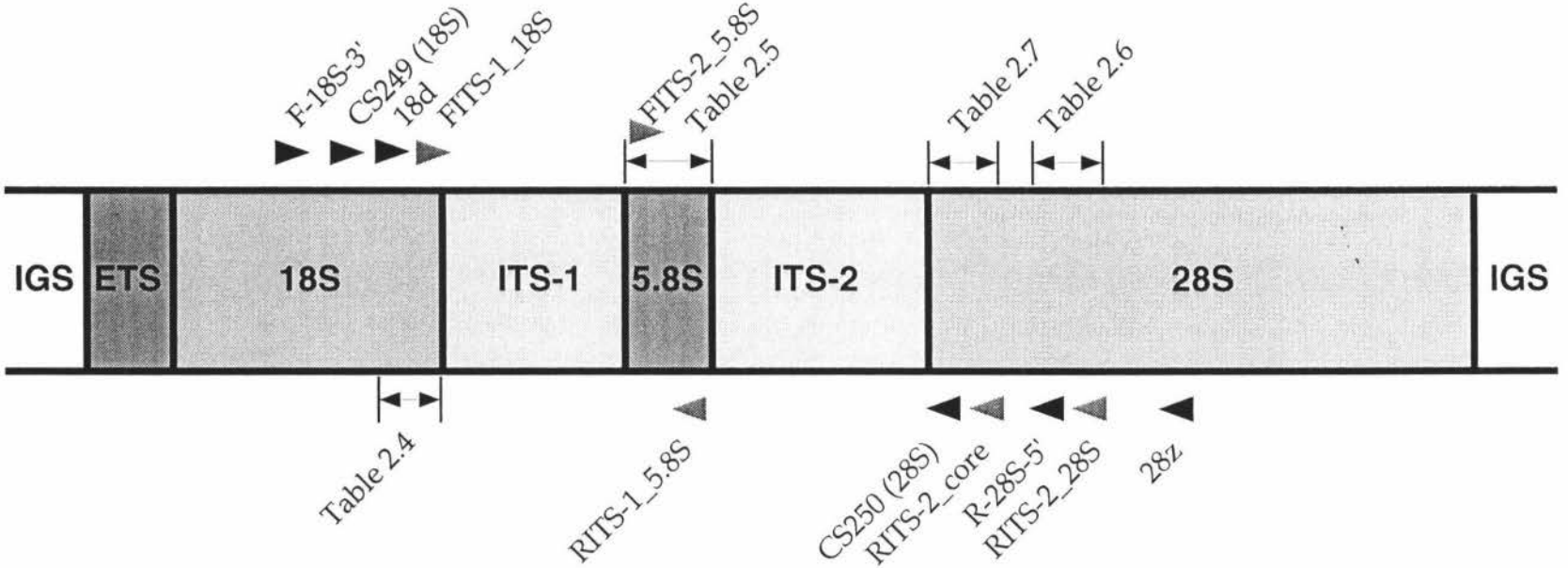

Figure 2.1 Primer map showing relative position and orientation of PCR primers along a single repeat of the rDNA (not drawn to relative scale). Dark arrows ► denote primers used in the preliminary design of the ITS-specific primers (see table 2.2) Light arrows ► denote the ITS-specific primers (see table 2.3). Regions of multiple sequence alignments referred to in tables are denoted by |◄ ►|.

Polymerase chain reaction (PCR) amplification of double-stranded DNA was performed, with slight modification to Kocher et al. (1989), in 25µl volumes containing 67mM Tris (pH 8.8), 1.5mM $MgCl_2$, 16.6mM $(NH_4)_2SO_4$, 10mM β-mercaptoethanol, 0.25µl glycerol, each dNTP at 200µM, each primer at 0.6µM, ~5ng genomic DNA, and 0.5 units of *AmpliTaq*® DNA polymerase (Perkin-Elmer, Branchburg, NJ). The typical thermocycling profile consisted of an initial denaturation cycle at 94°C for 2 min, followed by 30 cycles at 94°C for 1 min, 67°C for 1 min, and 72°C for 2 min, then a final extension at 72°C for 4 min.

These conditions were modified for the ITS-specific primers. Reaction volumes were reduced to 10µl and changes were made to the following: glycerol (0.1µl), *AmpliTaq*® DNA polymerase (0.25 units), and each primer (0.5µM). The thermocycling conditions were one cycle at 94°C for 2 min, followed by 20-25 cycles at 94°C for 1 min, 60°C for 1 min, and 72°C for 2 min, then one cycle at 72°C for 4 min.

Purification of amplified DNA

Double-stranded PCR products were purified prior to DNA sequencing using a QIAquick™ PCR purification kit (QIAGEN, Hilden, Germany) according to the manufacturers instructions. The DNA was eluted into 20-30µl EB buffer (10mM TrisCl, pH 8.5) and quantified visually using ethidium bromide fluorescence. For each sample 2µl of purified PCR product was mixed with 2µl of a 6X loading buffer containing bromophenol blue (10% w/v bromophenol blue, 15% w/v Ficoll, 1 X TA buffer). Samples were electrophoresed on 1-2% agarose gel containing ethidium bromide (5µg) in a minigel chamber (Horizon®58 GibcoBRL, Gaithersburg, MD) with 1X TA buffer (10mM Tris, 11.4mM acetic acid). DNA standards (1 Kb plus DNA ladder and λ *Hin*d III fragments; GibcoBRL) were run alongside the samples for sizing and/or quantifying the DNA. UV trans-illuminated products were photographed using polaroid film

or printed using a gel documentation system (Insta Doc™ System, Bio-Rad, Hercules, CA).

Cloning PCR products

PCR products were purified and cloned using the pGEM®-T vector system (Promega, Madison, WI) according to the manufacturers instructions. PCR products were ligated overnight at room temperature (10-12°C) in 10µl volumes containing T4 DNA ligase (3 units), 10X T4 DNA ligase buffer (GibcoBRL), 50ng pGEM®-T vector, 20-30ng PCR product, and dH_2O.

Competent *E. coli* JM109 cells (Promega) were transformed with the ligated DNA. Two microlitres of each ligation reaction was transferred to a sterile 1.5ml Eppendorf tube and placed on ice. 50µl of competent cells were added and left on ice for 20 min. This mixture was heat shocked in a water bath at 42°C for 50 seconds and returned to ice for 2 min. 950µl of LB broth (percentages are w/v: 1% peptone, 1% NaCl, 0.5% yeast extract, pH brought to 7.2 with NaOH, then autoclaved) was added and each tube incubated at 37°C for 1.5 hours with gentle shaking. The cells were pelleted by centrifugation at 5000 rpm for 5 min. 850µl of the supernatant was discarded and the cells resuspended in the remaining 100µl. The cells were spread onto agar plates (percentages are w/v: 1% peptone, 1% NaCl, 0.5% yeast extract, 2.2% Davis agar) containing ampicillin (100µg/ml), BCIG (20µg/ml) and IPTG (0.4µg/ml) and were grown overnight at 37°C.

Recombinant white colonies were screened by PCR amplification with the primers 18d and 28z to determine the presence and size of the DNA. A small sample from each colony was added to a sterile Eppendorf tube containing 20µl dH_2O. Samples were boiled at 94°C for 5 min, then cooled on ice before centrifugation at 2,500 g for 2 min. Two microlitres of this crude plasmid

extraction was then amplified as outlined above in 'primers and PCR amplification'. The PCR products were then size fractionated in 2% agarose gels.

Plasmid DNA preparation

Positive colonies with the correct insert size (~2100 bp) were grown overnight at 37°C in 2ml LB broth containing ampicillin (100µg/ml). DNA was extracted using the QIAprep miniprep kit (QIAGEN) as instructed by the manufacturers. DNA was eluted in 100µl EB buffer (10mM TrisCl, pH 8.5) and stored at -20°C.

Cycle sequencing and purification of extension products

All ABI PRISM® cycle sequencing reactions were carried out in 10µl volumes containing; 4µl BigDye™ termination mix (Perkin-Elmer), 3.2µl of either a forward or reverse primer (10µM/µl), and 2.8µl amplified DNA (20-60ng) overlaid with a drop of mineral oil. In all reactions a Hybaid Omn-E (Hybaid Ltd, Middlesex, UK) thermal cycler was used. The thermocycling conditions consisted of an initial 10 min pre-heating step then 25 cycles at 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 min. The extension products were purified by ethanol precipitation. Each sample (minus the oil) was placed into a new Eppendorf tube containing 25µl of chilled 100% ethanol and 1µl of 3M Sodium Acetate (pH 4.6). These were incubated on ice for 5 min then centrifuged at 15,500 g for 10 min to pellet the DNA. The ethanol was carefully removed with a fine pipette tip then 250µl of 70% ethanol was added and the sample centrifuged again at 15,500 g for 1 min. The ethanol was removed and the pellet was dried under vacuum for 5 min before finally being stored at -20°C.

Sequencing of DNA template

An ABI 377 (Perkin-Elmer) automated DNA sequencing system (MuSeq DNA sequencing facility, Massey University) was used to sequence PCR amplified and cloned rDNA products. Either both strands of the rDNA (forward and reverse) or two replicates of one strand were sequenced from each sample.

Computer programs

OLIGO 4.0-s and Amplify 1.2 (Madison, WI) were used as aids to design the primers. The rDNA sequence data was edited and/or aligned using EditView 1.1.1 (Perkin-Elmer), Amplify 1.2 (Madison, WI), PAUP 3.1.1, CLUSTAL W(1.4) (Higgins & Sharp, 1988), and MacClade 3.06 (Maddison & Maddison, 1992).

2.3 Results

Primer design in conserved rDNA regions

Both the Schlotterer et al. (1994) and Kjer primers consistently produced multiple PCR products despite trying a range of PCR conditions. However, the universal primers 18d and 28z designed by Hillis and Dixon (1991) worked better, yielding single amplification products that were easily sequenced. Sequences of the 18S and 28S gene-coding regions were obtained by direct sequencing of PCR products from *T. obsoleta*, *T. dolichos*, *Triplectidina mosleyi* and *Heliothis armigera*. Unfortunately, no sequences could be obtained from *T. cephalotes*, nor a 28S sequence from *H. armigera*. Figure 2.2 shows the amplified rDNA fragments from these species. Eight clones from a single individual *T. obsoleta* were also sequenced, using the 18d and 28z primers to yield rDNA sequence from 18S, ITS-1 and 5.8S. Figure 2.3 shows the eight *T. obsoleta* clones that were successfully ligated into the pGEM®-T vector.

The homology of these sequences was tested via a BLAST search on the GenBank database. The sequences were shown to be very similar to rDNA

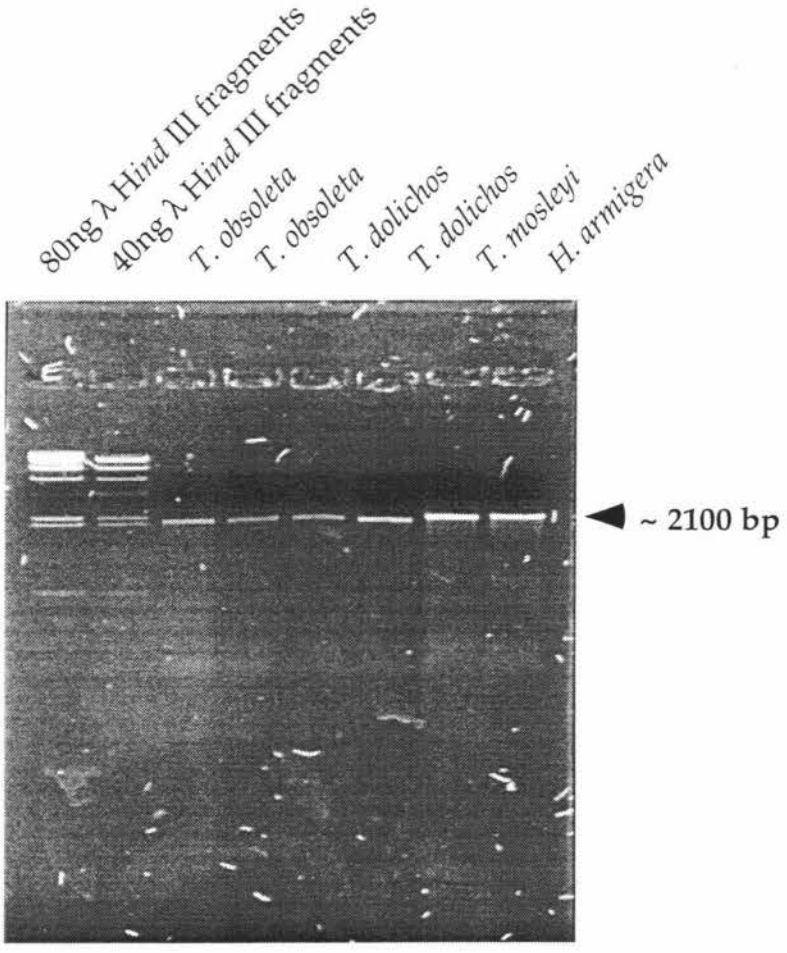

Figure 2.2 PCR amplified rDNA fragments of *Triplectides obsoleta, T. dolichos, Triplectidina mosleyi* and *Heliothis armigera*. Primers 18d and 28z (Hillis & Dixon, 1991) were used to amplify these fragments.

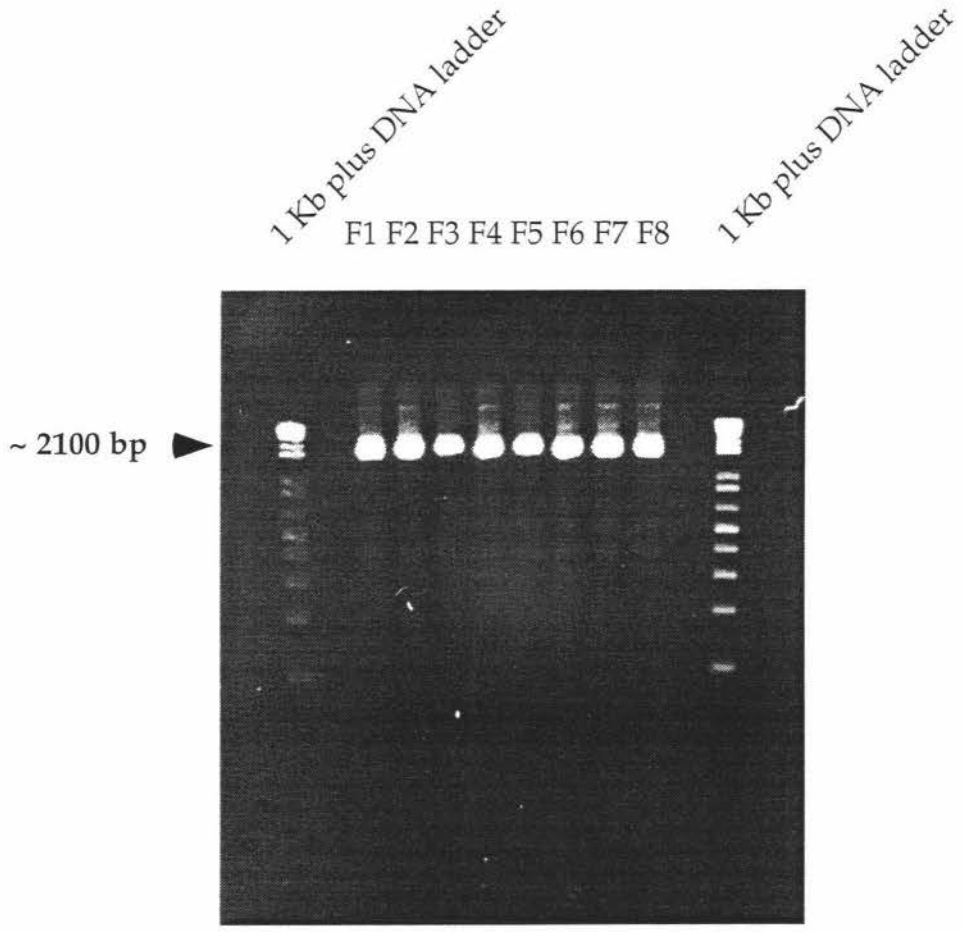

Figure 2.3 PCR amplified inserts of the cloned *Triplectides obsoleta* individual (F10PN). The eight clones are labelled F1 to F8. Primers18d and 28z (Hillis & Dixon, 1991) were used to amplify inserts.

sequences from several related insect taxa. GenBank sequences from the following taxa were used as outgroups to determine the identity of the sequenced PCR products as rDNA (accession nos. are given in brackets): *Hydropsyche sp.* (X89483), *Aedes aegypti* (M95126), *Drosophila melanogaster* (M21017), *Chironomus tentans* (X99212), *Sciara coprophila* (V01340), *Musca domestica* (Z28417), *Anopheles albimanus* (L78065), *Bombyx mori* (X01527) and *Vespa crabro* (S76777).

ITS-specific primers were designed in conserved regions within the rRNA genes at least 70 bp from the 5′ and 3′ end of a spacer. The exact positions of the gene-coding regions, including the internal 28S structures (core and expansion segments), were determined by comparison to the secondary structure of rRNA from *D. melanogaster* (Hancock et al., 1988). Multiple alignments of the rDNA gene-coding regions and exact ITS-specific primer positions are shown in tables 2.4, 2.5 and 2.6 for 18S, 5.8S and 28S rDNA respectively. The priming site for the R-28S-5′ primer was located in the 28S core sequence data and showed no mismatches. The primer (RITS-2_28S) was designed further downstream from this to incorporate more of the core region.

Further 28S sequence was obtained using the newly designed FITS-2_5.8S and RITS-2_28S ITS-2 primers (see figure 2.1). This sequence data was obtained from the caddisflies *O. fimbriata, H. spatulata* and the cloned *T. obsoleta*, and aligned to the flies *A. albimanus, C. tentans* and *D. melanogaster*. The multiple alignment of these sequences is shown in table 2.7. The priming site for the CS250 (28S) primer was located in the 28S core sequence data and showed one mismatch (see table 2.7). A PCR simulation using Amplify 1.2 showed this primer bound with low efficiency (64%) to *T. obsoleta* 28S. Therefore a new primer (RITS-2_core) was designed in this region to pair with FITS-2_5.8S. These primers efficiently amplified a smaller ITS-2 product which was sequenced in both directions.

```
                1                                  27                                  50
T. obsoleta     ?????????? ?????????? ?????????? ?????????? ?GTCTTCGGA
T. dolichos     ?????????? ?????????? ?????????? ?????????? ??????????
T. mosleyi      ?????????? ?????????? ?????????? ?????????? ??????????
Hydropsyche sp  CACACCGCCC GTCGCTACTA CCGATGG-GT GATTTGGTGA G.........
H. armigera     ?????????? ?????????? ?????????? ?????????? ??????????
A. aegypti      ?????????? ??CGCTACTA CCGATGG-AT TATTTAGTGA G.....T.A.
D. melanogaster CACACCGCCC GTCGCTACTA CCGATTGAAT TATTTAGTGA G....C....

                51                                  79                                  100
T. obsoleta     CTGGTGCGCG -GAAGCTCTT CGCGGGCATC CGACGCGTCG GAAAGTTTGA
T. dolichos     ?????????? ?????????? ??........ .......... ..........
T. mosleyi      ?????????? ?????????? ???....... .......... ..........
Hydropsyche sp  .......... -........-. ..-....G.. .......... .G........
H. armigera     ?????????? ??CTCGG..G TCG.C.TTG. ATGG.AA--- -----G....
A. aegypti      GGT.AA.ATT T.CTAG..CC TCG..ATTA. ATTT.AA... CTG.AG....
D. melanogaster .GT.AT.A.T -.TGA.G.C. T...T.TTA. G.TT.TT... C...AG....

                101                                                                 150
T. obsoleta     CCGAATTTGA TCATTTAGAG GAAGTAAAAG TCGTAACAAG GTTTCCGTAG
T. dolichos     .......... .......... .......... .......... ..........
T. moselyi      .......... .......... .......... .......... ..........
Hydropsyche sp  .......... .......... .......... .......... ..........
H. armigera     ..A..C.... .......... .......... .......... ..........
A. aegypti      .....C.... .G........ .......... .......... ..........
D. melanogaster .....C.... .T........ .......... .......... ..........

                151                 173
T. obsoleta     GTGAACCTGC GGAAGGATCA TTT
T. dolichos     .......... .......... ...
T. mosleyi      .......... .......... ...
Hydropsyche sp  .......... A......... ..A
H. armigera     .......... .......... ..A
A. aegypti      .......... .......... ..A
D. melanogaster .......... .......... ..A
```

Table 2.4 Multiple alignment of a partial 18S sequence (5′ to 3′) of *T. obsoleta, T. dolichos, T. mosleyi* and *H. armigera* to GenBank published sequences of a caddisfly, *Hydropsyche sp.*, and the flies *A. aegypti* and *D. melanogaster*. Position 173 denotes the 3′ end of 18S and the beginning of ITS-1. The forward primer used to amplify this region (18d) is underlined at position 1 to 27 and the primer designed in this region (FITS-1_18S) is underlined between positions 79 and 103. Matching characters = (.), missing characters = (?) and gaps = (-). See figure 2.1 for the sequence position within a rDNA repeat unit.

```
                  1                                   28                                  50
T. obsoleta       TACCCTGAAC GGCGGATCAC TCGGCTCGTG GGTCGATGAA GAACGCAGCT
B. mori           .......G.. ..T..... .T......C. .......... ........T.
A. aegypti        A.....AGG. A.G..... .......... .A........ ..C.......
C. tentans        A.....AG.. A.G..... .T.....A.. .......... ..C......A
S. coprophila     A.....A.G. ..G..... .T..T.T... .......... .........A
M. domestica      C..T..A.G. ..T..... .......A.. .......... .........A
D. melanogaster   A..T..A.G. ..T..... .......A.. .......... .........A

                  51                                                  94   100
T. obsoleta       AAATGCGCGT CAGAGTGTGA ACTGCAGGAC ACA--TGAGC ATCGACATTT
B. mori           ..C..... ..T..... .......... ...TT...A. ..........
A. aegypti        .......... ....A..... .......... ...--...A. .C.....CG.
C. tentans        ..C..... .GCCA..... .......... ...--...T. ..T.....G.
S. coprophila     ..C....T.. TGACA..... .......... ...--...A. ..T.......
M. domestica      ..C..T.... ..TC...... .......-.. ...--...A. ..........
D. melanogaster   ..C..T.... ..TC...... .......... ...--...A. ..........

                  101          115         125
T. obsoleta       CGAACGCACA TTGCGGTCCG AGCAA
B. mori           .......... .......... T.G.G
A. aegypti        T.......T. ....ACAT.. TA.T.
C. tentans        T.......T. ..???????? ?????
S. coprophila     T.......T. .........A TA.TG
M. domestica      T.......T. .C..A....A T..TG
D. melanogaster   T.......T. .C..A....A T..TG
```

Table 2.5 Multiple alignment of the complete 5.8S sequence (5′ to 3′) of *T. obsoleta* to GenBank published sequences of the silkworm moth, *B. mori* and the flies; *A. aegypti, C. tentans, S. coprophila, M. domestica* and *D. melanogaster.* Primers designed in this region are underlined. Primer FITS-2_5.8S lies between positions 28 and 49. Primer RITS-1_5.8S lies between positions 94 and 115. Matching characters = (.), missing characters = (?) and gaps = (-). See figure 2.1 for the sequence position within a rDNA repeat unit.

```
                  1                                                  34                  50
T. obsoleta       AGAGTCGGGT TGCTTGAGAG TGCAGCCTTA AGTGGGTGGT AAACTCCATC
T. dolichos       .......... .......... .......... .......... ..........
T. mosleyi        .......... .......... .......... .......... ..........
C. tentans        G.....AT.. .......T.. ......A... ......A... .......T..
A. albimanus      G......T.. .......T.. ......AC.. ......A... .......T..
D. melanogaster   .......T.. .......T.. ......AC.. .......... ..........
V. crabro         .......... .......... ......TC.. .......... ..........

                  51 55                        74                           100
T. obsoleta       TAAGGCTAAA TACTACCACG AGACCGATAG CGAACAAGTA CCGTGAGGGA
T. dolichos       .......... .......... .......... .......... ..........
T. mosleyi        .......... .......... .......... .......... ..........
C. tentans        .......... ..T.....T. .......... .......... ..........
A. albimanus      ...A...... ..TC....T. .......... .......... ..........
D. melanogaster   ...AA..... ..TA....T. .......... TA........ ..........
V. crabro         .......... ..TG..... .......... .......... ..........

                  101                                                                 150
T. obsoleta       AAGTTGAAAA GAACTTTGAA GAGAGAGTTA AAGAGTACGT GAAACCGTTC
T. dolichos       .......... .......... .......... .......... ..........
T. mosleyi        .......... .......... .......... .......... ..........
C. tentans        .......... .C...C.... T.......C. ..T....... .....T..CT
A. albimanus      .......... .C...C.... T.......C. .......... .....T.CCT
D. melanogaster   .......... .....C.... T......... ..C....... .....T.C.T
V. crabro         .......... .......... .........C .......... ..........

                  151    160
T. obsoleta       AGGGGT--AA
T. dolichos       ......--..
T. mosleyi        ......--..
C. tentans        .....CTC..
A. albimanus      ......GC..
D. melanogaster   ..A...TA.G
V. crabro         ......--..
```

Table 2.6 Multiple alignment of the 28S core region between D1 and D2 expansion segments (5′ to 3′) for *T. obsoleta*, *T. dolichos* and *T. mosleyi* to GenBank published sequences of the flies *A. albimanus*, *C. tentans* and *D. melanogaster* and the European hornet, *V. crabro*. The reverse primer 28z was used to amplify this region. Primer R-28S-5′ is underlined between positions 34 and 55 and the primer designed in this region, RITS-2_28S, is underlined between positions 74 and 98. Matching characters = (.) and gaps = (-). See figure 2.1 for the sequence position within a rDNA repeat unit.

```
                  1                                                    50
T. obsoleta       ATACAGTCGA CCTCAAATCA GGAGAGATCA CCCGCCGAAC TTAAGCATAT
O. fimbriata      CGTAT..... .......... .A......T. .......... ..........
H. spatulata      ..TGT..... .......... ........T. .......... ..........
A. albimanus      ..CA...G.G ........A. T.T.T..CT. ...C.TA..T ..........
C. tentans        .ATTTCG... ......C... T.T....CT. ...C.T...T ..........
D. melanogaster   T..T.TA.A. ......C... TAT.G..CT. ...C.T...T ..........

                  51      60               * 76           87           100
T. obsoleta       CAGTAAGCGG AGGAAAAGAA ACTAACGAGG ATTCCCCGAG TAGCGGCGAG
O. fimbriata      .......... .......... .......... .......... ..........
H. spatulata      .......... .......... .......... .......... ..........
A. albimanus      T.A....G.. .....G.... ......A... ......T... ....T.....
C. tentans        T.A.T..G.. .......... ..C...AG.. ......TT.. ...T.....A
D. melanogaster   T.A.T..G.. .......... ......A... ...TT.TT.. ..........

                  101  108         119
T. obsoleta       CGAAACGGGA AGAGCCCAG
O. fimbriata      .......... .CG......
H. spatulata      .......... .C.......
A. albimanus      .......... .....T...
C. tentans        .......... TC......T
D. melanogaster   .....A.AA. .C.. TT...
```

Table 2.7 Multiple alignment of the 28S core region between ITS-2 and the D1 expansion segment (5′ to 3′) for *T. obsoleta, O. fimbriata* and *H. spatulata* to GenBank published sequences of the flies *A. albimanus, C. tentans* and *D. melanogaster.* The reverse primer RITS-2_28S was used to amplify this region. Primer CS250 (28S) is underlined between positions 60 and 76 and the primer designed in this region, RITS-2_core, is underlined between positions 87 and 108. The asterix * at position 73 denotes a primer mismatch. Matching characters = (.) and gaps = (-). See figure 2.1 for the sequence position within a rDNA repeat unit.

ITS products

Initial analysis of the PCR products showed very little variation in length across the ITS-1 and ITS-2 regions for a range of Leptocerid species (six Triplectinae; five *Triplectides* and *Triplectidina mosleyi*, and one Leptocerinae: *Oecetis unicolor*). The amplified ITS-1 and ITS-2 fragments are shown in Figure 2.4 and 2.5 respectively. Several of these samples were later cloned and sequenced to determine the homology of these as ITS fragments (data presented in chapter 3). No significant length variation was found among individuals from the Triplectinae in either spacer. By comparison, *O. unicolor* from the subfamily Leptocerinae has significantly shorter ITS spacers. Adjusting for portions of flanking 18S, 5.8S and 28S sequences, the estimated length of the respective ITS-1 and ITS-2 regions in the Triplectinae is ~440 bp and ~400 bp, and in *O. unicolor*; ~200 bp and ~200 bp.

The specificity of the ITS primers was tested by applying them to a wide taxonomic range of trichopteran species. The PCR protocol for the primer pair FITS-1_18S and RITS-1_5.8S transferred well to other caddisflies and ITS-1 fragments were amplified from all. This was not the case for the primer pair FITS-2_5.8S and RITS-2_core. ITS-2 fragments were able to be amplified in three species outside the superfamily to which *T. obsoleta* belongs, though time constraints prevented further optimization. No ITS-1 products were able to be amplified in the moth *H. armigera* but multiple products were obtained using the ITS-2 primers. Amplified ITS-1 and ITS-2 fragments for these taxa are shown in figure 2.6 and 2.7 respectively.

ITS products from *H. spatulata* and *O. fimbriata* were sequenced directly to determine their homology. These were found to have ~100% sequence identity to the flanking gene-coding regions (results not shown). Adjusting for portions of flanking 18S, 5.8S and 28S sequences, the estimated length of ITS-1 varied

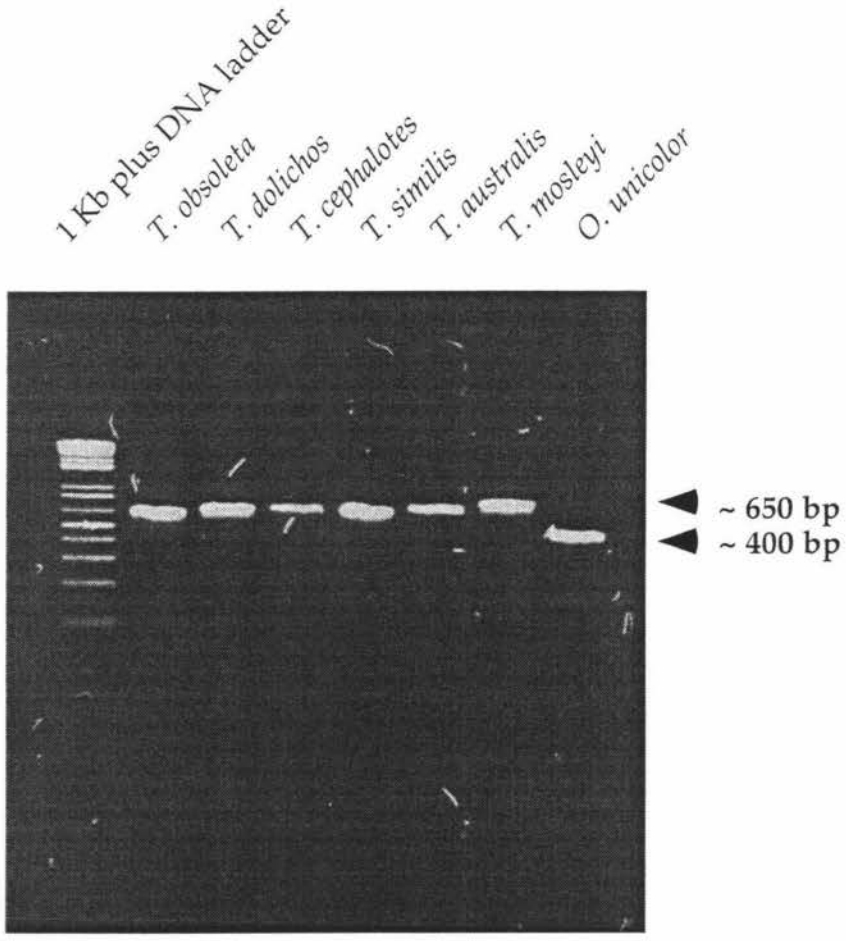

Figure 2.4 PCR amplified ITS-1 fragments of *Triplectides obsoleta, T. dolichos, T. cephalotes, T. simils, T, australis, Triplectidina mosleyi* and *Oecetis unicolor*. Primers FITS-1_18S and RITS-1_5.8S were used to amplify these fragments.

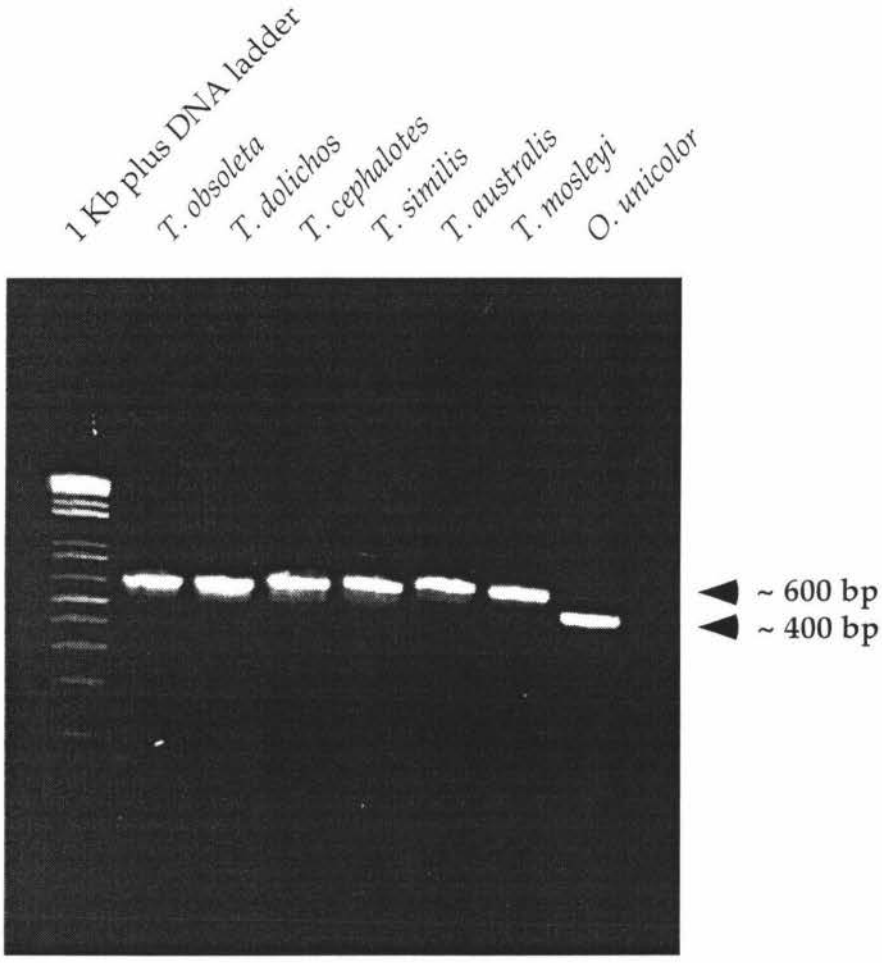

Figure 2.5 PCR amplified ITS-2 fragments of *Triplectides obsoleta, T. dolichos, T. cephalotes, T. simils, T, australis, Triplectidina mosleyi* and *Oecetis unicolor.* Primers FITS-2_5.8S and RITS-2_core were used to amplify these fragments.

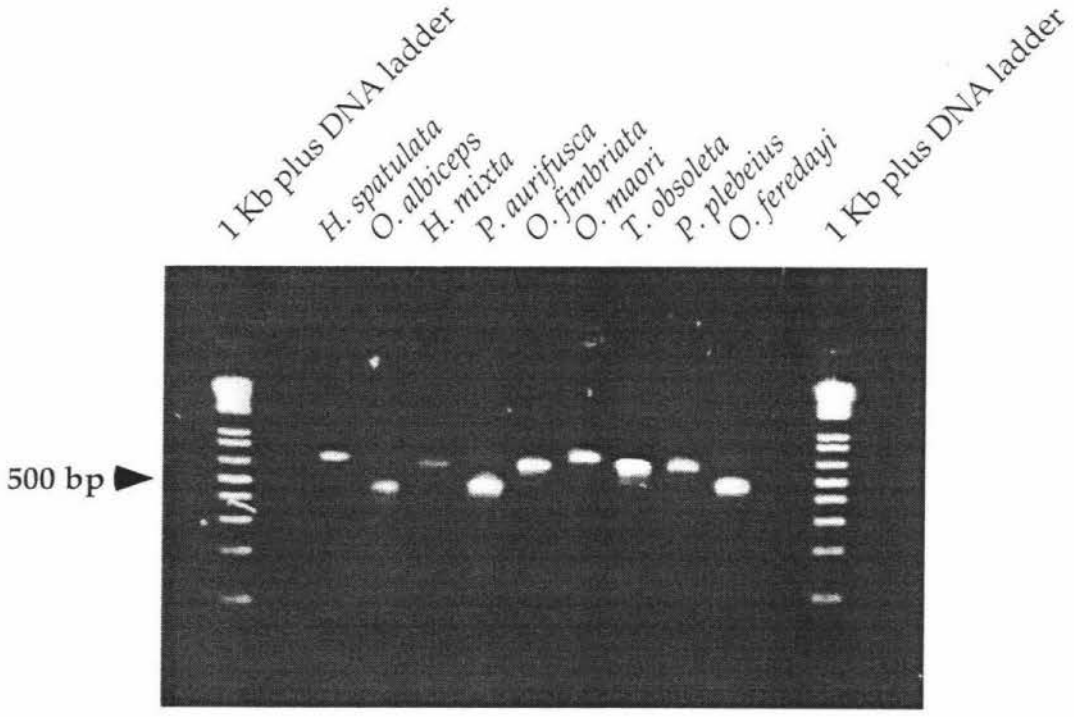

Figure 2.6 PCR amplified ITS-1 fragments of nine trichopteran species (see table 2.1 for taxonomic placement). Primers FITS-1_18S and RITS-1_5.8S were used to amplify these fragments.

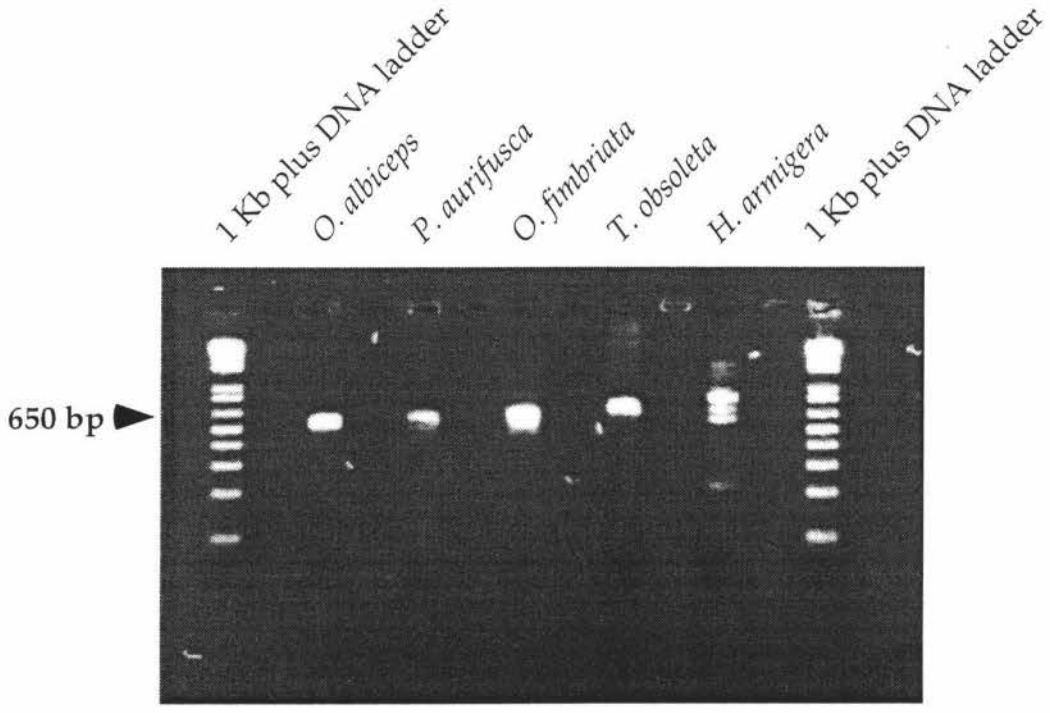

Figure 2.7 PCR amplified ITS-2 fragments of four trichopteran species including multiple products from the silkworm moth, *Heliothis armigera*. Primers FITS-2_5.8S and RITS-2_core were used to amplify these fragments.

between ~250 bp and ~500 bp and between ~350 bp and ~450 bp in ITS-2 in this trichopteran sample.

Analysis of ITS-1 and ITS-2 in T. obsoleta: a pilot study

Complete ITS-1 and ITS-2 sequence data were obtained from eight clones of an individual *T. obsoleta*. Two clones were sequenced from each of four independent PCR reactions (parallel samples). F1 and F2 clones were from one PCR reaction, F3 and F4 from another, as were F5 and F6, and F7 and F8. Two distinct ITS-1 types were found in these clones. Six ITS-1 clones were 395 bp in length and differed by only three substitutions. The other two (F4 and F7) were 391 bp in length and differed from each other by one substitution. The two ITS-1 types differed from each other by two substitutions and 14 insertions/deletions (indels). Only one ITS-2 type was found. Each ITS-2 clone was 451 bp in length and differed at most by two bases. The ITS-1 is AT-rich; 57.5% in the more common ITS-1 type and 58.1% in F4 and F7 clones. The ITS-2 however, is GC-rich with A+T at just 45.9%. A multiple alignment of the eight ITS-1 clones is shown in table 2.8 and a consensus sequence of ITS-2 (the two substitutions omitted) is shown in table 2.9.

```
      1                                                            50
F1    CCTGTGTTAT ATTTTATATA CATCCGTATA TAAAAATATA TACATCAAAA
F2    .......... .......... .......... .......... ..........
F3    .......... .......... .......... .......... ..........
F5    .......... .......... .......... .......... ..........
F6    .......... .......... .......... .......... ..........
F8    .......... .......... .......... .......... ..........
F7    .......... .......... ....T..... .......... ..........
F4    .......... .......... ....T..... .......... ..........

      51                                                           100
F1    TATATTTGAA CGCTTATAAT ATCGCTCGCT TGAACGCTTC TAGTGAGAGG
F2    .......... .......... .......... .......... ..........
F3    .......... .......... .......... .......... ..........
F5    .......... .......... .......... ......A... ..........
F6    .......... .......... .......... .......... ..........
F8    .......... .......... .......... .......... ..........
F7    .......... .......... .......... .......... ..........
F4    .......... .......... .......... .......... ..........

      101                                                          150
F1    AACCTTCAGT GG-ACGGCCC GACAGAGCGA GGGCCCCACC GCCCCAGGGT
F2    .......... ..-....... .......... .......... ..........
F3    .......... ..-....... .......... .......... ..........
F5    .......... ..-....... .......... .......... ..........
F6    .......... ..-....... .......... .......... ..........
F8    .......... ..-....... .......... .......... ..........
F7    .......... ..G....... .......... .......... ..........
F4    .......... ..G....... .......... .......... ..........

      151                                                          200
F1    GTGGTTCGGT GATGAGCGGA TTTCGTCGTG T--CTGTAAG GCGGTGGTGT
F2    .......... ..C....... .......... .--....... ..........
F3    .......... .......... .......... .--....... ..........
F5    .......... .......... .......... .--....... ..........
F6    .......... .......... .......... .--....... ..........
F8    .......... .......... .......... .--....... ..........
F7    .......... .......... .......... .GT......- --........
F4    .......... .......... .......... .GT......- --........
```

Table 2.8 Multiple alignment of the complete ITS-1 sequence (5′ to 3′) from eight clones (F1 to F8) of a single *T. obsoleta* individual. Matching characters = (.) and gaps = (-).

```
      201                                                        250
F1    TTTGTCGTGG TGTGCGAACG CACGCACGCC GCACACACAC ACACACGCCG
F2    .......... .......... .......... .......... ..........
F3    .......... .......... .......... .......... ..........
F5    .......... .......... .......... .......... .......T..
F6    .......... .......... .......... .......... ..........
F8    .......... .......... .......... .......... ..........
F7    .G........ .......... .......... .........- -----.....
F4    .G........ .......... .......... .........- -----.....

      251                                                        300
F1    CACGGCGTTC CGCATCTGTT AACCGATTTT TTTT-GGGTA ACGCGTTTCT
F2    .......... .......... .......... ....-..... ..........
F3    .......... .......... .......... ....-..... ..........
F5    .......... .......... .......... ....-..... ..........
F6    .......... .......... .......... ....-..... ..........
F8    .......... .......... .......... ....-..... ..........
F7    .......... .......... .......... ....T....G ..........
F4    .......... .......... .......... ....T..... ..........

      301                                                        350
F1    CGTTTACGAG AAATTCCCTT TACATTTTCG ATTGTATATA GAATATGAAA
F2    .......... .......... .......... .......... ..........
F3    .......... .......... .......... .......... ..........
F5    .......... .......... .......... .......... ..........
F6    .......... .......... .......... .......... ..........
F8    .......... .......... .......... .......... ..........
F7    .......... .......... .......... .......... ..........
F4    .......... .......... .......... .......... ..........

      351                                                        400
F1    TAATTTTTTT TT-ATATAAT ATCTGTTGAT TTTACAACAT GAAAAATCAT
F2    .......... ..-....... .......... .......... ..........
F3    .......... ..-....... .......... .......... ..........
F5    .......... ..-....... .......... .......... ..........
F6    .......... ..-....... .......... .......... ..........
F8    .......... ..-....... .......... .......... ..........
F7    .......... ..T....... .......... .......... ..........
F4    .......... ..T....... .......... .......... ..........
```

Table 2.8 (continued) Multiple alignment of the complete ITS-1 sequence (5′ to 3′) from eight clones (F1 to F8) of a single *T. obsoleta* individual. Matching characters = (.) and gaps = (-).

```
1                                                          50
AACGTAGTTT GTCGACGGAC CACTCCTGTT TGAGGGTCGC GCACGTAAAC

51                                                        100
GTGACCAGAC TCCGAACTTC GGTTCGGGAG AGTTGTCGGT TTCGGTTCAC

101                                                       150
GGGTTTTATC CCCCGGGGGC CGTCCGATGA AATCGAACGT ACGAACACCG

151                                                       200
CGCGTTTTTT TTTGTTAAAG CGCGCGTTCT TGTTCGTTTT ACCGAGGCGT

201                                                       250
GTACCGATCT GGTATCGCGT TCCTCTTCGA ACGATTCAGT GCACTAGCGC

251                                                       300
GCTCCGGCGC CCGCACGCGT TACGGAGAAC GTGTACACCA TCGACCGCTA

301                                                       350
GCGATGACGC TGCATACTGG AGATTTTTCA GATGAAAAAT TTTCCGGGCG

351                                                       400
ATCGTCGCGC GGTCTTCGAA CGAGGTGACG TTCTCGTCGC GTCAGCGGGT

401                                                       451
ACGCCGTACG GGCCTATCGA GTTACGTGTA CGATATATAT GACATTTTTTA
```

Table 2.9 Complete consensus sequence of ITS-2 (5′ to 3′) from eight clones of a single *T. obsoleta* individual.

2.4 Discussion

Determining homology

The Trichoptera together with their sister order Lepidoptera (butterflies and moths) form the Amphiesmenoptera (Hennig, 1981). Recent phylogenetic reconstructions using 18S rDNA sequences support this relationship and suggest that the Diptera (flies) are their closest relatives (Pashley et al., 1993). BLAST searches on the GenBank database of sequenced caddisfly PCR products yielded high sequence identity to rDNA from members of these orders. Therefore, dipteran and lepidopteran taxa were used as outgroups to design the ITS-specific primers. These primers were designed to amplify a minimum of 70 bp of the conserved gene-coding regions which flank the spacers. These regions were subsequently used to identify PCR products as rDNA.

There is greater length variation in the larger rRNA genes (18S and 28S) than the smaller 5.8S. This length variability results from the presence of regions known as 'expansion segments' that show less sequence similarity between eukaryotic species than do the 'core' segments (Hancock et al., 1988) (see figure 2.1). Two conserved core regions and the flanking expansion regions were identified in the 28S sequence data (tables 2.6 and 2.7) by comparison to the rDNA secondary structure analysis of *D. melanogaster* (Tautz et al., 1988; Hancock et al., 1988). Expansion segments of the 28S (e.g., D1, D2 and D3) have been shown to diverge more rapidly between species than core segments (Hancock et al., 1988). Therefore two 28S ITS-specific primers were designed to the conserved core sequence with the hope that these primers may be applied to other caddisflies.

The 28S primers (R-28S-5', CS250 and 28z) used in the preliminary work were designed to core segments. Primer CS250 (28S) is located in the core between the 3' end of ITS-2 and the D1 expansion segment, primer R-28S-5' is located

between expansion segments D1 and D2, and primer 28z between expansion segments D2 and D3. Although designed in conserved regions, primer/template mismatches may have accounted for the failure of the Schlotterer et al. (1994) and Kjer primer pairs to produce a single PCR product. The reduction in the stability of primer/template pairing caused by a single mismatch in primer CS250 (28S) may have caused primer failure (see table 2.6). For both CS250 and R-28S-5', priming failure may have occurred in the 18S region. These primers lie outside the region amplified with primer 18d, as verified by comparison to published sequences of *D. melanogaster* (Tautz et al., 1988) and the mosquito *A. aegypti* (Wesson et al., 1992). Therefore a mismatch in this region could not be determined from the caddisfly sequence data.

Length variation in ITS-1 and ITS-2

No significant within-species length polymorphisms were found in the ITS of seven *Drosophila* species (Schlotterer et al., 1994). Although the length of the IGS varies both within and between *Drosophila* species, within each species the ITS length appears invariant (Coen et al., 1982b). *Drosophila* have been the subject of many classic rDNA studies, recent findings however, demonstrate the presence of intra-specific ITS length polymorphism in other insects. A 52 bp ITS-1 length polymorphism has been found in the black fly *Simulum damnosum* (Tang et al., 1994). ITS-1 length varies between 419 and 426 bp, and the ITS-2, between 190 and 201 bp in the mosquito *A. aegypti* (Wesson et al., 1992). Estimates from multiple individuals of seven Triplectinae species suggests they share similar ITS lengths, though it is necessary to sequence individual clones to determine if any minor length polymorphisms exist (see chapter 3).

The range of ITS length variation in Trichoptera is comparable to that of other insect taxa. In the genus *Drosophila* ITS-1 ranges between 529 and 851 bp in length and ITS-2, between 353 and 410 bp (Schlotterer et al., 1994). In the mosquito genus *Culex* ITS-1 ranges between 184 and 572 bp and the ITS-2,

between 204 and 333 bp (Miller et al., 1996). Both ITS of *O. unicolor* are shorter than those of the other trichopterans surveyed but its estimated length is also within the range found in *Drosophila* and *Culex*.

Molecular taxonomy

Many insects (including caddisflies) are morphologically indistinguishable from congeners at one or more stages of their life cycle. This taxonomic problem is often exasperated by the lack of identification keys to identify eggs or early instars (Hiss et al., 1994). However, by comparing molecular characters in unknown larvae against known adults, species assignments often can be made (Avise, 1994). This can be particularly useful in cases where knowledge of a species presence or absence in a locality is required as is the case with caddisfly larvae which are used as bio-indicators to monitor levels of freshwater pollutants. The ITS-specific primers designed in this work, coupled with a rapid screening technique (e.g., RFLP or SSCP) may provide an effective method for the identification of caddisfly larvae.

ITS-1 and ITS-2 of T. obsoleta

There have been a few recent reports of intraspecific polymorphisms in the ITS which warn against the assumption that these spacers are homogenized (Rich et al., 1997; Miller et al., 1996; Tang et al., 1996; Volger & DeSalle, 1994; Wesson et al., 1992). Two ITS-1 types were present in the eight clones of an individual *T. obsoleta*. This heterogeneity consisted mainly of indels (14 out of 16 variable sites), with only two differences due to substitutional point mutations. Length variation was minor among the clones (391-395 bp). These differences amount to 4% of the sites in the ITS-1 alignment. The minor substitutional differences between some clones may have been *Taq* polymerase errors, or perhaps plasmid replication errors. This appears to be the case by comparison between parallel samples. The sequence differences that delimit the two ITS-1 types are likely to be genuine because they were isolated from different PCR replicates.

The presence of only one ITS-2 type suggests that this spacer may be more homogenized within the individual and perhaps within the species. No length variation was found in ITS-2, nor were there as many compositional differences as in ITS-1. The heterogeneity found in ITS-1 presents itself as an opportunity for further study. Intraspecific polymorphism in the ITS-1 of *T. obsoleta* may be found to be greater if more individuals are sampled and if more ITS-1 copies are isolated from individuals. A study into the nature of the variation in ITS-1 may determine whether this heterogeneity affects its usefulness as a genetic marker for population studies.

Chapter 3

Variation in the rDNA ITS-1 of Two Caddisflies

3.1 Introduction

Genetic markers and genetic dissection

Griffiths et al. (1993) described the use of two widely successful techniques in the field of genetics. The first is 'genetic dissection'; the genetic analysis of variants to identify a particular gene that is important for a biological process (e.g., organismic development). The second is the use of specific genes or nucleotide sequences as 'markers'. A genetic marker is a variant allele that is used to label a biological structure or process throughout the course of an experiment (Griffiths et al., 1993). The latter technique is applied here, using ITS-1 to 'label' coevolutionary processes in the rRNA gene family of two caddisflies: *Triplectides obsoleta* and *T. dolichos*.

Concerted evolution and molecular drive

ITS-1 undergoes concerted evolution (Dover, 1982) such that individual copies in a gene family do not evolve independently, but interact through non-Mendelian DNA transfer, during replication and recombination. Molecular drive is the cohesive process by which mutations are able to spread through a gene family (homogenization) and through a population (fixation) as a consequence of these interactions within and between chromosomes (Dover, 1986). The 'shuffling' of chromosomes in sexual populations is considered to occur faster than the rate at which new mutant copies arise in an individual, during its lifetime. The disparity in rates between genomic turnover (in an individual), and recombination through sexual reproduction ensures that the

genetic cohesion of a population is maintained throughout a period of change under molecular drive (Dover, 1982). The utility of ITS-1 as a population genetic marker is dependent on whether this 'genetic cohesion' partitions sequence variation predominantly within interbreeding populations.

Paralogy and orthology

The homology of ITS-1 copies is complicated because they arise by gene duplication. The multiple copies are termed paralogues; homologues that reflect gene history rather than organismal genealogy (Fitch, 1970). If they undergo independent evolution, phylogenetic analysis would infer the history of these duplication events. However, through concerted evolution they behave like orthologous sequences (formed by speciation); they show little divergence within species, but evolve rapidly between species (Hillis, 1994). Paralogous copies that are homogenized (identical) within a taxon can be used to infer phylogenetic relationships of taxa without fear of reconstructing gene duplication events. In a simulation, Sanderson and Doyle (1992) found that at either very low or very high levels of homogenization, the correct organismal tree can be inferred from multigene family data; only intermediate levels cause confusion.

Phylogenetic and population studies

Both spacers (ITS-1 and ITS-2) have been shown to be appropriate for investigating relationships at or below the genus level, for a wide range of taxa (e.g., Lee & Taylor, 1992; Wesson et al., 1992; Ritland et al., 1993). Coleman and Mai (1997) discuss the value of these spacers as guides in searching for closely related organisms and contaminants by comparison to published DNA sequences. Several recent studies (Odorico & Miller, 1997; Quijada et al., 1997) demonstrate their usefulness in tracing hybridization events and reticulate evolution. ITS sequence variation has also been used to determine relationships among conspecific populations (Fritz et al., 1994; McLain et al., 1995).

However, there are a growing number of studies (Kuperus & Chapco, 1994; Miller et al., 1996; Tang et al., 1996) that have faced difficulties in using ITS because of heterogeneity among isolated (cloned) copies. Miller et al. (1996) used ITS to produce a phylogeny for the *Culex pipiens* complex of mosquitoes where ITS variation indicated polyphyly among taxa traditionally considered to be monophyletic. Volger and DeSalle (1994) found that phylogenetic resolution of ITS-1 in the tiger beetle *Cicindela dorsalis* was low, and that variation within individuals was almost as high as variation within entire regional lineages. Williams et al. (1988) considered the assumption that high levels of intra-specific sequence homogeneity are a 'universal' feature of the rRNA gene family, to be overstated. Low levels of homogenization may obscure phylogenetic resolution among the units of a study. Indeed, if variability exists among the paralogous copies, these cannot serve as markers of lineage, unless it can be determined that the variation is distributed in a population-specific (diagnostic) manner (Rich et al., 1997).

Sampling problems

Certain problems apply to the practicalities of sampling a gene family, apart from the primary assumption that the intended product has been amplified (see chapter 2). PCR primers designed to hybridise with a number of different though very similar genes, runs the very high risk of producing recombinant artefacts (Zylstra et al., 1998). These are produced when partially extended DNA from one primer attachment, attaches at a second site during a subsequent extension cycle (Hillis et al., 1996). Zylstra et al. (1998) suggest that using the minimum number of cycles required to obtain a clonable product virtually eliminates recombination. The DNA extraction process produces linear DNA fragments that can be nicked in the region intended for amplification. These also produce recombinants though the problem is reduced by extracting high molecular weight DNA and amplifying smaller fragments.

Using primers designed to hybridise to multiple targets also raises the problem of PCR drift; a sampling bias where presumptive random events during early PCR cycles leads to preferential amplification of certain members of a gene family. The bias is not repeatable, that is, separate PCR experiments in general do not produce biases towards the same member of the gene family (Wagner et al., 1994). Running parallel PCR reactions for each sample will reduce the possibility of such an amplification bias. In addition, the accuracy of a particular gene sequence is guaranteed by re-isolation of the identical sequence from an independent amplification (Zylstra et al., 1998).

A study into the nature of variation in ITS-1

For a genetic marker to be employed routinely, reasonable steps must be taken to assure that variability within the units of study is not extensive (Rich et al., 1997). In this study, intragenomic variability was assayed to determine whether ITS-1 variation is distributed in a population-specific manner for two closely related caddisflies; *Triplectides obsoleta* and *T. dolichos*. Intra-specific variation was assayed by sampling among geographic populations and within individuals. However, this is neither a population nor phylogenetic study. These sampling levels were treated as integral parts of an hierarchical study into the 'nature of the variation' in a non-coding spacer region; ITS-1.

3.2 Materials and Methods

Samples

Adult caddisflies were collected using a UV light trap (see chapter 2). 12 *Triplectides obsoleta* and 5 *T. dolichos* were assayed for ITS-1 variation from several locations in both the North and South Islands of New Zealand. The likelihood of sampling from genetically isolated populations was maximised by sampling from locations that covered the full geographic range of New Zealand. Collection localities and individual codes (subsequently referred to in

the text) are shown figure 3.1. ITS-1 nucleotide sequence data were also isolated from a single individual of *T. cephalotes*.

DNA extraction

High molecular weight genomic DNA (~20 kb) was isolated from caddisflies preserved in 70% ethanol (see chapter 2). An example of the products of several DNA extractions is shown in figure 3.2.

PCR amplification

The ITS-specific primers FITS-1_18S (5'-TCC GAC GCG TCG GAA AGT TTG ACC G-3') and RITS-1_5.8S (5'-CGC AAT GTG CGT TCG AAA TGT C-3') (see chapter 2) were used to amplify ITS-1 for cloning, SSCP analysis, and DNA sequencing. Standard M13 forward and reverse primers were used to screen for cloned inserts. PCR amplification of double-stranded DNA was performed, with slight modification from Kocher et al. (1989) in 10µl volumes containing 67mM Tris (pH 8.8), 1.5mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 10mM β-mercaptoethanol, 0.1µl glycerol, each dNTP at 200µM, each primer at 0.5µM, ~5ng genomic DNA, and 0.25 units of *AmpliTaq*® DNA polymerase (Perkin-Elmer). The thermocycling conditions for cloned products were one cycle at 94°C for 2 min, followed by 16-18 cycles at 94°C for 1 min, 60°C for 1 min, and 72°C for 2 min (increased to 30 cycles for colony screening with M13 primers), then one cycle at 72°C for 4 min.

ITS-1 fragments for each individual were amplified in four parallel PCR reactions. Reactions were allowed to proceed for 16-18 cycles in order to reduce the likelihood of recombination amongst PCR products. Products from each PCR reaction were cloned and then used for SSCP and sequencing analysis.

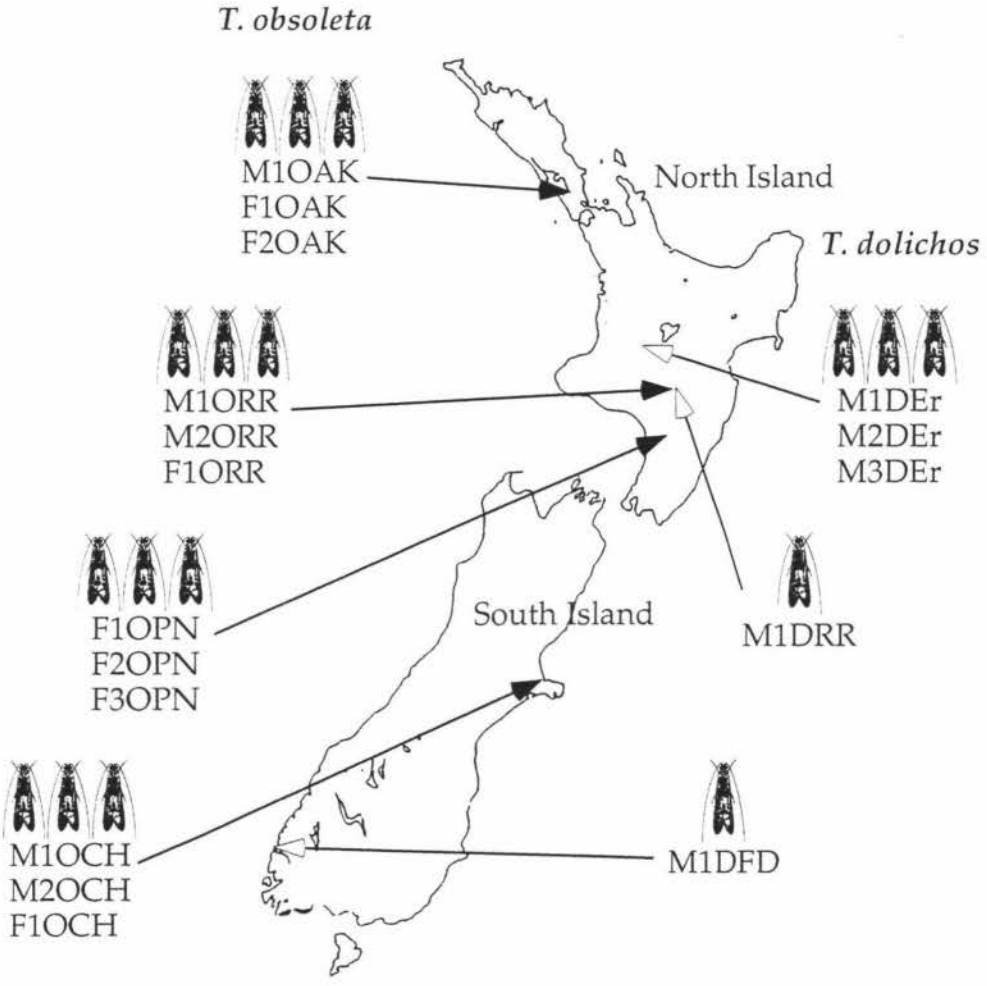

Figure 3.1 A map of New Zealand showing the origin of *T. obsoleta* and *T. dolichos* samples (closed and open arrows respectively). Individual codes lie below caddisfly figures. The prefix 'M' and 'F' denote male and female individuals. Collection sites are denoted by a two letter suffix. 'AK' = Auckland, 'RR' = Ruahine Ranges, 'Er' = Erua, 'PN' = Palmerston North, (North Island locations), 'CH' = Christchurch, and 'FD' = Fiordland (South Island locations). The one *T. cephalotes* sample (F1CRR) is not shown.

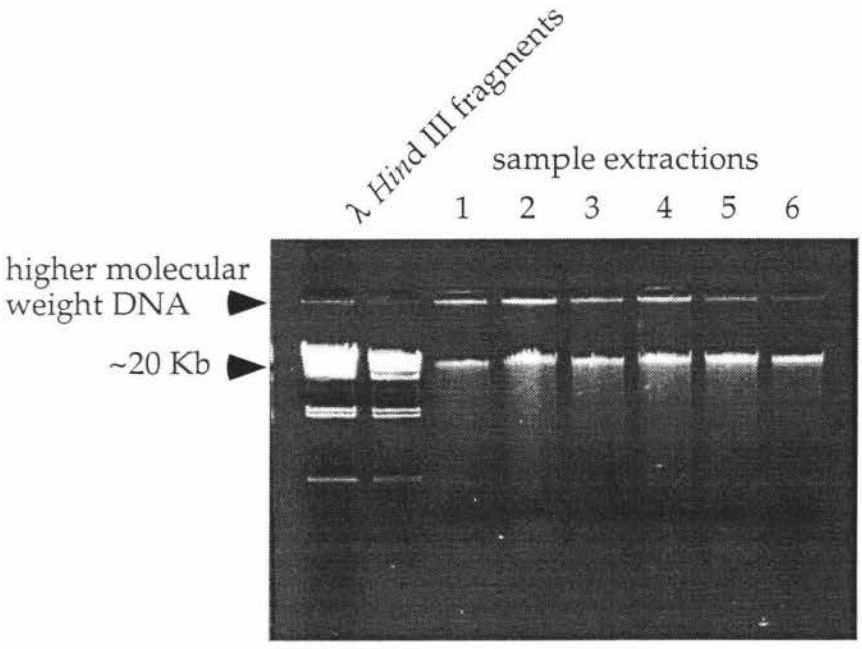

Figure 3.2 A 1% agarose gel showing an example of the average molecular weight (~20 Kb) of six genomic DNA extractions. Higher molecular weight DNA is present in the wells (zone of limited mobility).

Purification of amplified DNA

Before purification, 1µl of each PCR product was checked for size and DNA concentration by electrophoresis on 2% agarose gel along with a DNA standard size marker (1 Kb plus, GibcoBRL). The remainder of the PCR product was

purified using PCI (25:24:1) and Sephacryl® S-400 HR (Pharmacia Biotech, Uppsala, Sweden). The PCR product, minus the overlying oil, was placed into a sterile Eppendorf tube with an equal volume of PCI. Samples were then centrifuged at 15,500 g for 1 min. The aqueous phase was added to a column containing sephacryl. These columns were made by putting a small pipette tip through an incision made into the top of a 1.5ml Eppendorf tube. A siliconised glass bead and 200µl of sephacryl was added to the column and then this was centrifuged at 2,500 g for 2 min to remove the water. PCR samples were centrifuged through the column at 2,500 g for 2 min and the purified DNA was placed into a new tube. Aliquots from each sample were electrophoresed in 1% agarose with λ *Hin*d III fragments to check DNA recovery.

Cloning

PCR products were made blunt ended by treatment with T4 DNA polymerase I (GibcoBRL). Reactions were carried out in 10µl volumes containing 1 X T4 DNA ligase buffer (GibcoBRL), each dNTP at 200µM, 1 unit T4 polynucleotide Kinase (GibcoBRL), 0.5 units T4 DNA polymerase I, and purified PCR product (between 5-40ng). Reactions were incubated at 37°C for 20 min, then inactivated at 70°C for 5 min in a heating block. Ligations were carried out in the same tubes by increasing the volume to 15µl with the addition of 1 X T4 DNA ligase buffer, 0.5 units T4 DNA ligase (GibcoBRL), and 50ng of pUC18 *Sma* I (Pharmacia Biotech). Ligation were carried out overnight at room temperature (10-12°C).

Five microlitres of each ligation reaction was used for the transformation of competent *E. coli* cells and the remainder was stored at 4°C. The DH5α strain of *E. coli* was made competent according to the method of Chung et al. (1989). *E. coli* cells were grown in LB broth (see chapter 2) to mid-exponential phase (OD_{600} 0.5-0.8) and harvested by centrifugation at 1,000 g for 5 min. The pellet was gently resuspended in 1ml of ice-cold transformation and storage solution (TSS; LB broth with 10% polyethyleneglycol (PEG) 6000) and 70µl dimethyl-sulphoxide (DMSO). A 100µl aliquot of cells was added to 5µl of ligation reaction then mixed and the solution incubated on ice for 30 min. Cells were heat shocked at 42°C for 1 min before the addition of 900µl of LB broth. Cells were then grown at 37°C with agitation for 40 min to allow expression of the antibiotic resistance gene. Cells were centrifuged at 1,000 g for 1 min and 900µl of the supernatant was removed. After mixing, 100µl of the cell suspension was then spread onto agar plates (see chapter 2) and incubated overnight at 37°C.

Colony screening

Recombinant white colonies were screened by PCR amplification using standard M13 forward and reverse primers. Figure 3.3 shows an example of positive (insert size ~800 bp) and negative colonies (<800 bp). For each caddisfly, cells from 12 positive colonies (three from each initial PCR reaction) were streaked out onto new agar plates (see chapter 2) containing ampicillin (100µg/ml) and were grown overnight at 37°C. Crude plasmid extractions (see chapter 2) were made from these new colonies. The primers FITS-1_18S and RITS-1_5.8S were used to check for cloned ITS-1 inserts. The PCR products were electrophoresed on 2% agarose gel to confirm amplification and then analysed for sequence variation by SSCP analysis.

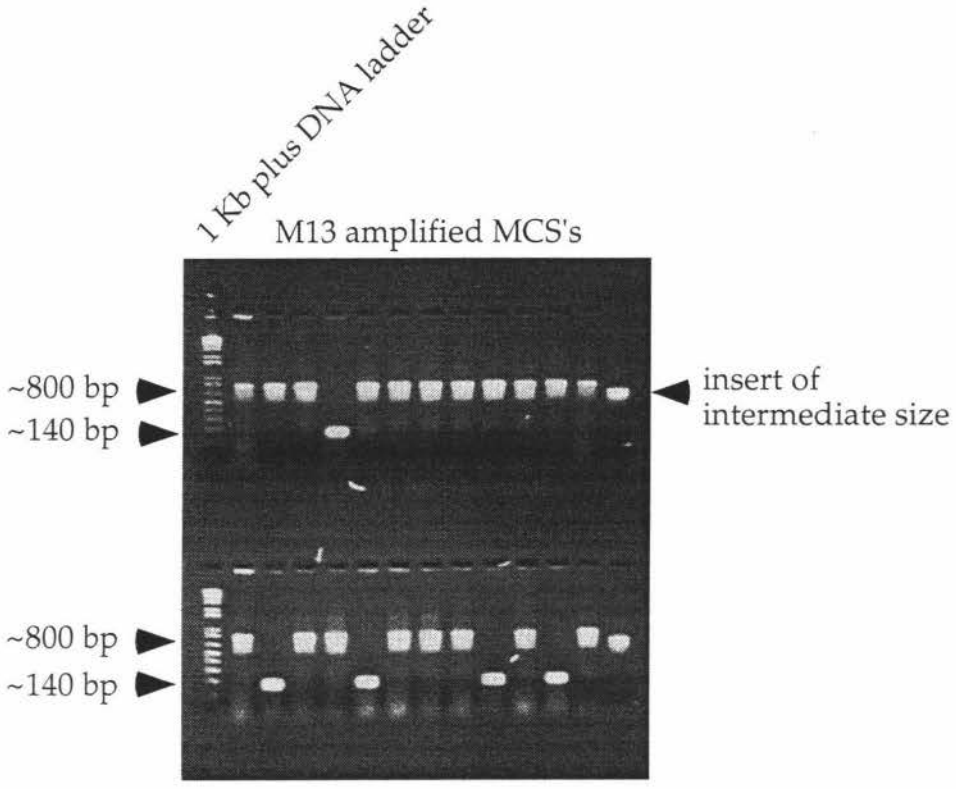

Figure 3.3 A 2% agarose gel showing an example of colony screening using PCR. Standard M13 primers were used to amplify across the pUC18 multiple cloning site (MCS). Clones with positive inserts are ~800 bp, those without are ~140 bp or of intermediate size.

Restriction enzyme digestion of PCR amplified ITS-1 clones

PCR-amplified ITS-1 clones were cleaved with restriction endonucleases (RE's) to circumvent having to analyse short PCR fragments in SSCP gels (Lee et al., 1992). All *T. obsoleta* ITS-1 clones were cleaved with *Hae* III. Clones from the *T. dolichos* individuals M1DFD and M1DRR were cleaved with *Hinf* I. However, some clones from M1DRR lacked this restriction site and *Dde* I was used thereafter for all other *T. dolichos* clones. Reaction buffer (1 X REact2) was used with *Dde* I digestions, but was not necessary for *Hae* III and *Hinf* I. PCR-amplified clones were digested with 1 unit of RE and incubated at 37°C for 1.5 h (all reagents from Boehringer Mannheim, Germany). *Hae* III digestion of all *T. obsoleta* yielded two fragments approximately 400 bp and 220 bp in length. *Hinf* I digestion of *T. dolichos* yielded two fragments approximately 440 bp and 200 bp and *Dde* I, two fragments both approximately 320 bp in length. An example of typical restriction digests are shown in figure 3.4.

Single-strand conformation polymorphism (SSCP)

For all clones, nonisotopic SSCP analysis ('cold SSCP') was conducted on the restriction digested PCR-amplified ITS-1 fragments following the protocol of Hongyo et al. (1993), with some modifications. Three microlitres of unpurified PCR product (~30ng of DNA) were mixed with 21µl of de-ionized formamide containing 0.05% bromophenol blue. This mixture was denatured for 5 min at 95°C and then chilled on ice before loading on the SSCP gel. 7% polyacrylamide (37.5: 1 acrylamide to bis-acrylamide, Bio-Rad) TBE gels (17cm x 15cm x 3cm) containing 2% glycerol, 30µl Temed and 250µl of 10% ammonium persulphate (APS), were run with 1 X TBE buffer (0.89M Tris, 0.89M Boric acid, 25mM EDTA) on a vertical electrophoresis system (model V16, GibcoBRL). Gels were run at constant power (7 W) for 3.5 h in a 4°C cool room. Gels were stained for 20 min with 0.2µg/ml ethidium bromide solution, and destained for 5 min with

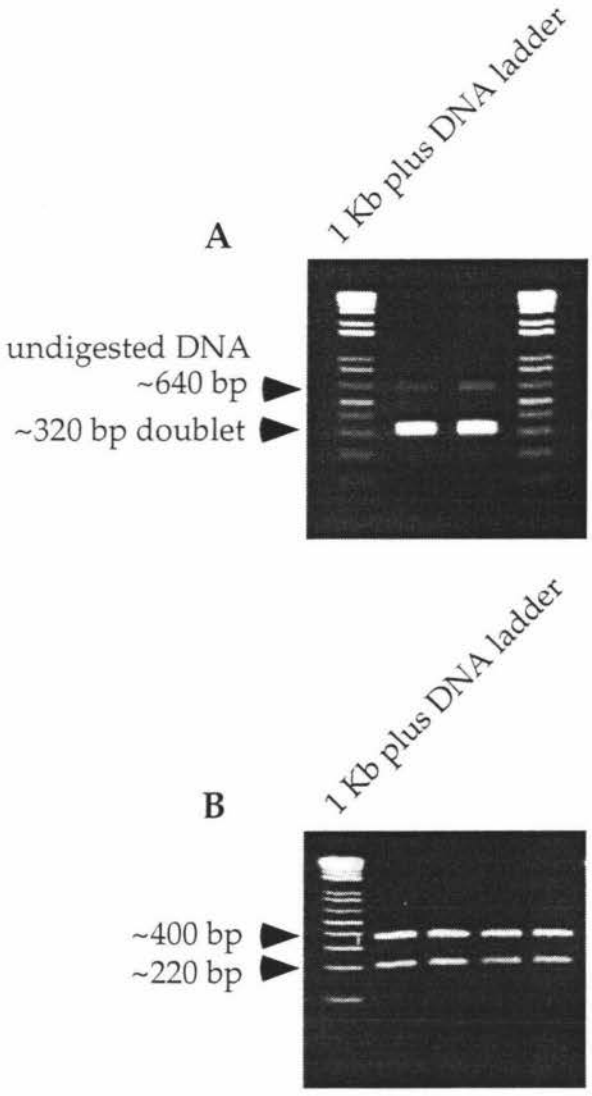

Figure 3.4 Two examples of restriction enzyme digested PCR-amplified ITS-1 clones. Gel 'A' shows the ~320 bp doublet formed by *Dde* I digestion in two *T. dolichos* samples. Gel 'B' shows the two RE fragments formed by *Hae* III digestion in *T. obsoleta* samples.

dH_2O. Bands were visualised under UV light and scanned using a gel documentation system (Insta Doc™ System, Bio-Rad).

Although formamide was used to as a denaturant, its effectiveness varied with DNA concentration. Hongyo et al. (1993) have shown that methylmercury hydroxide is a more effective denaturant, but we were unable to obtain this chemical, therefore large volumes of formamide were used. Figure 3.5 shows the effect of varying formamide:DNA ratios on the maintenance of denatured single-strand DNA (DSS). Renatured double-stranded DNA (RDS) was produced at lower formamide:DNA ratios and is intermediate in mobility between renatured single-strand DNA (RSS) and DSS (Hiss et al., 1994).

Plasmid DNA extraction and sequencing

A subset of the SSCP screened ITS-1 clones were chosen for plasmid extraction (those isolated from separate PCR's) using the QIAprep miniprep kit (QIAGEN) as instructed by the manufacturers. ITS-1 clones (common and rare ITS-1 variants) were sequenced from two independent PCR amplifications. For each clone, both forward and reverse strands were sequenced. See chapter two for DNA sequencing details.

Computer programs

The Repeat and StemLoop programs of the GCG software package (Wisconsin Genetics Group) were used to identify regions of direct and inverted repeat DNA >7 bp in length. Sequence data were edited using EditView 1.0.1 (Perkin-Elmer), Amplify 1.2 (Madison, WI), and PAUP v 3.1.1 (Swofford, 1990). Sequences were aligned by the method of Higgins and Sharp (1988), using CLUSTAL W(1.4). Alignments were formatted in MacClade 3.06 (Maddison & Maddison, 1992) then checked and adjusted manually, if necessary.

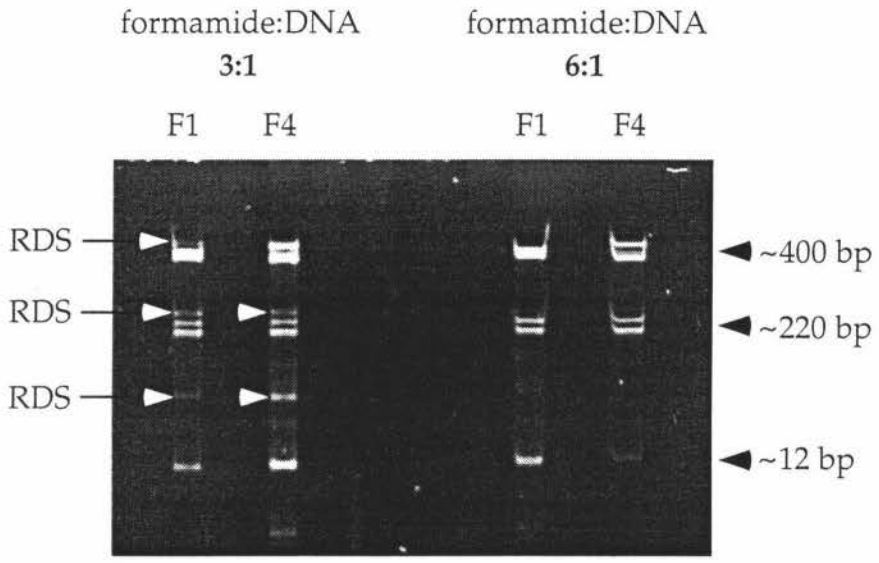

Figure 3.5 An SSCP gel showing the effect of two varying formamide:DNA ratios on the maintenance of denatured single-strand DNA (DSS). A 3:1 ratio represents 9µl formamide:3µl DNA (30ng) and a 6:1 ratio represents18µl formamide:3µl DNA (30ng). At 3:1 some single-strand DNA reanneals to form renatured double-stranded DNA (RDS). RDS bands are denoted by open arrows. All other bands are DSS. F1 and F4 represent two *T. obsoleta* ITS-1 clones that differ in composition by 14 bp. The approximate size of the DSS *Hae* III fragments is shown on the right hand side.

3.3 Results

SSCP analysis

SSCP analysis was employed as a method to screen intra-individual sequence variation and formed the basis for a DNA sequencing strategy. Separate SSCP gels were run for each individual and ITS-1 variation was determined by band mobility. The 12 clones from each individual were grouped by shared band mobility and variants were sequenced, if isolated from separate PCR reactions. All individuals screened, 11 *T. obsoleta* (132 clones) and 5 *T. dolichos* (60 clones), had three to five groups of clones that shared similar band mobility. No individuals were homogenized for ITS-1 variation, as determined by SSCP analysis.

SSCP gels were produced for all *T. obsoleta* and *T. dolichos* individuals except F1OPN which was used in the ITS pilot study (see chapter 2). SSCP gel examples for *T. obsoleta* and *T. dolichos* are shown in figures 3.6 and 3.7 respectively. The frequency of ITS-1 variants were recorded from each gel and a graph was produced for each species, though no comparisons could be made between gels for these variants. Bar graphs are presented in figures 3.8 and 3.9 (*T. obsoleta* and *T. dolichos* respectively) that show the distribution of ITS-1 variants within the 12 clones from each individual. In all cases, a single common variant was present at five to ten copies and two to four rarer variants at one to four copies each from the 12 screened. This pattern of the frequency of variants within individuals was consistent in both species. Assuming that the ITS-1 clones are an unbiased sample of the variation, the common variant may be present at a level of 42% to 83% in the rRNA gene family, and rarer variants between 8% to 33%, as estimated from both graphs.

A χ^2 test was performed to determine whether three clones taken from a single PCR would sample as many variants as one clone from each of three PCR's.

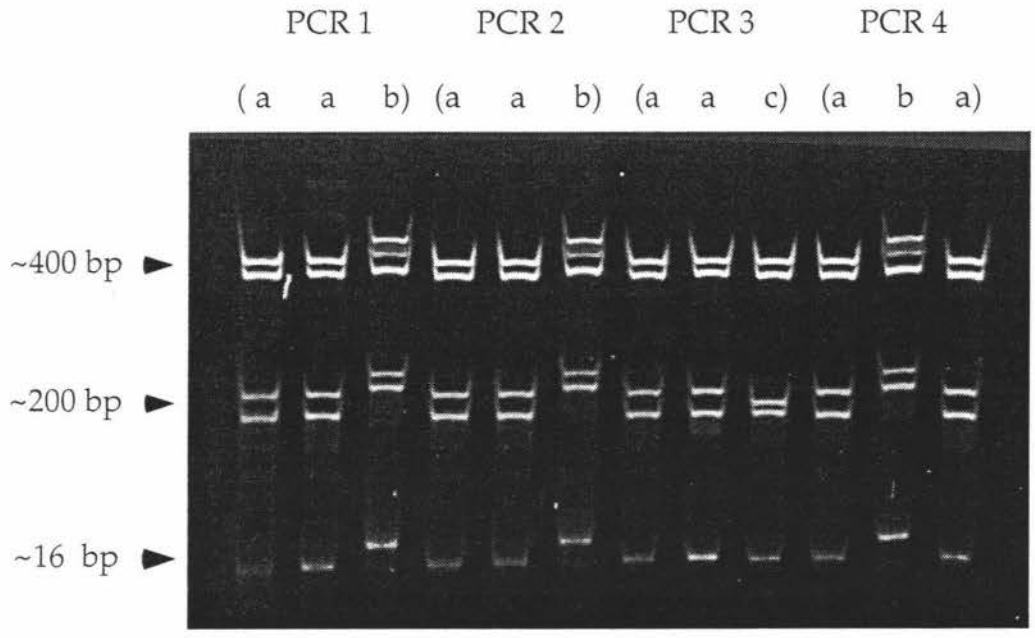

Figure 3.6 An SSCP gel of 12 cloned ITS-1 *Hae* III fragments of a single *T. obsoleta* individual (M1ORR) from four parallel PCR reactions. Clones from each PCR are shown within brackets. Three ITS-1 variants (a, b and c) were scored that share similar DSS band mobility. The common variant 'a' is represented in eight out of the 12 clones screened. The variant 'b' occurs three times, once in three separate PCR's. The rare variant 'c' occurs just once. Approximate DSS *Hae* III fragment sizes are shown.

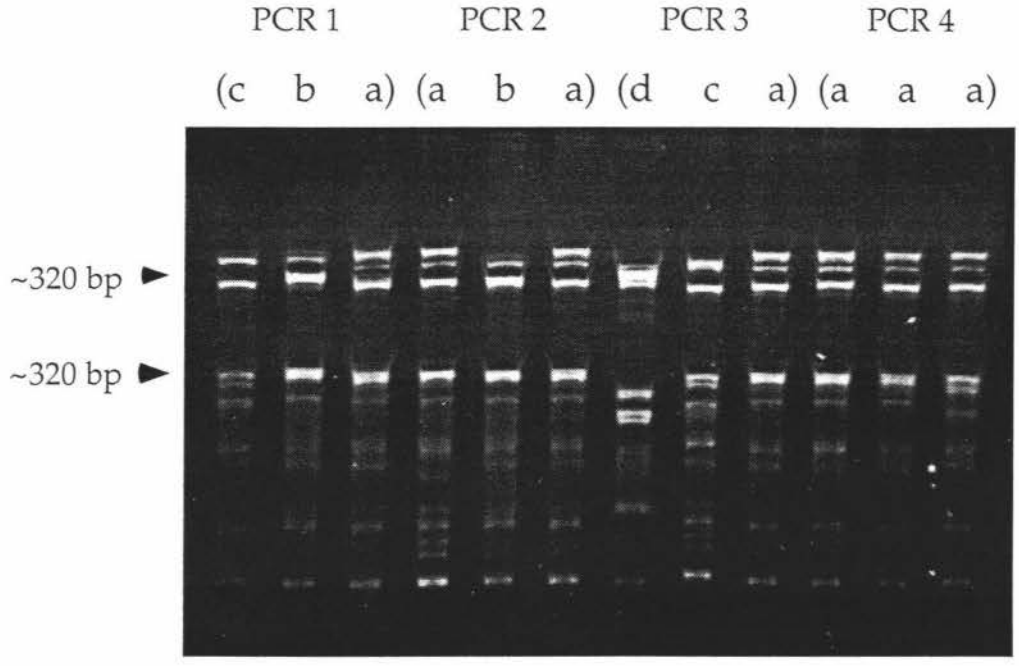

Figure 3.7 A SSCP gel of 12 cloned ITS-1 *Dde* I fragments of a single *T. dolichos* individual (M2DEr) from four parallel PCR reactions. Clones from each PCR lie within brackets. Four ITS-1 variants (a, b, c and d) were scored that share similar DSS band mobility. The common variant 'a' is represented in seven out of the 12 clones screened. Variant 'b' and 'c' occur twice, in separate PCR's. The rare variant 'd' occurs just once. Approximate DSS *Dde* I fragment sizes are shown. Note that although the RE fragments are the same size, they differ in conformation.

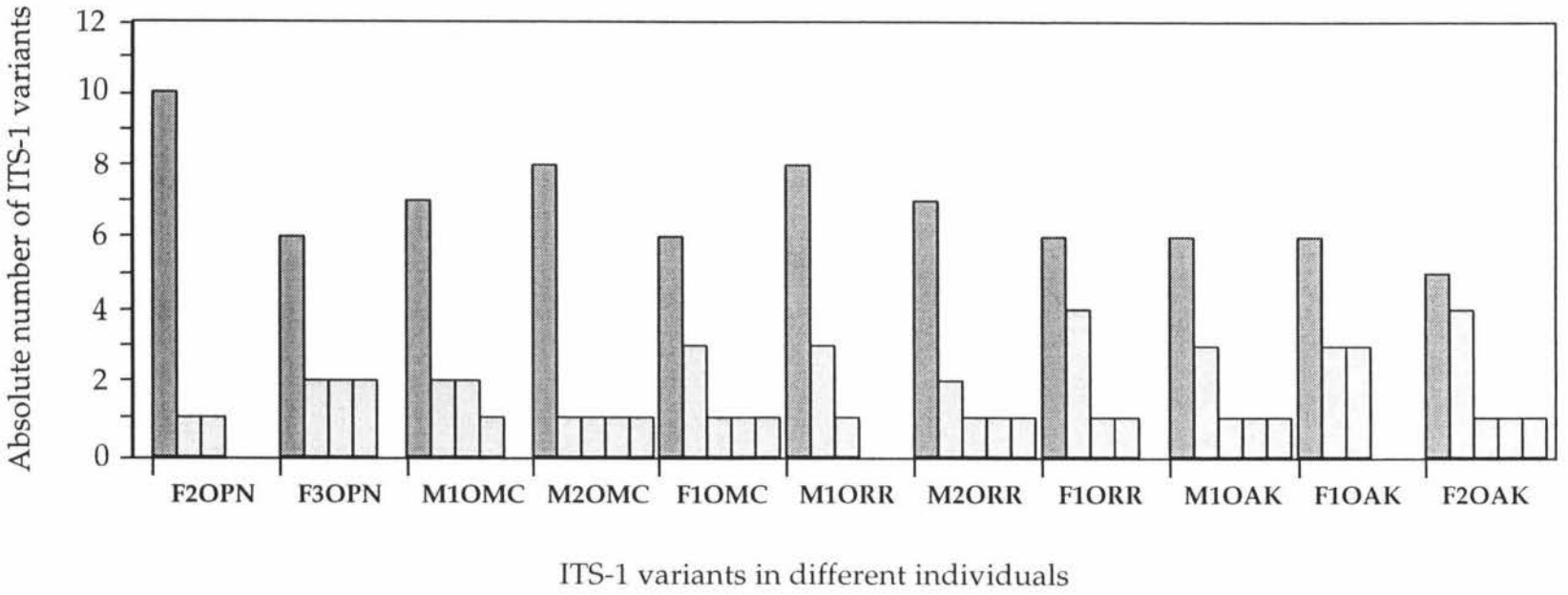

Figure 3.8 Bar graph of the frequency of ITS-1 variants scored from SSCP gels for 11 *T. obsoleta* individuals. Individual codes are shown in bold (see figure 3.1). Each individual has a single common variant ▩ and several rare variants □.

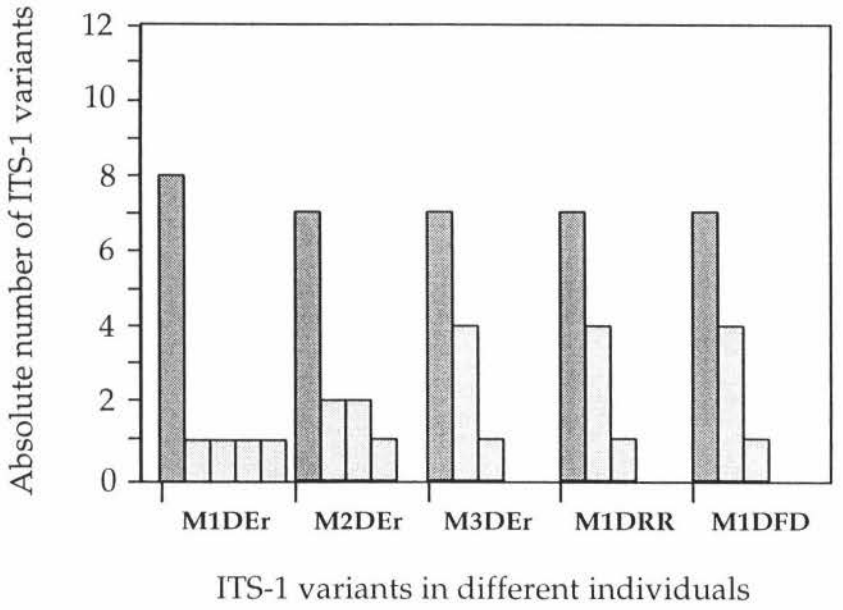

Figure 3.9 Bar graph of the frequency of ITS-1 variants scored from SSCP gels for 5 *T. dolichos* individuals. Individual codes are shown in bold (see figure 3.1). Each individual has a single common variant ▩ and several rare variants □.

The mean number of variants found in the first PCR and among the first clone of three other independent PCR's from all 16 SSCP gels was calculated (data from both species were pooled). There was no significant difference ($\chi^2 = 0.584$ $P<0.01$, 2 df) between the number of variants isolated within a single PCR reaction versus three PCR's. As many variants were isolated from three clones of a single PCR as from a single clone taken from three PCR reactions. Therefore, at this level of sampling, PCR drift was not found to be operating.

Sequence Variation

A minimum of 50 bp from both 18S and 5.8S rDNA were present in each sequence. The gene coding regions contained no variable sites within or between species therefore, all sequence analyses are based solely on the rDNA ITS-1 region. The extent of similarity between the two copies of each ITS-1 variant (sequenced from independent PCR's) was determined by DNA sequence comparison. In most cases no differences were found, but some varied by one or two sites. After alignment of all sequences, one of the two copies of these variants were removed from further analysis, if the differences were shown to be unique with respect to all other sequences. Such differences may have been the result of *Taq*-incorporated error during PCR amplification, which would inflate the level of perceived variation. Their removal did not affect the cladistic relationship of the remaining ITS-1 clones.

The multiple sequence alignment of 22 clones from *T. obsoleta* (see table 3.1) contains 429 sites, of which 80 are variable (18.6%) and 43 (10%) are phylogenetically informative. On average, the AT content was 57.7% and length heterogeneity was 388 bp to 405 bp. The multiple sequence alignment of ten clones from *T. dolichos* (see table 3.2) contains 468 sites, of which 87 are variable (18.6%) and 57 (12.2%) are phylogenetically informative. On average, the AT content was 57.4% and length heterogeneity was 437 bp to 456 bp.

No variation was found among three ITS-1 clones sequenced from a *T. cephalotes* individual (results not shown). These were 472 bp in length with an AT content of 43.4%.

The average percentage of variable sites among ten randomly selected sequences from *T. obsoleta* and the ten *T. dolichos* sequences were compared to determine if ITS-1 variability between the species differed. *T. dolichos* was more variable with 18.6% of the ITS-1 having variable sites and for the same number of *T. obsoleta* clones, this figure was just 12.2%. The ITS-1 of these two species were aligned for comparison (data not shown). Only 48% of the 482 sites were shared. The ITS-1 from all three *Triplectides* species show very little similarity, but comparison is difficult due to the presence of multiple gaps in the alignment.

For both species, a majority of the variable sites (59% for *T. obsoleta* and 57% for *T. dolichos*) involve insertion/deletion events (indels). In *T. obsoleta* a total of 20 indels occur over 46 sites, with a majority of these events involving two nucleotides. In *T. dolichos* 14 indels occur over 51 sites and a majority of these involve at least three nucleotides. Single base substitutions accounted for only 41% (32 sites) of the variation in *T. obsoleta* and 43% (38 sites) for *T. dolichos*. Both species have a transition:transversion ratio of 1:1.9.

There are no absolute statistical tests for randomness of distribution of mutations in an alignment, for all rely on arbitrarily chosen 'windows' of comparison (Strachan et al., 1985). However, mutations appear to be clustered in their distribution within the ITS-1 and contain a high proportion of the indels. These regions contain AC or TA repeat motifs and poly(T) sequences. Patterns of repeat DNA were analysed to determine whether a correlation existed between their location and the clusters of mutations in the ITS-1.

Some of these repeats are shown as highlighted regions in the multiple sequence alignments (see table 3.1 and 3.2).

In *T. obsoleta* four variable regions (7 to 15 bp in length) appear to result from replication slippage and contain the majority of indels. Two of these regions are flanked by inverted repeats which are possible base-pairing stem structures, 10 and 11 bp in length. The associated variable regions of slippage are non-base-pairing loops. Although ten more inverted repeats were identified, seven of these overlap and none are characterised by intervening regions of variation. Two dispersed direct repeats (as opposed to tandem repeats) were identified that have adjacent indels associated with them. Another three direct repeats were identified but these all overlap a larger inverted repeat and are not shown.

In *T. dolichos* three variable regions (9 to 15 bp in length) contain the majority of indels and also appear to be regions of replication slippage. One of these regions is flanked by an inverted repeat 10 bp in length. Five more inverted repeats were identified with four of these overlapping and none characterised by an intervening variable region. A single dispersed direct repeat (10 bp in length) was identified that has a large (15 bp) intervening region of indels. Two small direct repeats overlap this larger repeat. Finally, one tandem direct repeat 7 bp in length was identified, indicative of replication slippage.

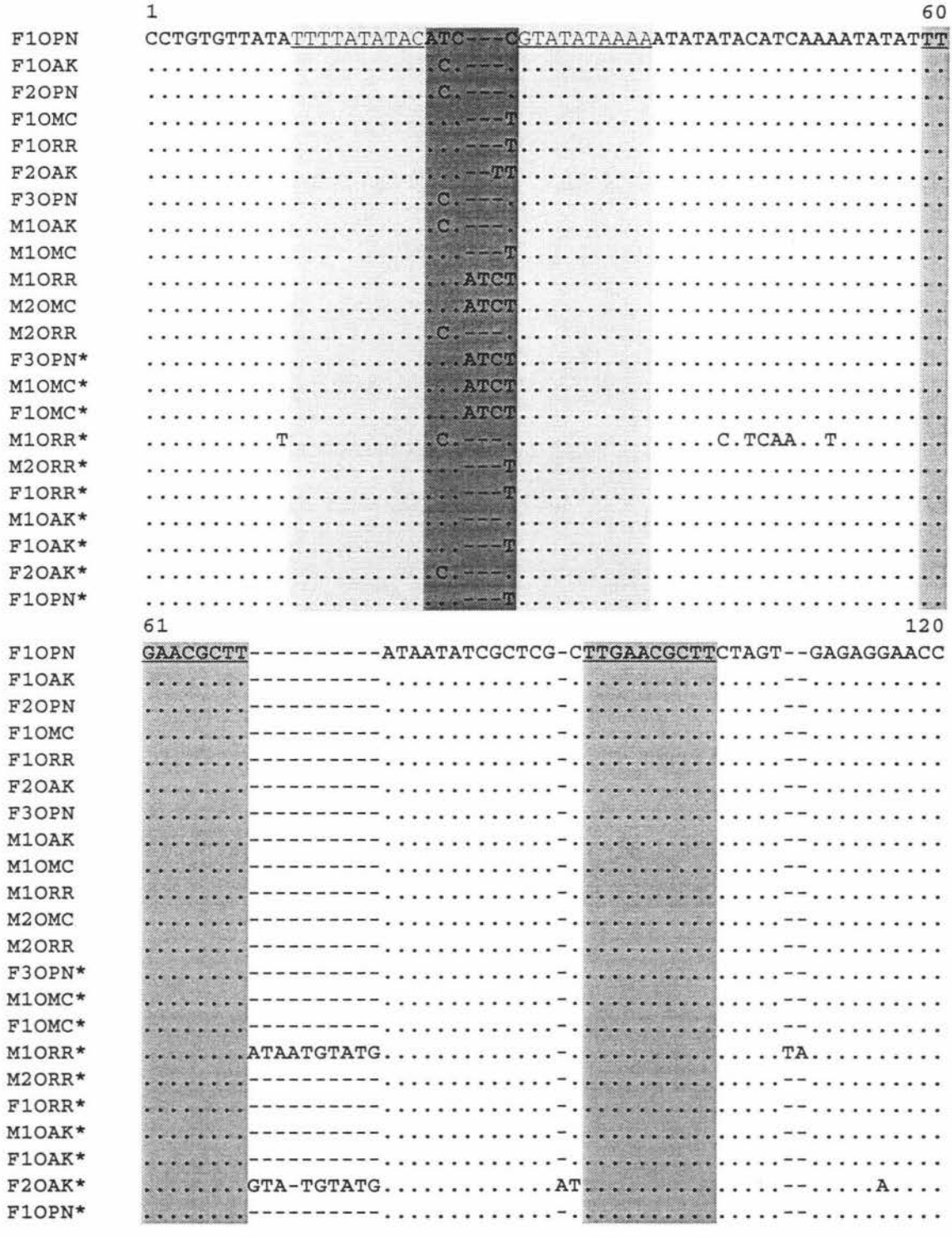

Table 3.1 Multiple ITS-1 sequence alignment (5′ to 3′) of 22 clones from 12 *T. obsoleta* individuals. An * in the individual code denotes a rare variant within an individual, all others were common. Regions of replication slippage, direct and inverted repeats are highlighted in dark grey, medium grey, and light grey respectively.

```
          121                                                                 180
F1OPN     TTCAGTGGA--CGGCCCGACAGAGCGAGGGCCCCACCGCCCCAGGGTGTGGTTCGGTGAT
F1OAK     .........-GA.....................................................
F2OPN     .........-GA.....................................................
F1OMC     .........--.....................................................
F1ORR     .........--.....................................................
F2OAK     .........--.....................................................
F3OPN     .........--.....................................................
M1OAK     .........-GA.....................................................
M1OMC     .........-GA.....................................................
M1ORR     .........-GA.....................................................
M2OMC     .........-GA.....................................................
M2ORR     .........-GA.....................................................
F3OPN*    .........--.....................................................
M1OMC*    .........-GA.....................................................
F1OMC*    .........-GA.....................................................
M1ORR*    .........--......A...............................................
M2ORR*    .........-GA.....................................................
F1ORR*    .........--.....................................................
M1OAK*    .........--.....................................................
F1OAK*    .........GA......................................................
F2OAK*    .........--.....................................................
F1OPN*    .........-GA.....................................................
          181                                                                 240
F1OPN     GAGCGGATTTCGTCGT--GTCTGTAAGGCG-GTGGTGTTTTGTCGTGGTGTGCG-AACGC
F1OAK     ................GT.........T.-.....G.............-.....
F2OPN     ................GC.........T.-.....G.............-.....
F1OMC     ................--.........-.....................-.....
F1ORR     ................--.........-.....................-.....
F2OAK     ................--.........-.....................-.....
F3OPN     ................--.........-.....................-.....
M1OAK     .G.............GTC.G.AAG...T.-.....G.............-.....
M1OMC     T...............GT.........T.-.....................-.....
M1ORR     ................GC.........T.-.....G.............-G....
M2OMC     ................GC.........T.G.....G.............-.....
M2ORR     ................GT.........T.-.....CG............-.....
F3OPN*    ................--.........-.....................-.....
M1OMC*    ................GC.........T.-.....G.............-.....
F1OMC*    ................GC.........T.G.....G.............-.....
M1ORR*    ................--.........T.-.....G.............-.....
M2ORR*    ................GT.........T.-.....G.............-.....
F1ORR*    ................--.........-.....................-.....
M1OAK*    ................--.........-.....................G.....
F1OAK*    ................--.........-.....................-.....
F2OAK*    ................--.........T.-.....G.............-.....
F1OPN*    ................GT.........T.-.....G.............-.....
```

Table 3.1 (continued) Multiple ITS-1 sequence alignment (5′ to 3′) of 22 clones from 12 *T. obsoleta* individuals. An * in the individual code denotes a rare variant within an individual, all others were common. Regions of replication slippage, direct and inverted repeats are highlighted in dark grey, medium grey, and light grey respectively.

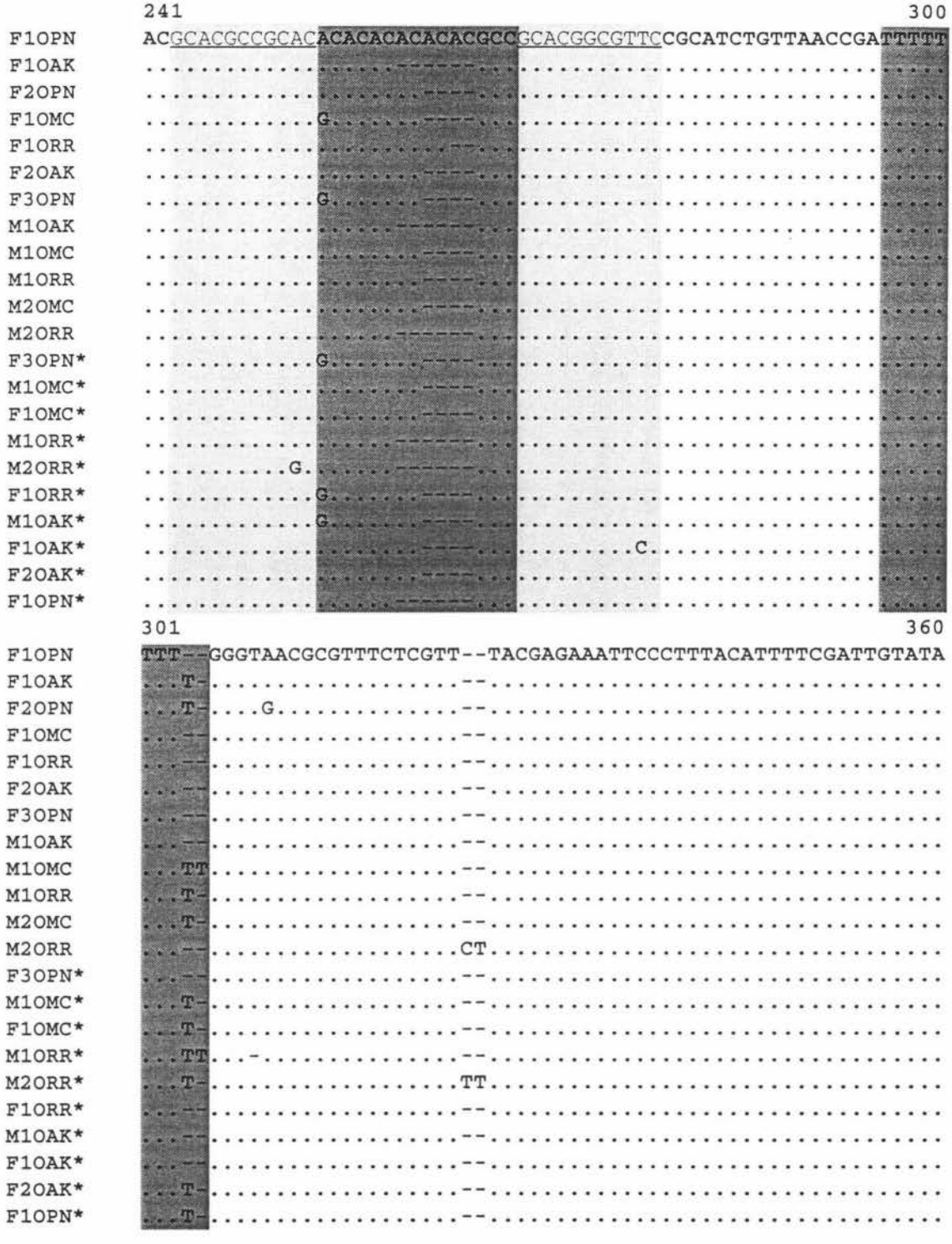

Table 3.1 (continued) Multiple ITS-1 sequence alignment (5′ to 3′) of 22 clones from 12 *T. obsoleta* individuals. An * in the individual code denotes a rare variant within an individual, all others were common. Regions of replication slippage, direct and inverted repeats are highlighted in dark grey, medium grey, and light grey respectively.

```
          361                                                                 420
F1OPN     TAGAATATGAAATAATTTTTTT-TTT-----ATATAATATCTGTTGATTTT--ACAACATG
F1OAK     ..............-...-...TT---.....................--........
F2OPN     ..............-...-...T--AT.....................--........
F1OMC     ..............-...-...-----.....................--........
F1ORR     ..............-...-...-----.....................--........
F2OAK     C.............-...-...T----.....................--........
F3OPN     ..............-...-...TTT--.....................--........
M1OAK     C.............-...-...T----.....................--........
M1OMC     ..............-...-...T----.....................--........
M1ORR     ..............-...-...T----..-..................--........
M2OMC     ..............-...-...T----.....................--........
M2ORR     ..............-...-...TT---.....................--........
F3OPN*    ..............-...-...-----.....................--........
M1OMC*    ..............-...-...T----.....................--........
F1OMC*    ..............-...-...T----.....................--........
M1ORR*    ..............A...-...TTTTT.....................--........
M2ORR*    ......G.......-...-...TT---.....................--........
F1ORR*    ....G.........-...-...TT---.....................--........
M1OAK*    ......G.......-...-...-----.....................TT........
F1OAK*    C.............-...-...T----...C.................--........
F2OAK*    ..............-...-...TT---.....................--........
F1OPN*    ..............-...-...TT---.....................--........
          421   429
F1OPN     AAAAATCAT
F1OAK     .........
F2OPN     .........
F1OMC     .........
F1ORR     .........
F2OAK     .........
F3OPN     .........
M1OAK     .........
M1OMC     .........
M1ORR     .........
M2OMC     .........
M2ORR     .........
F3OPN*    .........
M1OMC*    .........
F1OMC*    .........
M1ORR*    ..C......
M2ORR*    .........
F1ORR*    .........
M1OAK*    .........
F1OAK*    .........
F2OAK*    .........
F1OPN*    .........
```

Table 3.1 (continued) Multiple ITS-1 sequence alignment (5′ to 3′) of 22 clones from 12 *T. obsoleta* individuals. An * in the individual code denotes a rare variant within an individual, all others were common. Regions of replication slippage, direct and inverted repeats are highlighted in dark grey, medium grey, and light grey respectively.

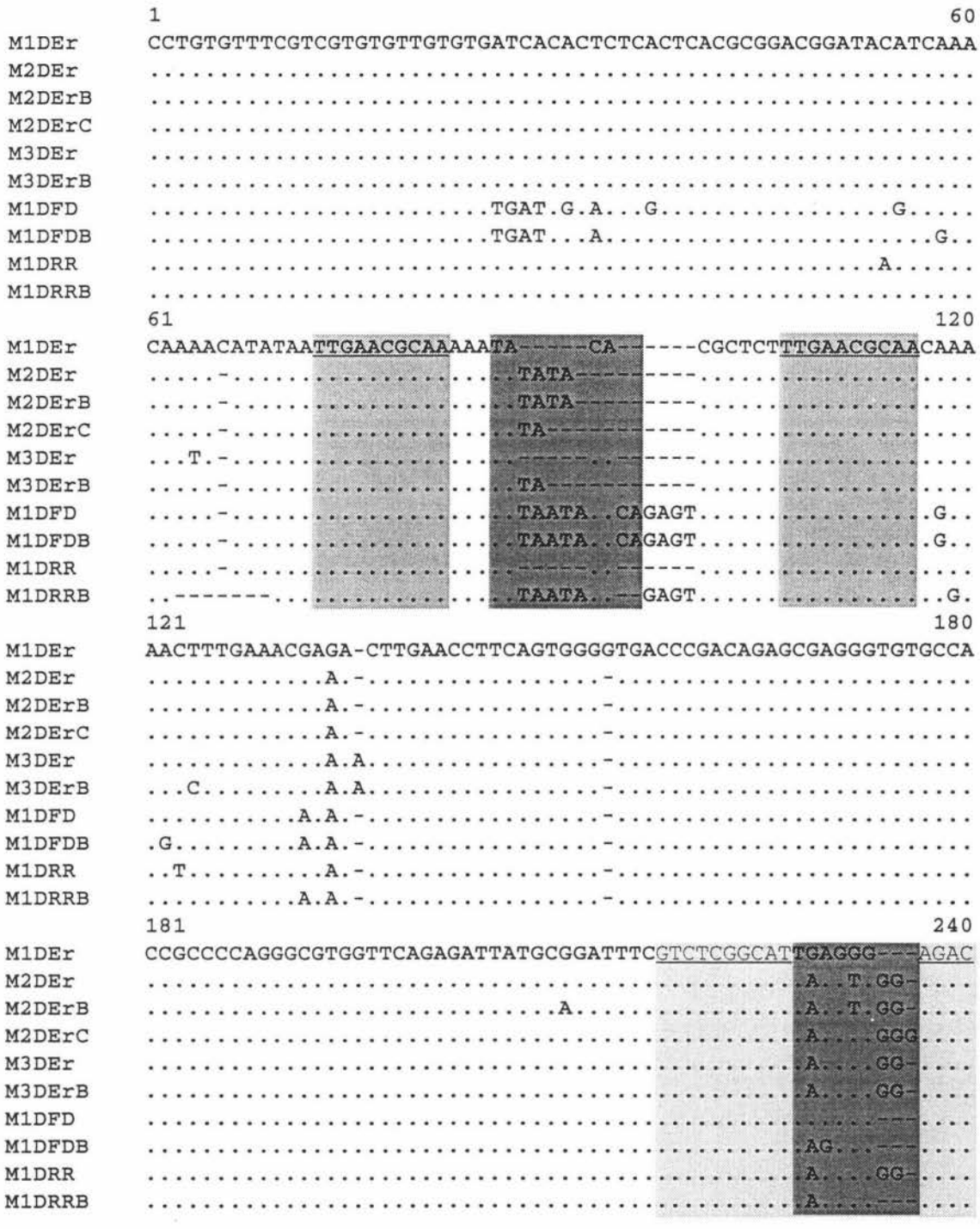

Table 3.2 Multiple ITS-1 sequence alignment (5′ to 3′) of 10 clones from 5 *T. dolichos* individuals. A 'B' or 'C' and the end of an individual code denotes a rare variant within an individual, all others were common. Regions of replication slippage, direct and inverted repeats are highlighted in dark grey, medium grey, and light grey respectively.

```
          241                                                                 300
M1DEr     GCTGAGATATCGCTGGCATTCCCGATTGCGGAGTGCGCGATTTTTCGCGTTTCT--CGTC
M2DEr     ..........................................A.........--....
M2DErB    ..........................................A.........--....
M2DErC    ....................................................--....
M3DEr     ....................................................--....
M3DErB    ..................................................CT....
M1DFD     ....................................................--..C.
M1DFDB    ....................................................--..C.
M1DRR     ....................................................--....
M1DRRB    ....................................................--....
          301                                                                 360
M1DEr     A-GGCGTTCCGCTATAATTTTTAATTTTGAACGAACTAAAGGGT---------TCTCTCT
M2DEr     .A.....................................................ACACACGTT.......
M2DErB    .A.....................................................ACACACGTT.......
M2DErC    .-.....................................C.......ACACACGTT.......
M3DEr     .-.....................................---------.......
M3DErB    .-.....................TT.....................ACACACGTT.......
M1DFD     .-.....................TT..--.................GCACACGTT.......
M1DFDB    .-.................A...TT..--.................GCACACGTT.......
M1DRR     .-.....................................---------.......
M1DRRB    .-.....................TT...-........C.......ACACACGTT.......
          361                                                                 420
M1DEr     GGAGAAATTCCCTAAACATTTTTTTATGTATATAGAATTTACAATGTTTTTATAAAAA--
M2DEr     ..................................................................AA
M2DErB    ..................................................................AA
M2DErC    ..............................G...................................AA
M3DEr     ..................................................................AA
M3DErB    ......................C...........................................AA
M1DFD     ......G.....................................................T..T...AT
M1DFDB    ..........................................................T..T...AT
M1DRR     ..................................................................AA
M1DRRB    ..................................................------..T...AT
          421                                                       468
M1DEr     TTTTTGAACGCTTTCGTGTGATGATA-TATTATCGCGC-GTAAATCAT
M2DEr     --...........C........A.-.-----------.---------
M2DErB    --...........C....C...A.-.-----------.C........
M2DErC    --.....................A.-.-----------.---------
M3DEr     A-.....................A.-.-----------.---------
M3DErB    A-.....................CA.-.-----------.---------
M1DFD     ..........................A...---.........-.---------
M1DFDB    ..........................A...A...........-.---------
M1DRR     A-.........................-...........-.---------
M1DRRB    ..........................A...----..G.......-.C.......
```

Table 3.2 (continued) Multiple ITS-1 sequence alignment (5′ to 3′) of 10 clones from 5 *T. dolichos* individuals. A 'B' or 'C' and the end of an individual code denotes a rare variant within an individual, all others were common. Regions of replication slippage, direct and inverted repeats are highlighted in dark grey, medium grey, and light grey respectively.

Phylogenetic analysis

Phylogenetic analysis by parsimony was performed using the exhaustive search and heuristic search methods in PAUP v 3.1.1 (Swofford, 1990). The analysis was used to determine whether variation in ITS-1 is restricted at the individual and population level. A monophyletic grouping of clones from individuals and populations would indicate such a restriction and provide evidence of homogenization. An initial analysis was performed to root the trees for both species using an ITS-1 sequence from the other species. Once the root was found, this was added to trees constructed without the outgroup in order to make the ingroup monophyletic.

A comparison was made between trees built with topological constraints; clones from individuals or clones from populations were forced to be monophyletic, and trees without constraints. If ITS-1 evolution followed the pattern set by the constraints, then constrained and unconstrained trees should be of similar length. The parameters used were as follows; gaps in the sequence alignment were treated as missing data, thereby excluding these characters from the tree construction, and branches of zero length were collapsed to yield polytomies.

A maximum parsimony (MP) analysis using the heuristic procedure with 100 random replicates was performed on 22 aligned ITS-1 sequences from *T. obsoleta*. Trees constructed with no constraints were of length = 55, a consistency index (CI) of 0.80, and a rescaled consistency index (RC) of 0.72. A 50% majority-rule consensus of 32,700 trees is shown in figure 3.10. The deeper clades are more defined than terminal clades, which place 10 of the 22 clones into undefined polytomies. All 10 individuals that had two clones present were polyphyletic and none of the four geographic locations formed a monophyletic clade. Trees constrained to group clones to individuals were of length = 86 and CI = 0.51. Trees constrained to group populations were of length = 76 and CI = 0.58.

A rare ITS-1 variant may be common or rare in another individual from the same or different population. For example, the common variant from two individuals M1ORR and M2OMC that are geographically separated by ~500km differ in just 3 sites of the ITS-1 alignment. In comparison, the common and rare variants of individual M1ORR differ by 41 sites. The absence of a monophyletic grouping of individuals and populations indicates a lack of homogenization within both individuals and populations.

A maximum parsimony (MP) analysis using an exhaustive search was performed on 10 aligned ITS-1 sequences from *T. dolichos*. With no constraints, 13 equally parsimonious trees of length = 47, CI of 0.94, and RC of 0.846 were found. A 50% majority-rule consensus of the 13 trees is shown in figure 3.11. Half (5) of the clones form a basal polytomy. Only the two clones from individual M1DFD are monophyletic. All other individuals and the only 'population' present (three Erua individuals) do not form a resolved monophyletic group. Trees constrained to individuals were of length = 55 and CI = 0.80. When the 6 clones from the Erua population were constrained to a single clade, trees were of length = 54 and CI = 0.82. These results are similar to *T. obsoleta*.

A comparison was also made to two other methods of treating gaps in the alignments. Firstly, each gap was treated as an independent character, thus increasing the relative weight of gaps in the phylogenetic analysis (Volger & DeSalle, 1994). For *T. obsoleta*, this resulted in 89 trees of length = 113, CI = 0.752, and RC = 0.611, and for *T. dolichos*; 2 trees of length = 106, CI = 0.887, and RC = 0.777. Secondly, each gap was treated as a single character (a single mutational event), regardless of the length of the gap. If the length of a gap was not identical in all sequences, then each length variant was coded as a single character. For *T. obsoleta*, this resulted in 18 trees of length = 78, CI = 0.795, and RC = 0.652, and for *T. dolichos*; 1 tree of length = 72, CI = 0.903, and RC = 0.771.

Trees constructed by the method 'gaps treated as missing data' had similar homoplasy measurements (CI and RC) and topologies to trees constructed by the two methods that included gaps as characters. This indicates that there is phylogenetically useful information contained in the gaps. Furthermore, characters involved in substitutional and indel changes showed similar high character consistency.

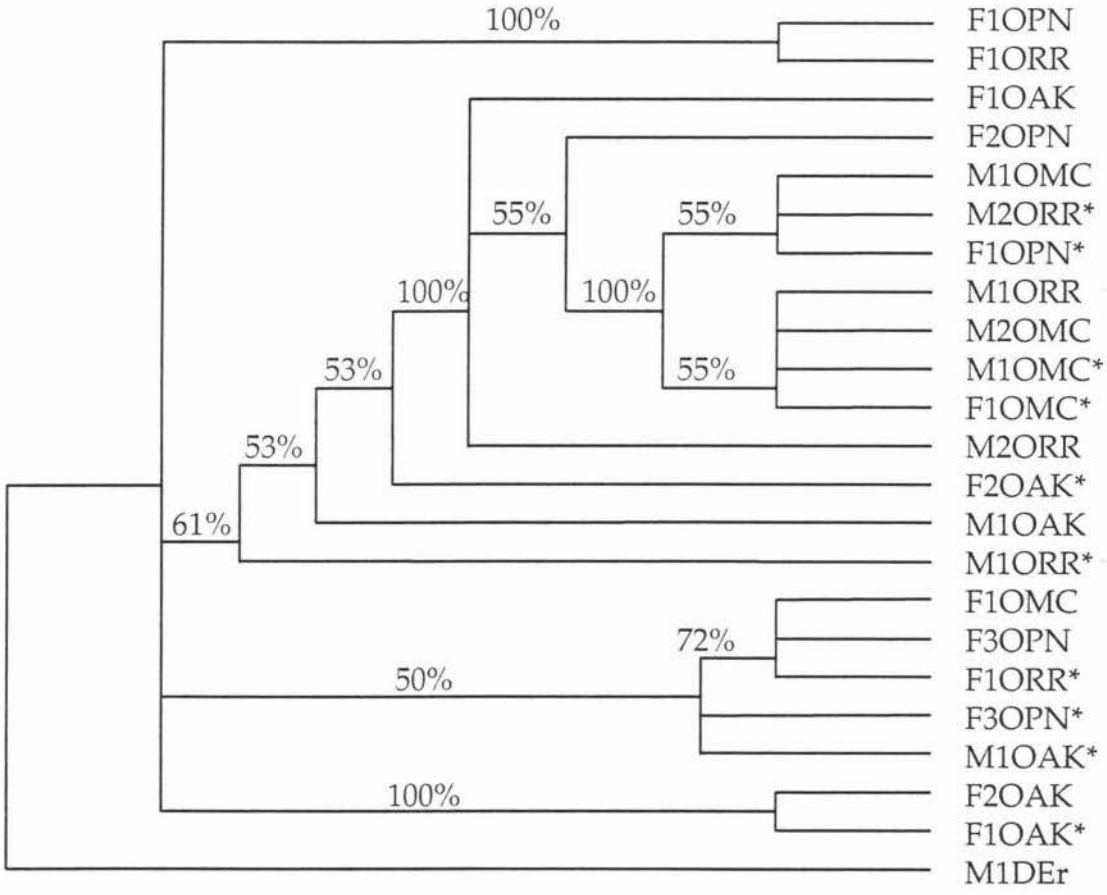

Figure 3.10 The MP tree (50% majority-rule consensus) of 32,700 trees for *T. obsoleta*, rooted to T. dolichos (individual M1DEr). Based on the alignment of 22 ITS-1 clones (table 3.1). Topological constraints not enforced. Tree length = 55 and CI = 0.80. Rare ITS-1 variants are denoted with an * at the end of the individual code.

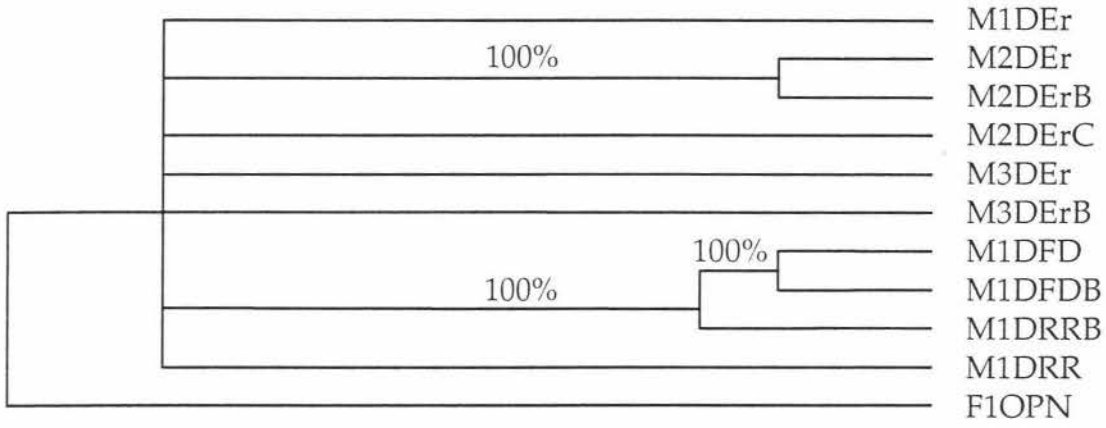

Figure 3.11 The MP tree (50% majority-rule consensus) of 13 trees for *T. dolichos*, rooted to *T. obsoleta* (individual F1OPN). Based on the alignment of 10 ITS-1 clones (table 3.2). Topological constraints not enforced. Tree length = 47. CI = 0.936. Rare ITS-1 variants are denoted with a 'B' or 'C' at the end of the individual code.

3.4 Discussion

SSCP: the frequency of variants

The presence of multiple ITS-1 variants from all individuals in both species indicates that they are heterogeneous for ITS-1 variation. In an extensive study of the sensitivity of the SSCP technique, Hayashi (1991) concluded that 99% of point mutations in DNA fragments 100-300 bp in length and 89% of mutations in fragments 300-450 bp in length could be detected. Although groups of variants scored within gels as the same differed at most by 2 bp, variation between groups could be under-estimated by the presence of larger restriction enzyme fragments (400-450 bp). From Hayashi's (1991) work, 400 bp fragments differing by up to 40 bp could have the same conformation (band mobility) in an SSCP gel. In figure 3.6 there is a clear difference in the conformation of the 400 bp fragments between variant 'a' and 'b'. The DNA sequence of these showed they differ by only 27 bp in this region. Therefore, the sensitivity of the SSCP technique may be greater than thought.

The absence of PCR drift suggests that similar levels of ITS-1 variation can be found within a single individual by sampling either multiple clones from single PCR reactions or single clones from multiple PCR reactions. Screening fewer than 12 clones per individual may reduce the number of variants thought to be present within an individual. However, intra-individual ITS-1 heterogeneity can also be detected as long as multiple individuals are screened. Wagner et al. (1994) suggest that pooling the products from multiple PCR reactions would improve the sampling of a gene family. However, the accuracy of a particular gene sequence can only be guaranteed by the re-isolation of the identical sequence from an independent PCR amplification (Zylstra et al., 1998). In this regard, the parallel sampling carried out in this study was superior to Wagner et al.'s (1994) suggestion of cloning from a pool of multiple PCR reactions.

A common (>50%) ITS-1 variant was present in all individuals sampled and between two to four rarer variants were also present. Although only 12 ITS-1 copies were isolated from each individual, an estimate of the copy number of these variants in the haploid genome may be made by comparison to other insects. Using dot-blot hybridization, Kumar and Rai (1990) surveyed the rRNA genes of 13 mosquito species and estimated the average copy number per haploid genome to be 440, when averaged over the species. Polanco et al. (1998) stated that ~250 copies are present in *Drosophila melanogaster*. Long and Dawid (1980) estimate ~240 copies are present in the silkworm moth, *Bombyx mori*, and 100 copies in the fly *Chironomus tentans*. Given the above figures and the assumption that the ITS-1 clones were an unbiased sample of the genome, approximately 2.7% to 12% of the copies present in the gene family may have been sampled from the two *Triplectides* species. Employing the SSCP technique meant that rarer variants were given a higher chance of being isolated, therefore enabling a more representative sample of the variation present in the gene family.

Sequence heterogeneity and replication slippage

The type of variation was very similar in these presumed sister taxa with neither being homogenized for ITS-1 variation. In both cases, length heterogeneity was similar (*T. obsoleta*: 17 bp and *T. dolichos*: 19 bp) and approximately 19% of sites within aligned sequences were variable. Transitions and transversion are close to a 1:2 ratio in both species, which is expected when equal probability of change for all nucleotides is assumed (Volger & DeSalle, 1994). A majority of the variable sites are indels containing contiguous short repeats which are clustered in their distribution. This type of variation is considered to arise by replication slippage.

Slippage can occur because of mispairing between neighbouring repeats during DNA replication but can also occur in nonreplicating DNA (Li & Graur, 1991). Surveys of sequence databases show that the frequencies of different simple sequences in non-coding regions of eukaryotes are very similar between species and this reflects a bias inherent in replication slippage of short AT-rich motifs (Hancock, 1996). In both species, regions of proposed slippage have a preponderance of AC-rich or AT-rich motifs, and poly(T) sequences (see tables 3.1 & 3.2). This is reflected in the fact that the ITS-1 of both species is AT-rich (~57%) though this is not so for their more distant relative, *T. cephalotes*, which is GC-rich (56.6%). However, slippage cannot be a general feature of ITS evolution, as the AT-rich ITS of *D. melanogaster* does not show elevated levels of sequence simplicity due to replication slippage (Torres et al., 1990; Tautz et al., 1988).

There are two broad categories of sequence simplicity; regions of pure simple sequence (pure simplicity) and regions of less regularity but with a clear bias in nucleotide composition (cryptic simplicity) (Tautz et al., 1986, 1988). Sequence repetition in these regions is over and above that expected by chance alone. The presence of numerous direct and inverted repeats (7 to 10 bp in length), the repeat DNA in regions of slippage, and AT richness indicate that the ITS-1 of *T. obsoleta* and *T. dolichos* may be a region of cryptic simplicity. Simple repetitive sequences are considered to represent a natural ground state of DNA unselected for coding functions such as the ITS-1 (Levinson & Gutman, 1987).

Two likely consequences of slippage events between noncontiguous (dispersed) sequences is the formation of palindromes and the deletion of sequences between direct repeats (Levinson & Gutman, 1987). It is thought that the secondary structure formed by a palindrome brings direct repeats into close proximity, allowing deletion of DNA between repeats during DNA replication; an hypothesis called palindome-stimulated replication slippage (Pinder et al.,

1998). In the sequence alignments, three regions of slippage are flanked by dispersed inverted repeats which may form palindromes (see tables 3.1 & 3.2). In addition, there are two cases (one in each species) in which dispersed direct repeats have an intervening region characterized by indels, with deletions being more common in that region of the alignment.

T. obsoleta and *T. dolichos* were chosen for this study because they are presumed sister species and because they are likely to have different population structures due to habitat differences. The spatial distribution of the habitat of *T. dolichos* (seepages) is more isolated than that of *T. obsoleta* (rivers) (see chapter 1). Although the type of ITS-1 variation is very similar between these two species, *T. dolichos* has a higher percentage of variable sites in the aligned sequences. This factor may be related to differences in population dynamics, such as the effective population size which may affect the fixation of variants (see chapter 4). This result was not otherwise expected and deserves further investigation.

Paralogous sequences

Non-diagnostic variation exists among the ITS-1 sequences of *T. obsoleta* and *T. dolichos*. Variation was not restricted among individuals nor among populations. An ITS-1 variant predominating within an individual can be common or rare in another, either from the same or different population. The heterogeneity among the clones cannot be casually rejected. Moreover, in this case, the heterogeneity precludes the reliability of ITS-1 as a genetic marker for a population study. The monophyletic grouping of clones from a population would have provided strong evidence for homogenization (Sanderson & Doyle, 1992). The absence of such groups indicates a lack of homogenization in both species. The paralogous ITS-1 copies do not behave as a single orthologous sequence. Therefore, the MP trees (figures 3.10 & 3.11) should be interpreted as 'gene trees' to graphically display the paralogy among these sequences (Rich et al., 1997).

Other studies

This study represents a clear case of ITS-1 heterogeneity, a result that has been recorded in other studies. Significant intra-individual variation exists in the ITS-1 of the *Simulium damnosum* species complex (black flies) with the degree of homogenization varying among the sibling species (Tang et al., 1996). In two deer tick species (*Ixodes*), intra-individual variation in the ITS-2 is thought to result from the maintenance of ancestral polymorphisms which hinder the phylogenetic analysis among ticks of geographically disparate populations (Rich et al., 1997). Other studies indicate the presence of high levels of homogenization. Putative cryptic species of the mosquito *Anopheles nuneztovari* were unable to be distinguished by ITS-2 sequence data because of a lack of variation. Low intra-specific variation has also been demonstrated in both ITS-1 and ITS-2 among other mosquitoes species (*Aedes*) and phylogenetic analyses of these support traditional morphological classification (Wesson et al., 1992).

It is difficult to make comparisons between the above studies because of differences in sampling strategies. In this study, SSCP analysis allowed for a more efficient means of maximizing the variation present. Intra-individual and intra-specific variation is often not assayed as thoroughly as it perhaps should be, perhaps due to the dogma of homogenization (Williams et al., 1988). Differences in the results of studies of ITS variation may also be due to variable rates of homogenization. The result of studies demonstrating ITS heterogeneity should be taken as evidence that high rates of homogenization cannot be assumed. Therefore, sampling should be more of a concern for studies utilizing highly variable portions of the rDNA, such as the ITS-1 (Rich et al., 1997).

Chapter 4

Evolution of Repeat DNA: 'the Nature of the Variation' in ITS-1

Modes of evolution

The chance or fortuitous nature of the origin of both genetic and phenotypic variation is an inherent feature to both contemporary neo-Darwinian and neutralist concepts. However, chance alone cannot explain the non-independent (concerted) evolution of families of repeat DNA sequences. It is by the process of molecular drive that mutations are able to spread through a gene family (homogenization) and through a sexual population (fixation) as a consequence of DNA turnover (Dover, 1986). Molecular drive operates on both protein coding and non-coding repetitive DNA and has the potential to affect the evolution of organisms in decidedly non-Mendelian and non-Darwinian ways (Dover, 1982). In studies of molecular drive, the demonstration of the processes which generate variation are primary in explaining evolution; accidental variation and selective advantage are typically relegated to a minor role (Ho & Saunders, 1984).

Natural selection, neutral drift, and molecular drive are considered by some authors to represent 'modes' of evolution which interact to some degree (Dover, 1982). Others have suggested that molecular drive is not a mode of evolution, but rather an extension of modes (selection and drift) that already exist (Williams, 1990). However, the processes inherent in molecular drive are operationally different from those of selection and drift. The latter operate within Mendelian populations and represent adaptive and non-adaptive modes

of evolutionary change, respectively (Dover, 1982). It is common for workers to discuss the action of both selection and molecular drive in considering gene family variation. In this chapter I discuss whether such a pluralistic approach is tenable given their inherent differences.

ITS-1 variation in two caddisflies

In this study, rDNA ITS-1 copies were isolated from two presumed sister species; *Triplectides obsoleta* and *T. dolichos*. A representative sample of ITS-1 variation was assayed at the individual and population levels, via a sequence screening technique (SSCP) with cloned ITS-1 copies. Parallel sampling was employed to offset the possibility of PCR-drift and to guarantee the isolation of genuine sequences. These were aligned to determine sequence variation and their cladistic relationships via parsimony analysis.

The nature of the ITS-1 variation was very similar in both species. The ITS-1 is AT-rich, which suggests a bias in the establishment of new variation, and contain repeat DNA; characteristic of cryptically simple non-coding DNA. Variable sites accounted for nearly a fifth of the ITS-1. A majority of this variation is due to indels and the remainder to substitutions. This variation appears to have arisen via replication slippage and point mutation. Intra-individual variation is also high, with multiple clones taken from an individual sometimes containing as much variation as was detected in the whole species. The two most striking forms of variation found in both species, are the presence of *extensive ITS-1 heterogeneity* and the *absence of population-specific variation*. These two forms of variation are discussed below in the context of different explanations offered by neo-Darwinism and molecular drive.

Natural selection

ITS-1 is not expressed. Hence it is unlikely that natural selection acts on this region directly, though it has been argued that it is the indirect subject of

selection (Williams et al., 1985; Ridley, 1993). Transcribed spacers (ETS, ITS-1 and ITS-2) are indispensable for pre-rRNA processing. It has been demonstrated in yeast that deletions from the 5' end of ITS-1 abolish the production of mature 17S rRNA (Musters et al., 1990). The maintenance of secondary structure is also considered to be of functional importance because it has been demonstrated that both intraspecific and interspecific variable sites are concentrated in non base-pairing structures (McLain et al., 1995; Schlotterer et al., 1994; Wesson et al., 1992). This is taken as evidence for functional constraints and in turn, of some selective (deterministic) force which is responsible for its stability. In presumed cases of minor selective constraints, sequences are said to experience 'relaxed selection'. For example, the heterogeneous ITS-2 of deer ticks (*Ixodes spp.*) are said to experience relaxed selection for sequence conservation (McLain et al., 1995).

The general view is that selection does not operate at the level of the individual repeat because the accumulation of independent mutation cannot account for the level of similarity among all members of a gene family (Li, 1997). Despite this, natural selection is often thought to affect concerted evolution by acting at some higher level. For example, it has been suggested that selection may act at the chromosome level, rather than on individual repeats (Polanco et al., 1998).

New types of selection are sometimes inferred to explain the discovery of new phenomena (Lima-de-Faria, 1988). For example, 'drive intensified selection' is said by McLain et al. (1995) to impede fixation of pre-rRNA processing-ineffective mutants at some level of homogenization, but would not hinder the random fixation of one among a subset of processing-effective rDNA units that posses a long-term, 'ultimate' advantage.

Directional selection is variously thought to limit variation by maintaining high homogeneity among repeat members within a species (negative selection), or

increase variation by accelerating the differentiation among repeat members between different species (positive selection) (Li, 1997; Hoelzel & Dover, 1991). The maintenance of base pairing by the production of compensatory base mutations or a bias towards the presence of one repeat over others is considered evidence for positive selection (Wheeler & Honeycutt, 1988). The existence of a bias toward the continual accumulation of simple repeat DNA is evidence for negative selection (Levinson & Gutman, 1987). This form of 'push-me-pull-you' selection is consistent with the heterogeneity present in this study.

Stabilizing selection (Simpson, 1953), is the state where a number of genotypes are maintained because together their average fitness value is higher than any smaller set of genotypes. If this could somehow be demonstrated among the ITS-1 variants, then stabilizing selection could explain the heterogeneity found in this study and why variants are not restricted geographically.

Molecular drive

Predictions based on mathematical models imply that gene families in populations that are relatively isolated genetically should undergo independent concerted evolution (Elder & Turner, 1995). In this study, molecular drive was not seen to be acting in a population-specific manner. However, another genetic marker is required to determine whether the caddisfly populations employed here are truly genetically distinct. SSCP analysis and sequencing demonstrated that ITS-1 variants were not at the same frequency among individuals of a population. This is despite Ohta and Dover's (1984) expectation that within a population, individuals should vary little in average copy number or sequence variation for a given repeat family. Individual genomes are expected to evolve in unison relative to one another in the population, just as individual repeat units are seen to evolve in unison within genomes, provided homogenization is strong (Elder & Turner, 1995).

However, under some parameters the variance-to-mean ratio of the copy number of a gene family variant (repeat type) in a population can change over the time to fixation. Both large population size and large gene family size are considered to affect the variance of repeat types as they spread throughout a population (Ohta and Dover, 1984). Under simulation, models of concerted evolution predict a small relative variance in families of 50 to 200 members, under what are considered realistic rates of DNA turnover (Ohta, 1983; Ohta and Dover, 1984). Both *Triplectides* species studied are common aquatic insects and may have very large populations. By comparison to closely related insects from which the rDNA copy number have been estimated, it is possible that the rDNA copy number of *T. obsoleta* and *T. dolichos* exceeds 200 (see chapter 3).

The theory of molecular drive does predict that some heterogeneity can exist within gene families. This is because the degree of homogeneity found in a tandemly repetitive sequence depends on the balance between the rate of homogenization and the rate of new mutations (Ohta, 1983; Schlotterer et al., 1994). For example, it has been demonstrated that the generation and reshuffling of variation by point mutation and replication slippage can slow down homogenization (Tautz et al., 1986). In this study, slippage and point mutations are likely to account for variation present in the sequence alignments. Slippage and point mutation may be occurring faster than the rate of homogenization, in which case this could account for the ITS-1 heterogeneity.

Although Dover (1982) emphasized non-Mendelian inheritance within and among chromosomes, reciprocal recombination events during sexual reproduction can also increase the variance among repeat types within a population (Williams, 1990; Ganley, pers. comm). Little has been done to assess the relative roles that mitotic and meiotic recombination play in concerted evolution (Ganley & Scott, 1998). It has been suggested that meiosis can actually slow down or prevent homogenization thereby maintaining variation.

In a study of IGS (rDNA) diversity in parthenogenic greenbugs (*Schizaphis graminum*), Shufran et al. (1997) suggest that periodic sexual reproduction is a primary mechanism for the maintenance of genetic variability in populations. Meiosis may slow the rate of homogenization that would occur with mitosis alone (Elder & Turner, 1995; Ganley & Scott, 1998). That repeat DNA families of asexual organisms also undergo concerted evolution demonstrates the relative importance of mitosis in concerted evolution (Ganley, pers. comm). It has also been demonstrated that non-Mendelian segregation in rDNA copy number occurs prior to cell division (premeiotic) in crosses of pink mould (*Neurospora*) (Butler & Metzenburg, 1989). However, challenging the role of cell division in mutation and DNA turnover remains problematic, because the idea that cell division must occur first has become an entrenched paradigm (Blanden et al., 1998; Bridges, 1997).

Homogenization may occur faster between homologous chromosomes than between nonhomologous chromosomes (Dover, 1982). In *Drosophila melanogaster* the rDNA arrays that occur on both X and Y chromosomes have diagnostic sequence differences, indicating a higher rate of homologous chromosome exchange (Williams et al., 1985, 1987). It is possible that the rDNA of *Triplectides* occurs on two or more chromosome pairs. The absence of diagnostic sequence differences between the sexes (in *T. obsoleta*) indicates these may be on autosomes. However, ITS-1 heterogeneity among the *Culex pipiens* species complex occurs despite rDNA being clustered on just one chromosome pair (Miller et al., 1996).

Forms of explanation

Both molecular drive and selection (stabilizing selection in particular) are able to explain the low level of homogenization and the variance in the frequency of ITS-1 repeat types within populations. Molecular drive (a realist theory) and neo-Darwinism (an idealist theory) differ radically with regard to their logical

basis of explanation (see chapter one). Given this difference, rather than marrying these modes of evolution it would be best to choose one over the other, particularly when discussing the concerted evolution of non-coding repeat DNA. It is my opinion that only molecular drive is likely to reveal an adequate account of the nature of the variation in rDNA, by virtue of its emphasis on establishing the particular generative processes responsible for the variation observed.

There is no limit to the choice of selectionist explanation, since it is easier to create new metaphors than it is to demonstrate biological generative processes. If non-randomness automatically indicates selection and randomness indicates neutrality, this biases thinking about the actual biology involved. Workers concentrate on the state (selected or not selected) of a process rather than the process(es) underlying the production of that state. From a realist perspective, dwelling on metaphors such as 'relaxed selection' lacks explanatory power. One is left without an indication as to how the observed irregularity (pattern) may have arisen, and explanations are justified by retreat to empirical proofs of a model of evolution. This is a manifestation of a broad trend in biology gone wrong, one in which explanation resides, not in the logic of the underlying processes, but in historically imagined events which lead to an 'adapted' world (Lambert, 1995).

The dogma of homogenization

In an early discussion of the importance of concerted evolution, Williams et al. (1988) cautioned against the idea that rDNA homogenization within a species requires only short periods of evolutionary time. However, in reply, Hillis and Davis (1988) stated that the low intra-individual and intra-specific polymorphism observed in the rDNA arrays of most species suggests that rapid rates of homogenization are widespread, if not universal. The latter view comes from a misunderstanding of the rate at which homogenization occurs, if this at

all occurs at a fixed rate. In 1988, 'most species' were not sampled for rDNA variation, but the observations of homogenization were elevated to the level of empirical law. Our knowledge is still limited about the rate of unequal crossing-over, the rate of gene conversion, and the rate of concerted evolution in repeat DNA families (Li, 1997). It is clear however, that observed rates of DNA change are unlikely to represent the "monotonous measure of molecular metronomes" (Dover, 1986), consequently, homogenization is not expected to be universal.

Recent findings of intraspecific heterogeneity lead Rich et al. (1997) to conclude that the hypothesis that all portions of the multiple copies of rDNA have been homogenized in all taxa, has been overstated. Because it is formally defined in terms of species differences, concerted evolution is often viewed as the result of molecular drive operating at the species level (Elder & Turner, 1995). It is perhaps this view that has lead to the assumption that species are homogenized at all levels and it has only been recently that this has been called into question.

Homology and history

In some recent studies, the distribution of rDNA variation hindered the reconstruction of cladistic relationships among geographic populations (Rich et al., 1997; Crease, 1995). The general explanation offered is that a proportion of the variation is ancestral and the remainder derived. Some IGS variants among populations of the water flea, *Daphnia pulex*, are inferred to be quite old, at least predating recolonization of freshwater habitats after glaciation (Crease, 1995). Some of the ITS-2 variants found in individual deer ticks (*Ixodes*) were also considered to be 'ancestral' (Rich et al., 1997). By the same reasoning, the geographic distribution of ITS-1 variants in this study also suggests that some variants may be ancestral. A problem not to be over-looked is whether the paralogous copies of a gene family can be determined or even understood as being ancestral or derived given their non-independent evolution.

In chapters one and three, it was suggested that gene duplication necessitated the elaboration of the concept of homology to include paralogy and orthology. Continual homogenization obscures the history of paralogous relationships among the members of a gene family, such that they appear as if they were a single-copy gene (an orthologous sequence). In gene trees, molecular characters need to be independent in order to uncover the history of divergence among the repeat types (Doyle, 1992). But the assumption of independence does not hold for phylogenies based on rDNA, because it undergoes concerted evolution. Sanderson and Doyle (1992) stated that "the entire concept of descent is confounded here by recombination [in the gene family] and reticulation in the gene tree." The monophyletic grouping of paralogous genes only provides strong evidence of homogenization, not that the paralogous genes are actually descended from a recent common ancestor (Sanderson & Doyle, 1992).

Perhaps it is not common descent, but the DNA turnover processes that represent the true homologies which generate the similarity (sameness) among repeat types. To Lima-de-Faria (1988), all processes represent homologies, it is only the degree of homology that varies. This is not a minor point. If homology can be defined by a demonstrable process, such as homogenization, then this should be the focus of future rDNA population studies, rather than the reconstruction of unrepeatable historical events.

References

Avise, J. C. (1994) *Molecular markers, natural history and evolution.* Chapman & Hall, New York.

Bateson, W. (1894) *Materials for the study of variation, treated with especial regard to discontinuity in the origin of species.* Cambridge University Press, Cambridge.

Bhaskar, R. (1978) *A realist theory of science.* 2nd ed. The Harvester Press, Sussex.

Blanden, R. V., H. S. Rothenfluh, P, Zylstra, G. F. Weiller and E. J. Steele (1998) The signature of somatic hypermutation appears to be written into the germline IgV segment repertoire. *Immunological Reviews* 162: 117-132.

Bowen, T. and G. A. Dover (1995) PCR amplification of intergenic spacers in the ribosomal DNA of *Drosophila melanogaster* reveals high levels of turnover in length and copy-number of spacers in geographically separated populations. *Mol. Ecol.* 4: 419-427.

Bridges, B. A. (1997) DNA turnover and mutation in resting cells. *BioEssays* 19: 347-352.

Brown, D. D., P. C. Wensink and E. Jordan (1972) A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. Mol. Biol.* 63: 57-73.

Bunn, S. E. and J. M. Hughes (1997) Dispersal and recruitment in streams: evidence from genetic studies. *J. N. Am. Benthol. Soc.* 16: 338-346.

Butler, D. K. and R. L. Metzenburg (1989) Premeiotic change of nucleolus organizer size in *Neurospora. Genetics* 122: 783-791.

Cairns, J., J. Overbaugh and S. Miller (1988) The origin of mutants. *Nature* 335: 142-145.

Chung, C. T., S. Niemela, and R. H. Miller (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* 86: 2172-2175.

Coen, E. S., J. M. Thoday and G. Dover (1982a) Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster. Nature* 295: 564-568.

Coen, E., T. Strachan and G. Dover (1982b) Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. *J. Mol. Biol.* 158: 17-35.

Coleman, A. W. and J. C. Mai (1997) Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relationships. *J. Mol. Evol.* 45: 168-177.

Collier, K. J. and B. J. Smith (1996) Riparian vegetation use by adult Ephemeroptera, Plecoptera and Trichoptera alongside some central North Island streams. *NIWA Science & Technology Series* 34.

Crease, T. J. (1995) Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* 141: 327-1337.

Dahlberg, A. E. (1989) The functional role of ribosomal RNA in protein synthesis. *Cell* 57: 525-529.

Doolittle, W. F. and C. Sapienza (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603.

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

Dover, G. A. (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics* 2: 159-165.

Dover, G. A. and W. F. Doolittle (1980) Models of genome evolution. *Nature* 299: 646-647.

Doyle, J. J. (1992) Gene trees and species trees: molecular systematics as one character taxonomy. *Syst. Bot.* 17: 144-163.

Eldeman, G. M., and J. A. Gally (1970) Arrangement and evolution of eukaryotic genes. *In:* F. O. Schmitt (ed), *The Neurosciences: Second Study Program.* Rockefeller University Press, New York, pp. 962-972.

Elder, J. F. J., and B. J. Turner (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* 70: 297-320.

Fitch, W. M. (1970) Distinguishing homologous from analogous proteins. *Syst. Zool.* 19: 99-113.

References

Fritz, G. N., J. Conn, A. Cockburn and J. Seawright (1994) Sequence analysis of the ribosomal DNA internal transcribed spacer 2 from populations of *Anopheles nuneztovari* (Diptera: Culicidae). *Mol. Biol. Evol.* 11: 406-416.

Ganley, A. R. D. and B. Scott (1998) Extraordinary ribosomal spacer length heterogeneity in as Neotyphodium endophyte hybrid: implications for concerted evolution. *Genetics* 150: 1625-1637.

Gerbi, S. A. (1985) Evolution of ribosomal DNA. *In:* R. J. MacIntyre (ed), *Molecular Evolutionary Genetics*. Plenum, New York, pp. 419-517.

Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart (1993) *An introduction to genetic analysis*. 5th ed. W. H. Freeman and Company, New York.

Hancock, J. M., D. Tautz and G. A. Dover (1988) Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol. Biol. Evol.* 5: 393-414.

Hancock, J. M. (1996) Simple sequences and the expanding genome. *BioEssays* 18: 421-425.

Hayashi, K. (1991) PCR-SSCP: a simple and sensitive method for detection of mutations in genomic DNA. *PCR Meth. Appl.* 1: 34-38.

Hennig, W. (1981) *Insect Phylogeny*. John Wiley & Sons, New York.

Higgins, D. G. and P. M. Sharp (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73: 237-244.

Hillis, D. M. (1994) Homology in molecular biology. *In:* B. K. Hall (ed), *Homology: the hierarchical basis of comparative biology*. Academic Press, New York, pp. 339-368.

Hillis, D. M., C. Moritz and B. K. Mable (1996) *Molecular Systematics.*, 2nd ed. Sinauer Associates, Massachusetts.

Hillis, D. M. and M. T. Dixon (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* 66: 411-453.

Hillis, D. M. and S. K. Davis (1988) Ribosomal DNA: intraspecific polymorphism, concerted evolution, and phylogeny reconstruction. *Syst. Zool.* 37: 63-66.

Hiss, R. H., D. E. Norris, C. H. Dietrich, R. F. Whitcomb, D. F. West, C. F. Bosio, S. Kambhampati, J. Piesman, M. F. Antolin and W. C. Black (1994) Molecular taxonomy using single-stranded conformation polymorphism (SSCP) analysis of mitochondrial ribosomal DNA genes. *Insect Molecular Biology* 3: 171-182.

Ho, M.-W. and P. T. Saunders (1984) Pluralism and convergence in evolutionary theory. *In:* M.-W. Ho & P. T. Saunders (eds) *Beyond neo-Darwinism: an introduction to the new evolutionary paradigm.* Academic Press, London, pp. 3-12.

Hoelzel, A. R. and G. A. Dover (1991) *Molecular Genetic Ecology.* Oxford University Press, London.

Hongyo, T., G. S. Buzard, R. J. Calvert and C. M. Weghorst (1993) 'Cold SSCP': a simple, rapid and non-radioactive method for optimized single-strand conformation polymorphism analyses. *Nucleic Acids Res.* 21: 3637-3642.

Hood, L., J. H. Campbell and S. C. R. Elgin (1975) The organization, expression, and evolution of antibody genes and other multigene families. *Ann. Rev. Genet.* 9: 305-353.

Hornby, A. S. (1970) *The advanced learner's dictionary of English.* Oxford University Press, London.

Hughes, J. M., S. E. Bunn, D. M. Kingston and D. A. Hurwood (1995) Genetic differentiation and dispersal among populations of *Paratya australienis* (Atyidae) in rainforest streams in southeast Queensland, Australia. *J. N. Am. Benthol. Soc.* 14: 158-173.

Hughes, J. M., S. E. Bunn, D. A. Hurwood, S. Choy and R. G. Pearson (1996) Genetic differentiation among populations of *Caridina zebra* (Decapoda: Atyidae) in tropical rainforest streams, northern Australia. *Freshwater Biology* 36: 289-296.

Jackson, J. K. and V. H. Resh (1992) Variation in genetic structure among populations of the caddisfly *Helicopsyche borealis* from three streams in northern California, USA *Freshwater Biology* 27: 29-42.

Kauffman, S. (1995) *At home in the universe: the search for laws of complexity.* Penguin Books, Auckland.

Kocher, T. D., W. K. Thomas, A. Meyer, A. V. Edwards, S. Paabo, F. X. Villablanca and A. C. Wilson (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86: 6196-6200.

Kovats, Z. E., J. H. Ciborowski and L. D. Corkum (1996) Inland dispersal of adult aquatic insects. *Freshwater Biology* 36: 265-276.

Kumar, A. and K. S. Rai (1990) Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionary diverse mosquitoes (Diptera, Culicidae). *Hereditas* 113: 277-289.

Kuperus, W. R. and W. Chapco (1994) Usefulness of the internal transcribed spacer regions of ribosomal DNA in Melanopline (Orthoptera: Acrididae) systematics. *Annals of the Entomological Society of America* 86: 751-754.

Lambert, D. M. (1995) Biological function: two forms of explanation. *In:* D. M. Lambert & H. G. Spencer (eds), *Speciation and the recognition concept, theory and application.* The John Hopkins University Press, Baltimore, pp. 238-259.

Lee, H.-H., W.-J. Lo and K.-B. Choo (1992) Mutational analysis by a combined application of the restriction fragment single-strand conformation polymorphism and the direct linear amplification DNA sequencing protocols. *Anal. Biochem.* 205: 289-293.

Lee, S. B. and J. W. Taylor (1992) Phylogeny of five fungus-like protistan *Phytophthora* species, inferred from the internal transcribed spacers of ribosomal DNA. *Mol. Biol. Evol.* 9: 636-653.

Levinson, G. and G. A. Gutman (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4: 203-221.

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

Li, W. and D. Graur (1991) *Fundamentals of molecular evolution.* Sinauer Associates, Sunderland, Massachusetts.

Liebherr, J. K. (1988) Gene flow in ground beetles (Coleoptera: Carabidae) of differing habitat preference and flight-wing development. *Evolution* 42: 129-137.

Lima-de-Faria, A. (1988) *Evolution without selection, form and function by autoevolution.* Elsevier, Oxford.

Linares, A. R., T. Bowen and G. A. Dover (1994) Aspects of nonrandom turnover involved in the concerted evolution of intergenic spacers within the ribosomal DNA of *Drosophila melanogaster. J. Mol. Evol.* 39: 151-159.

Long, E. O. and I. B. Dawid (1980) Repeated genes in Eukaryotes. *Ann. Rev. Biochem.* 49: 727-764.

Lovtrup, S. (1976) On the falsifiability of neo-Darwinism. *Evol. Theory* 1: 267-283.

Lovtrup, S. (1977) *The phylogeny of Vertebrata.* John Wiley & Sons, New York.

Maddison, W. P. and D. R. Maddison (1992) *MacClade: analysis of phylogeny and character evolution, version 3.0.* Sinauer Associates, Sunderland, Massachusetts.

McFarlane, A. G. and B. Cowie (1981) Descriptions of new species and notes on some genera of New Zealand Trichoptera. *Records of the Canterbury Museum* 9: 353-385.

McLachlan, R. (1862) Characters of new species of exotic Trichoptera. *Trans. Entol. Soc. Lon.* 3: 301-311.

McLain, D. K., D. M. Wesson, F. H. Collins and J. H. Oliver Jr. (1995) Evolution of the rDNA spacer, ITS 2, in the ticks *Ixodes scapularis* and *I. pacificus* (Acari: Ixodidae). *Heredity* 75: 303-319.

Miller, B. R., M. B. Crabtree and H. M. Savage (1996) Phylogeny of fourteen *Culex* mosquito species, including the *Culex pipiens* complex, inferred from the internal transcribed spacers of ribosomal DNA. *Insect Molecular Biology* 5: 93-107.

Monod, J. (1972) *Chance and necessity.* Collins, Glasgow.

Morse, J. C. (1997) Phylogeny of Trichoptera. *Annu. Rev. Entomol.* 42: 427-450.

Mougey, E. B., L. K. Pape and B. Sollner-Webb (1996) Virtually the entire *Xenopus laevis* rDNA multikilobase intergenic spacer serves to stimulate polymerase I transcription. *J. Biol. Chem.* 271: 27138-27145.

Musters, W., K. Boon, C. Van der Sande, H. Van Heerkhuizen and R. J. Planta (1990) Functional analysis of transcribed spacers of yeast ribosomal DNA. *EMBO J.* 9: 3989-3996.

Nagylaki, T., T. D. Petes (1982) Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* 100: 315-337.

Neboiss, A. (1991) Trichoptera. *In:* I. D. Nauman, P. B. Carne, J. F. Lawrence, E. S. Nielson & J. P. Spradbury (eds), *The Insects of Australia: A Textbook for Students and Research Workers.* Cornell University Press, Ithaca, New York, pp. 787-816.

Odorico, D. M. and D. J. Miller (1997) Variation in the ribosomal internal transcribed spacers and 5.8S rDNA among five species of *Acropora* (Cnidaria: Scleractinia): patterns of variation consistent with reticulate evolution. *Mol. Biol. Evol.* 14: 465-473.

Ohno, S. (1972) So much "junk" DNA in our genome. *Brookhaven Symp. Biol.* 23: 366-370.

Ohta, T. (1976) Simple model for treating evolution of multigene families. *Nature* 263: 74-76.

Ohta, T. (1983) On the evolution of multigene families. *Theor. Pop. Biol.* 23: 216-240.

Ohta, T. and G. A. Dover (1984) The cohesive population genetics of molecular drive. *Genetics* 108: 501-521.

Orgel, L. E. and F. H. C. Crick (1980) Selfish DNA: the ultimate parasite. *Nature* 284: 604-607.

Pashley, D. P., B. A. McPheron and E. A. Zimmer (1993) Systematics of holometabolous insect orders based on 18S ribosomal RNA. *Mol. Phylogenet. Evol.* 2: 132-142.

Patterson, C. (1988) Homology in classical and molecular biology. *Mol. Biol. Evol.* 5: 603-625.

Pinder, D. J., C. E. Blake, J. C. Lindsey and D. R. F. Leach (1998) Replication strand preference for deletions associated with DNA palindromes. *Molecular Microbiology* 28: 719-727.

Polanco, C., A. I. Gonzalez, A. Fuente and G. A. Dover (1998) Multigene family of ribosomal DNA in *Drosophila melanogaster* reveals contrasting patterns of homogenization for IGS and ITS spacer regions: a possible mechanism to resolve this paradox. *Genetics* 149: 243-256.

Quijada, A., A. Liston, W. Robinson and E. Alvarez-Buylla (1997) The ribosomal ITS region as a marker to detect hybridization in pines. *Mol. Ecol.* 6: 995-996.

Rich, S. M., B. M. Rosenthal, S. R. Telford, A. Spielman, D L. Hartl and F. J. Ayala (1997) Heterogeneity of the internal transcribed spacer (ITS-2) region within individual deer ticks. *Insect Molecular Biology* 6: 123-129.

Ridley, M. (1993) *Evolution.* Blackwell Scientific Publications, Boston.

Ritland, C. E., R. Ritland and N. A. Straus (1993) Variation in the ribosomal internal transcribed spacers (ITS1 and ITS2) among eight taxa of the *Mimulus guttatus* species complex. *Mol. Biol. Evol.* 10: 1273-1288.

Ross, H. H. (1944) The caddisflies, or Trichoptera, of Illinois. *Bulletin of the Illinois Natural history Survey* 23: 1-326.

Sambrook, J., E. F. Fritsch and T. Maniatus (1989) *Molecular cloning, a laboratory manual.* 2nd ed. Cold Spring Harbor Laboratory Press, USA.

Sanderson, M. J. and J. J. Doyle (1992) Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Syst. Biol.* 41: 4-17.

Schlotterer, C., M. Hauser, A. Haeseler and D. Tautz (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila. Mol. Biol. Evol.* 11: 513-522.

Shufran, K. A., D. C. Peters and J. A. Webster (1997) Generation of clonal diversity by sexual reproduction in the greenbug, *Schizaphis graminum. Insect Molecular Biology* 6: 203-209.

Simpson, G. G. (1953) *The major features of evolution.* Columbia University Press, New York.

Smith, G. P. (1973) Unequal crossing over and the evolution of multigene families. Cold Spring Harbor Symposium. *Quant. Biol.* 38: 507-513.

Steele, E. J., R. A. Lindley and R. V. Blanden (1998) *Lamarck's signature: how retrogenes are changing Darwin's natural selection paradigm.* Allen & Unwin, Australia.

Strachan, T., D. Webb and G. A. Dover (1985) Transition stages of molecular drive in multiple-copy DNA families in *Drosophila. EMBO J.* 4: 1701-1708.

Swofford, D. L. (1990) *PAUP: phylogenetic analysis using parsimony, version 3.0.* Illinois Natural History Survey, Champaign III.

Tang, J., L. Toe, C. Back and T. R. Unnasch (1996) Intra-specific heterogeneity of the rDNA internal transcribed spacer in the *Simulium damnosum* (Diptera: Simulidae) complex. *Mol. Biol. Evol.* 13: 244-252.

Tautz, D., M. Trick and G. A. Dover (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652-656.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz and G. A. Dover (1988) Complete sequences of the rRNA genes of *Drosophila melanogaster. Mol. Biol. Evol.* 5: 366-376.

Torres, R. A., M. Ganal and V. Hemleben (1990) GC balance in the internal transcribed spacers ITS 1 and ITS 2 of nuclear ribosomal RNA genes. *J. Mol. Biol.* 30: 170-181.

Volger, A. P. and R. DeSalle (1994) Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis. Mol. Biol. Evol.* 11: 393-405.

Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof, P. Snow, G. P. Wagner, J. Bartels, M. Murtha and J. Pendleton (1994) Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Syst. Biol.* 43: 250-261.

Walker, F. (1852) Catalogue of the specimens of neuropterous insects in the collection of the British Museum. *London: British Museum (Natural History)* 1: 1-192.

Ward, J. B., I. M. Henderson, B. H. Patrick and P. H. Norrie (1996) Seasonality, sex ratios and arrival pattern of some New Zealand caddis (Trichoptera) to light-traps. *Aquatic Insects* 18: 157-174.

Webster, G. (1984) The relations of natural forms. *In:* M.-W. Ho & P. T. Saunders (eds) *Beyond neo-Darwinism: an introduction to the new evolutionary paradigm.* Academic Press, London, pp. 193-217.

Wesson, D. M., C. H. Porter and F. H. Collins (1992) Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). *Mol. Phylogenet. Evol.* 1: 253-269.

Wheeler, W. C. and R. L. Honeycutt (1988) Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Mol. Biol. Evol.* 5: 90-96.

Williams, S. M., R. DeSalle and C. Strobeck (1985) Homogenization of geographical variants at the nontranscribed spacer of rDNA in *Drosophila mercatorum. Mol. Biol. Evol.* 2: 338-346.

Williams, S. M., G. R. Furnier, E. Fuog and C. Strobeck (1987) Evolution of the ribosomal DNA spacers of *Drosophila melanogaster*: different patterns of variation on X and Y chromosomes. *Genetics* 116: 225-232.

Williams, S. M., R. W. DeBry and J. L. Feder (1988) A commentary on the use of ribosomal DNA in systematic studies. *Syst. Zool.* 37: 60-62.

Williams, S. M. (1990) The opportunity for natural selection on multigene families. *Genetics* 124: 439-441.

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

Zylstra, P., H. S. Rothenfluh, G. F. Weiller, R. V. Blanden and E. J. Steele (1998) PCR amplification of murine immunoglobulin germline V genes: strategies for minimisation of recombinant artefacts. *Immunol. Cell Biol.* 76: 395-405.