The trematode genera *Paragonimus* and *Schistosoma* in East Asia : molecular evolution, phylogeny and biogeography.

Thesis submitted by
Lynne van Herwerden BSc., BSc. (Med.) Hons., MSc. University of Cape Town.

for the degree of Doctor of Philosophy
in the School of Biological Sciences (Zoology)
James Cook University of North Queensland

Dedication -

For my family

"... bump, bump, bump ... . It is as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if only he could stop bumping for a moment and think of it. And then he feels that perhaps there isn't."

taken from *Winnie - the - Pooh*,

- A.A. Milne, illustration by E.H. Shepard

STATEMENT OF ACCESS

I, the undersigned, the author of this thesis, understand that James Cook University of North Queensland will make it available for use within the University Library and, by microfilm or other means, allow access to users in other approved libraries. All users consulting this thesis will have to sign the following statement:

> In consulting this thesis I agree not to copy or closely paraphrase it in whole or in part without the written consent of the author; and to make proper written acknowledgment for any assistance which I have obtained from it.

Beyond this, I do not wish to place any restriction on access to this thesis.

29-06-1998

(Name) *(Date)*

STATEMENT ON SOURCES
DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references given.

29-06-1998

(*Name*) (*Date*)

ACKNOWLEDGMENTS

I am indebted to Dr. Takeshi Agatsuma and his colleagues, also to Prof. Don McManus and Dr. Paul Brindley for donating specimens and for their interest in this work. I would like to thank Dr. Robin Gasser for doing SSCP analyses on *Paragonimus* samples for me (Chapter 8) and generally for the interest that he has shown, particularly for valuable comments made during the finalisation of the manuscript from Chapter 8. Dr. Mark Blows gave me a statistical solution to my queries (chapter 6) and Dr. Julian Caley wrote a suitable Basic program and did the statistical analyses for me, (chapter 6). In addition he has contributed substantially to the preparation of the manuscript, based on chapter 6, as a co-author and I gratefully acknowledge his contributions.

I would like to thank many unknown reviewers of work presented here, for their valuable comments during the process of manuscript revision for publication. Thanks to my fellow postgraduate students that have shared in the frustrations and the moments of elation with me, particularly Faye Christidis, Jess Morgan, Dani Tikel, Dave Slaney, Simon Cook, Amanda Brooker and Vicki Hall - thanks for many discussions and pots of coffee, not to mention the copious amounts of chocolate.

My family deserves more gratitude than I can express. Firstly, I wish to thank my parents for providing me with the foundations, upon which I have been able to build my dreams, of which the culmination of this thesis is one. They have provided me with a strong work ethic and are always there when support of any kind is needed, despite the great distance between us. Secondly, Carl, my husband has always had faith in me and has boosted me with his loving support at all times. Finally, my children, Calyn and Leon, deserve a very special thanks for keeping my feet on the ground and providing me with the other half of being.

This thesis would not have been possible without the support and patience of my supervisor, Assoc. Prof. David Blair. Thank you for being enthusiastic and interested in the project and for giving me the freedom to explore.

ABSTRACT

Trematode infections affect one fifth of the worlds' population, with more than 200 million Chinese alone being at risk (World Health Organisation 1997, Hotez *et al.* 1997). Of particular interest to this study are the Asian species of the trematode genera *Paragonimus* and *Schistosoma*, which infect many mammals, including humans. It is becoming clear that within both genera species complexes exist that require further characterisation. Molecular markers have given some insight into the nature of these complexes. In this thesis additional molecular markers are investigated as sources of information about the nature and distribution of the *S. japonicum*, *P. westermani* and *P. ohirai* species complexes. These two genera are of additional interest because they are thought to have shared a similar biogeographic history in Asia, which should be reflected in their molecular evolution.

I have obtained DNA sequences from one nuclear (ITS1) and one mitochondrial (ND1) marker of five *Paragonimus* species (*P. westermani*, *P. miyazakii*, *P. macrorchis*, *P. ohirai* and *P. iloktsuenensis*) present in East Asia, including numerous geographic isolates of the *P. westermani* and *P. ohirai* species complexes. In addition, I have obtained ITS1 sequences of three schistosome species (*S. japonicum*, *S. malayensis* and *S. mekongi*) present in East Asia, which comprise the *S. japonicum* species complex and an African schistosome species (*S. mansoni*) for comparative purposes. I could differentiate all species using ITS1 and ND1 sequence data.

ITS1 is contained within the ribosomal DNA gene cluster, which is repeated in tandem many times. ITS1 does not experience the constraints of coding regions and theoretically undergoes sequence homogenisation within individuals and species. Despite this, intra-individual sequence variation was sometimes greater than inter-species variation, particularly among northeast Asian *P. westermani* isolates. Such variation confounded phylogenetic inferences for this group. There are some interesting aspects relating to processes of concerted evolution generally, when considering the differences within and between the two trematode genera, *Paragonimus* and *Schistosoma*, in this study. Additionally, a transcriptional enhancer motif (TATAAT) is embedded within the repeats of the ITS1 of the Asian schistosomes. The repeats are the cause of size variation and given the improbability of such motifs occurring by chance, I propose that the

abundant, large variants containing multiple copies of repeats may have a role in the stage- or tissue-specific regulation of transcription of the ribosomal genes.

In contrast, the ND1 gene codes for a mitochondrial protein and is therefore functionally constrained. Phylogenetic inferences could be made from ND1 sequence data obtained from *Paragonimus*. However, many clones had to be sequenced per individual to achieve this, as multiple ND1 lineages occurred within individuals of all *Paragonimus* species and strains investigated. Presumed pseudogenes were identified, which may be present in either the nuclear genome or in different types of mitochondria, as *Paragonimus* species have two types of mitochondria which differ structurally. It seems likely that both nuclear pseudogenes and heteroplasmic mitochondrial genes occur, though this hypothesis remains to be tested.

The *Paragonimus westermani* species complex has been further studied, to determine the origins of parthenogenetic triploid forms. Ribosomal DNA-ITS restriction fragments, ND1 sequences and simple sequence repeat (SSRs) "fingerprints" of triploids from China, Korea and Japan all indicate that the triploids are genetically different from one another. I conclude from this that triploids may have arisen more than once independently, possibly by relatively rare matings between diploid and tetraploid *P. westermani* individuals, which occur in sympatry with the triploids in NE China. Alternatively, triploid lineages may have arisen once and diverged subsequently by mutation alone.

Despite their suggested shared biogeographic history, members of the two genera exhibit rather different properties with respect to their ND1 and ITS1 genes. This is likely to be a reflection of their different phylogenetic histories. Molecular techniques have been used successfully to infer phylogenies which could be used to evaluate a previously published biogeographic hypothesis, but only to a limited extent, using the markers. It is clear that the situation for *P. westermani* in particular is complex and requires further investigation. Molecular evolution of these markers has proved to be an interesting component of the study that has brought to light some novel ideas and applications of these markers, not as phylogenetic tools, but as tools to study processes of concerted evolution in the nuclear genome and mitochondrial evolution.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 2. Literature Review.

Chapter 3. General Materials and Methods

Chapter 4. Intra- and inter-specific variation in nuclear ribosomal Internal transcribed spacer 1 of the *Schistosoma japonicum* species complex.

Chapter 5 Intra - and inter- individual variation in ITS1 of *Paragonimus westermani* (Trematoda: Digenea) and related species

LIST OF PUBLICATIONS

The following papers have been produced from material within this thesis.

1) D. Blair, L. van Herwerden, H. Hirai, S. Habe, M. Hirata, K. Lai, S. Upatham, T. Agatsuma (1997) Relationships between *Schistosoma malayensis* and other Asian schistosomes deduced from DNA sequences. *Molecular & Biochemical Parasitology* **85**: 259 - 263.

2) L. van Herwerden, D. Blair & T. Agatsuma (1998a) Intra- and inter-specific variation in nuclear ribosomal internal transcribed spacer 1 of the *Schistosoma japonicum* species complex. *Parasitology* **116:** 311-317.

3) L. van Herwerden, D. Blair & T. Agatsuma (1998b) Intra - and inter-individual variation in ITS1 of *Paragonimus westermani* (Trematoda: Digenea) and related species: implications for phylogenetic studies. (In press) *Molecular Phylogenetics and Evolution.*

4) L. van Herwerden, D. Blair & T. Agatsuma (1998c) Clonal diversity in parthenogenetic triploid *Paragonimus westermani*. (submitted) *International Journal of Parasitology*

5) L. van Herwerden, J. Caley and D. Blair (1998) Multiple ITS1 variants and putative transcriptional enhancers in trematodes (Digenea). *Nucleic Acids Research* (in preparation)

6) L. van Herwerden & D. Blair. Multiple lineages of NADH Dehydrogenase subunit 1 (ND1) genes in some trematodes. (submitted) *Journal of Molecular Evolution.*

CONFERENCE PROCEEDINGS

1. L. van Herwerden, D. Blair, T. Watanobe & T. Agatsuma. (1996). Molecular phylogenetic studies of the *Paragonimus westermani* species complex. *PASEAN: [Parasites in South East Asia Now.*]The Australian Society for Parasitology and The Parasite Control Association of Indonesia. Bali, Indonesia, September 1996.

2. L. van Herwerden and D.Blair, (1997). Comparison of nuclear and mitochondrial sequences for phylogenetic studies of human lung flukes. *Genetics Society of Australia,* (GSA) Perth, Western Australia, September 1997.

3. L. van Herwerden and D. Blair, (1998) Multiple lineages of NADH Dehydrogenase subunit 1 (ND1) genes in some trematodes. *Genetics Society of Australia,* Sydney, New South Wales, July 1998.

DEPARTMENTAL CONFERENCES

1. L. van Herwerden (1996). ITS sequences as a tool in phylogenetic studies of some trematodes. *ZONQ: Zoology in North Queensland,* James Cook University, Townsville, North Queensland, April 1996.

CHAPTER 1

General Introduction

1.1 Background

The original aim of this study was to evaluate the use of molecular methods to distinguish among, and infer phylogenies of, trematodes found in humans. The Asian representatives of the genera *Schistosoma* and *Paragonimus* are suitable candidates for this study for three reasons: Firstly, the taxonomy and distribution of species in East Asia is generally well-known. Secondly, some molecular data is already available and thirdly, inferred phylogenies can be evaluated against an explicit biogeographical hypothesis relating to the origin and distribution of both genera in East Asia.

Species that have a shared evolutionary history are likely to share similar characters. For example *S. mekongi* and *S. malayensis,* which have both diverged from *S. japonicum* share many characters to the exclusion of *S. japonicum,* including the transmitting snail subfamily (Triculinae rather than the Pomatiopsinae). The existence of other species in the *S. japonicum* complex was predicted on the basis of different transmitting snail subfamilies (Davis 1980), but could not be tested until molecular phylogenetic studies were undertaken, as the species were morphologically indistinguishable (cryptic). This demonstrates the value of molecular phylogenetic studies. Ecological and biogeographical hypotheses about organisms can be proposed on the basis of observations, but it is only by determining the evolutionary relationships between them (phylogeny) that such hypotheses can be rigorously tested.

Similarly, the extensive genetic diversity of the strains & species of *Schistosoma* and *Paragonimus* and their candidate vaccine antigens must be catalogued prior to the effective development of vaccines (Hotez *et al.* 1997). Such diversity can be appreciated within a phylogenetic framework. The importance of such cataloguing can be demonstrated by the failure of an experimental vaccine in mice, based on irradiated metacercaria from Chinese *S. japonicum*, to prevent reinfection by a Philippine strain (Hope, Duke & McManus 1996). In addition, there are important biological differences between these strains of *S. japonicum*, particularly infectivity and pathogenicity to the definitive host and drug susceptibility (McManus & Hope, 1993 and references therein). Vaccine development is essential for the longterm

treatment of people living in endemic areas who are at risk of rapid re-infection after treatment with antihelminthic drugs. The ability of pathogens to develop drug resistance is well known (for example antibiotic resistance in bacteria and insecticide resistance in insects), therefore the development of vaccines is considered an important challenge, so that these often debilitating pathogens can be effectively managed.

Molecular sequence data from the mitochondrial gene, cytochrome oxidase subunit I (COI) and the nuclear ribosomal DNA spacer, the internal transcribed spacer 2 (ITS2) have already been used for phylogenetic inference of both *S. japonicum* (Bowles, Blair & McManus 1995a) and the *P. westermani* (Blair *et al.* 1997a) and *P. ohirai* (Blair, Agatsuma & Watanobe 1997) species complexes. Both ITS2 and COI sequence data differentiated among species of the complexes so that phylogenetic relationships could be inferred. However, relationships among strains of the species complexes were not adequately resolved. Because functional trematode primers exist for ITS1 of the nuclear ribosomal DNA (rDNA) and NADH Dehydrogenase subunit 1 (ND1) of the mitochondrial genome and because both these regions have been successfully used in the past for better resolution at the species level, these regions were targeted as molecular markers in this study, to obtain phylogenetic resolution within the species complexes.

However, both molecular markers exhibited intra-individual variation that reduced their value for phylogenetic inference. The original aim to evaluate phylogenetic inferences against an existing biogeographic hypothesis was therefore reduced somewhat and the properties of the markers themselves became a focus for much of the thesis.

1.2 Thesis structure

The thesis begins with a literature review in Chapter 2, which provides the foundation for the work undertaken in this study. What were the expectations, on the basis of other molecular studies of trematodes, using ITS1 and ND1 sequences? What would we expect on the basis of the biogeographic hypothesis for the Southeast Asian trematodes? Why use molecular techniques and which genes should we target? How can we infer phylogenies from sequence data and what are the methods for evaluating such phylogenies? Chapter 3 is methodological, describing general materials and

methods used throughout this study. Separate, chapter-specific methods are presented where relevant. The remainder of the thesis consists of six data chapters (Chapters 4 - 9) in which I present the observations of this study and how these findings relate to findings presented in the broader literature. The phylogenetic findings are discussed in the context of an existing biogeographic hypothesis (addressed in Chapter 10, which is the concluding chapter). This chapter draws from the preceding data useful information relating to the original aims of the study and extends this to the broader context of molecular phylogenetic studies generally.

Chapter 3. General Materials and Methods.

Chapter 4. Intra- and inter-specific variation in nuclear ribosomal internal transcribed spacer 1 of the *Schistosoma japonicum* species complex. This section presents a description of the ITS1 data obtained for the Asian schistosomes, particularly describing intra-individual variation detected, which is due to the presence of multiple repeat sequences.

Chapter 5. Intra- and inter- individual variation in ITS1 of *Paragonimus westermani* (Trematoda: Digenea) and related species. This extends the ITS1 descriptive work on the Asian schistosomes in Chapter 4 to the Asian *Paragonimus* species, describing intra-individual variation in this genus, which is also a result of the presence of multiple repeat sequences, but not exclusively.

Chapter 6. Multiple ITS1 variants and putative transcriptional enhancers detected in Trematodes (Digenea). This chapter deals with the further investigation of repeats detected in the ITS1 (Chapters 4 & 5) and the presence of putative transcriptional enhancers within the repeats. It extends the work on *Schistosoma* and *Paragonimus* to consider other digeneans as well as a medically important cestode genus and proposes functional significance for the repeat sequences.

Chapter 7. Intra- and inter- individual variation in ITS1 of two trematode (Digenea) genera, *Paragonimus* and *Schistosoma*: phylogenetic inferences. In this section, I consider the phylogenetic signal detected in ITS1 of the Asian *Schistosoma* and the *P. ohirai* species complex, but the lack of such signal in the *Paragonimus westermani* species complex.

Chapter 8. Multiple lineages of NADH Dehydrogenase subunit 1 (ND1) genes in trematodes. Here I focus on ND1 sequence data obtained for the *Paragonimus* species, highlighting the interesting aspects of molecular evolution of this mitochondrial gene in this and other trematodes and the implications for phylogenetic questions.

Chapter 9. Clonal diversity in triploid parthenogenetic *Paragonimus westermani* (Trematoda: Digenea). In this chapter I address the question about the possible single clonal origin of the parthenogenetic triploid *P. westermani*, using ITS-rDNA RFLPs, ND1 sequence and microsatellite fingerprinting data, indicating the possible multiple independent origins of different triploid isolates.

Chapter 10. General Discussion . The concluding chapter synthesises all the data presented above. It draws attention to the potential pitfalls as well as to the broader implications of these findings for phylogenetic studies. It indicates that additional systematic work is required for the snail hosts that transmit *Paragonimus* species, before the hypothesis of coevolution, to the exclusion of host-switching can be accepted for this group.

Tables and Figures are interspersed with the relevant text within each chapter and all references are presented in the final section of the thesis.

CHAPTER 2

Literature Review

2.1 Background - why study these trematodes?

2.1.1. The genus *Paragonimus*

Members of the genus *Paragonimus* are long-lived hermaphroditic, digenetic trematodes that live as adults in pairs in the lungs of carnivorous mammals. Over 40 species have been named to date, nine of which infect man. *Paragonimus* spp. infect more than 50 million people world-wide and are among the important foodborne trematodes infecting humans (WHO 1997). They have a three-host life cycle, typical for trematodes. Humans are infected by eating crabs or other crustaceans containing live metacercariae. In addition, immature juveniles can be transmitted from paratenic (non-essential, intermediate) hosts, such as wild boar if improperly cooked flesh containing immature worms is eaten. One species, *Paragonimus westermani* (Kerbert, 1878) Braun 1899 probably consists of a complex of cryptic species of which the exact geographic limits and identities remain unclear (see comments in Blair *et al.*, 1997a). *Paragonimus westermani* is identified morphologically by the distinctive lobular form of the testes and ova and by the presence of single spines all over the body (Miyazaki 1991). *P. westermani* is the best known of the species pathogenic to humans and exists in both sexually reproducing diploid (2n) and parthenogenetic triploid (3n) forms. The sexual diploid individuals are self-incompatible, as unpaired individuals release unfertilised eggs that are non-viable (Miyazaki, Habe & Terasaki 1981). Triploid individuals produce eggs that have failed to undergo reduction division and their sperm are not functional (Fujino & Ishi 1982). The triploid form of *P. westermani* is the more pathogenic, as adults grow substantially larger and produce 10-fold more and larger eggs than the diploid adults. *P. westermani* is geographically widely dispersed, from Russia, India, Nepal and China to the Philippines and Indonesia. Diploid forms occur throughout the species' range, whilst triploid forms are restricted to Taiwan, Korea, Japan and China. Morphologically these two forms are similar,

however they generally utilise different crustacean hosts. Possible origins of the triploid form are discussed in chapter 9.

There is another complex, the *Paragonimus ohirai* species complex, which includes *P. ohirai ohirai* Miyazaki, 1939 , *P.ohirai* forma *iloktsuenensis* Chen 1940 and *P. ohirai sadoensis* Miyazaki *et al.*, 1968 , which is only distributed in coastal regions of Japan, China, Korea and Taiwan. This species complex is morphologically distinguished by a delicately branched ovary and the presence of grouped cuticular spines (Miyazaki 1991). Polyploidy has not been observed and reproduction is sexual. Molecular studies to date have however, not been able to distinguish between these species on the basis of allozymes (Agatsuma & Habe 1985) or the second internal transcribed spacer (ITS2) and cytochrome oxidase subunit I (COI) sequence data (Blair, Agatsuma & Watanobe 1997) (Fig. 2.1). Members of this complex very rarely infect humans.

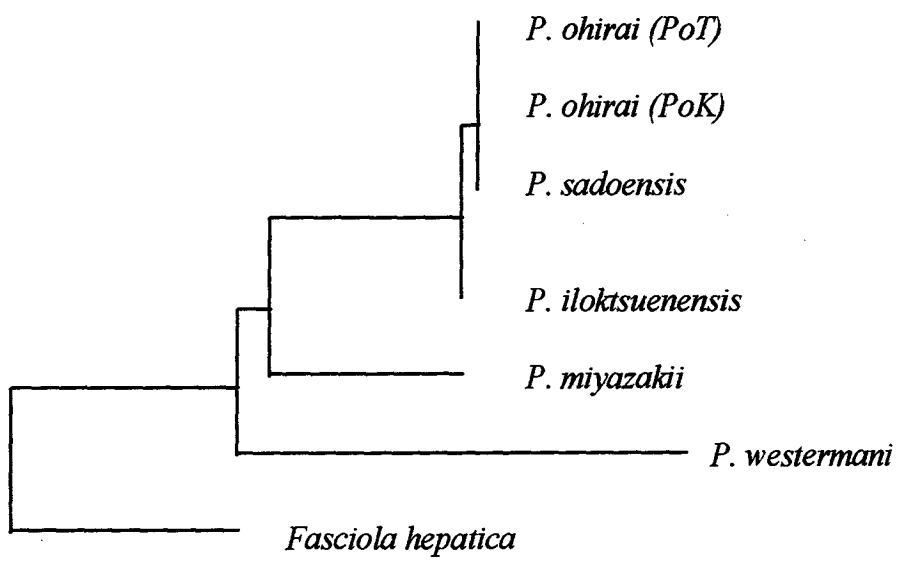


Fig. 2.1 Phylogenetic tree of species of *Paragonimus* inferred from mitochondrial CO I sequence data. Distance matrix based on the Kimura 2-parameter model. Tree constructed using Neighbour joining method (Blair, Agatsuma & Watanobe 1997) on PoT and PoK as described in Table 3.1.

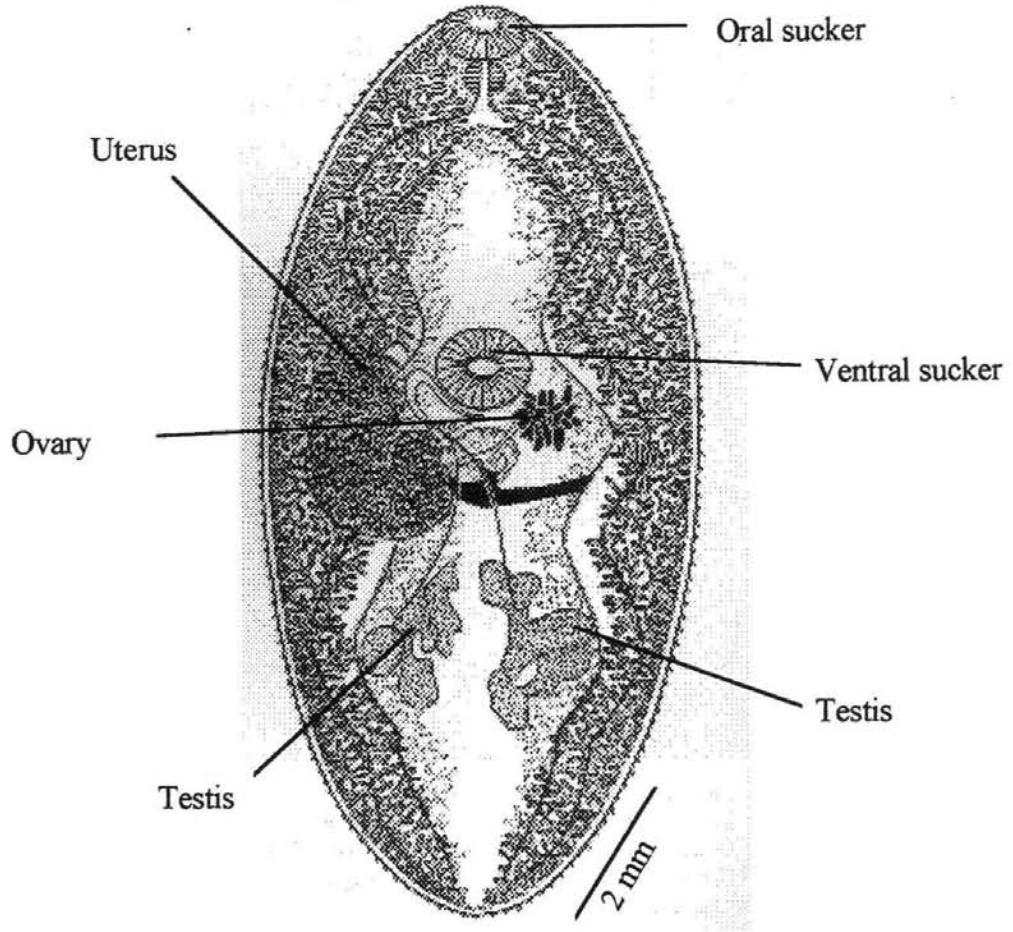


Fig. 2.2 Dorsal view of an adult *P. ohirai* (adapted from Miyazaki 1991).

Including members of the *P. westermani* and *P. ohirai* complexes, twenty five *Paragonimus* species have been recorded in China alone, some of which are probably synonyms. Some of the other species in Asia are *P. miyazakii* Kamo, Nishida, Hatsushika et Tomimura, 1961, *P. macrorchis* Chen, 1962 (both of which have been used as additional taxa for phylogenetic inference in this study), *P. heterotremus* Chen & Hsia, 1964 and *P. skrjabini* Chen, 1959 (neither of which has been used in this study). *Paragonimus miyazakii* and *P. skrjabini* belong to the same species complex (the *P. skrjabini* complex) (Blair *et al.* 1998). Both are transmitted by pomatiopsid snails. Nominal *Paragonimus skrjabini* is transmitted by 23 species of snails.

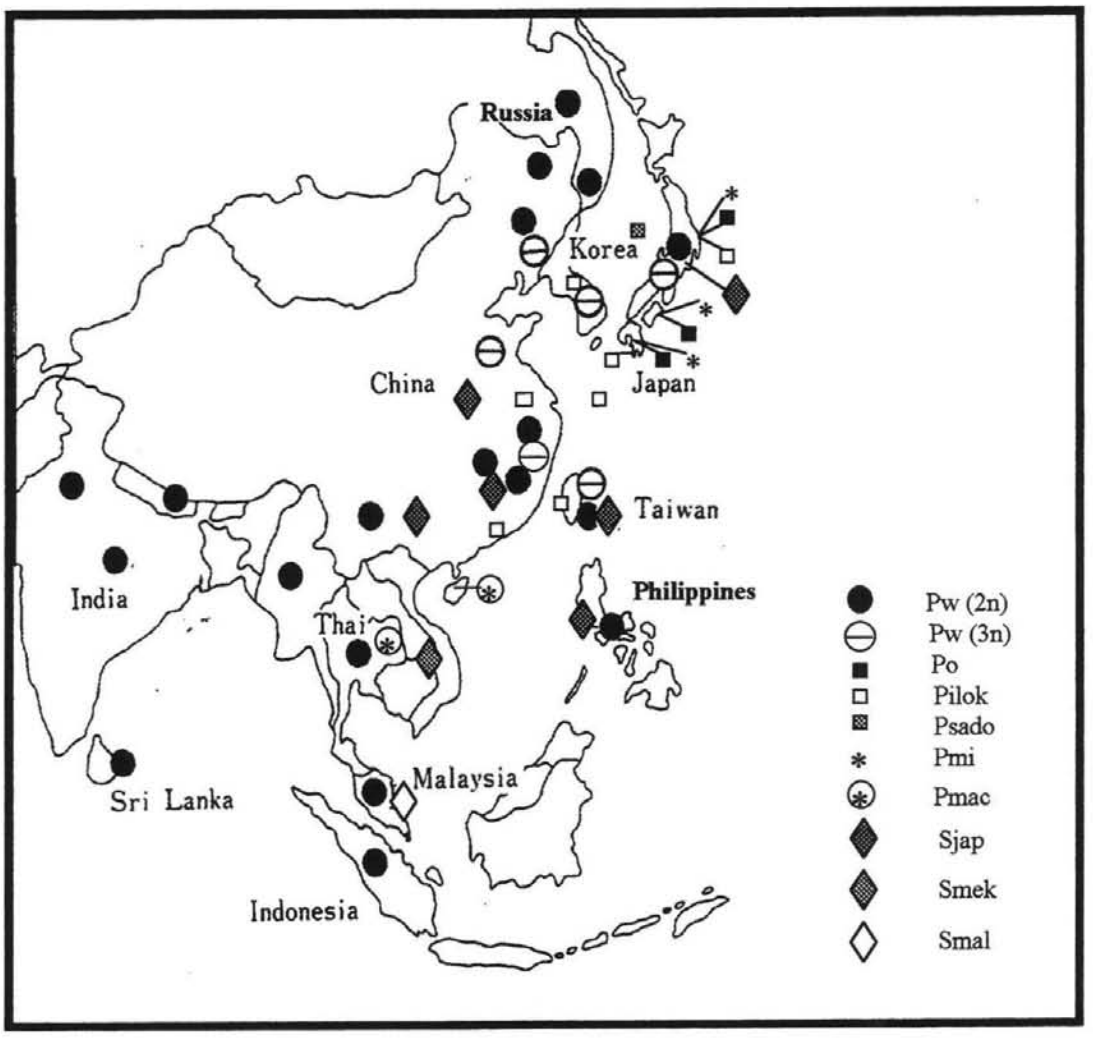


Fig. 2.3 Asian distribution of *Paragonimus* and *Schistosoma* species investigated in this study. Pw (2n) is diploid *P. westermani*, Pw (3n) is triploid *P. westermani*, Po is *P. ohirai*, Pilok is *P. iloktsuenensis*, Psado is *P. sadoensis*, Pmi is *P. miyazakii* and Pmac is *P. macrorchis;* Sjap is *S. japonicum*, Smek is *S. mekongi* and Smal is *S. malayensis*. (adapted from Miyazaki 1991).

2.1.2 The genus *Schistosoma*

Members of the genus *Schistosoma* are atypical, dioecious, digenetic trematodes that live as adults in pairs in the vasculature of the vertebrate hosts. Nineteen species have been recognised to date, seven of which infect man. *Schistosoma* spp. infect one in thirty people world-wide and are medically second in importance only to malaria among parasitic diseases. *Schistosoma* spp. have a two-host

life cycle, which is atypical for trematodes. Humans are infected by entering water containing live cercariae, which penetrate the skin directly. In Asia, *Schistosoma japonicum* (Katsurada, 1904) Stiles, 1905 is the most important and widespread pathogenic species. The *Schistosoma japonicum* species complex consists of widely distributed, biologically different strains of *S. japonicum* and two localised species (*S. malayensis* Greer *et al.* 1988 and *S. mekongi* Voge, Bruckner et Bruce, 1978). *S. japonicum* occurs throughout China to Taiwan, Japan, the Philippines and Indonesia (Hope, Duke & McManus 1996), where it is transmitted by the Pomatiopsid snail genus *Oncomelania* (subfamily Pomatiopsinae). *Schistosoma malayensis,* which is transmitted by *Robertsiella* (subfamily Triculinae), occurs in Malaysia and *S. mekongi* is restricted to the lower Mekong river, where it is transmitted by another Triculine, *Neotricula aperta* (see Fig. 2.3).

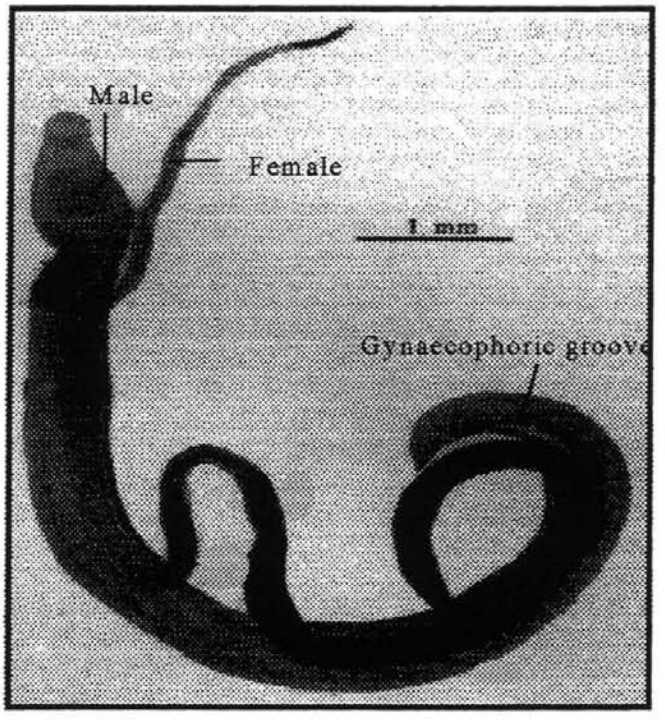


Fig. 2.4 Paired adult *S. japonicum* individuals, illustrating how the female fits into the gynaecophoric groove of the male (adapted from Miyazaki 1991).

2.2 Asian biogeography and the Davis Hypothesis (Davis 1980, 1992)

An explicit biogeographic hypothesis, focussing on the origins, distribution and evolution of the Asian pomatiopsid snails and their associated trematodes was proposed in the late 1970's and has been developed and refined during the subsequent two decades to the present (Davis 1979, 1980, 1992, 1995). Other snail families occurring in Asia may have different biogeographic histories. Most of what follows is from the papers by Davis.

The Mesozoic Pomatiopsidae and their trematode parasites started co-evolving in the Gondwanian regions which are currently the continents of Africa, India and South America (Davis 1980, 1992). India was adjacent to Africa prior to the break-up of Gondwana. Fossil evidence indicates that India was at that time extensively marshy (during the latter period of the Mesozoic, about 65 MYA (million years ago)). Few species of snails with non-specifically associated trematodes would have inhabited such a uniform, marshy environment (Davis 1992). This hypothesis is supported by an analysis of vicariant distribution patterns of modern snail faunas, generalised tracks (based on fossil distributions of ancestral snail species) and direction of dispersal and evolution together with the restriction of the trematodes to snail hosts of the most ancestral bodyplan (Davis 1980). The few prosobranch and pulmonate snail ancestors of extant species, together with their non-specifically associated ancestral trematodes, arrived on the Asian plate via the Indian plate. The initial contact between India and Asia (about 45 MYA) occurred long before the most intense phase of the Himalayan orogeny, which occurred during the middle Miocene, approximately 16 MYA (Adams 1981). Furthermore, the northwestern corner of the Indian plate connected with Asia long before the northeastern corner (25 MYA) (Davis 1980 and references therein). It was the upheaval of the Tibetan Plateau, during the Himalayan orogeny, that generated the drainage patterns, which gave rise to the major Asian rivers, the Yangtze, Mekong, Red, Salween and Irrawaddy, approximately 16 MYA (Fig. 2.6). As the river systems evolved south- and eastward from Yunnan, China and Tibet, the snails and their parasites evolved down the major rivers. New ecological niches were created for the snails as a result of geological upheaval and altered hydrology of the area. The fresh water niches were invaded by ancestors of the modern Triculinae. Other snails

(precursors of the modern *Oncomelania*) maintained an amphibious existence (Davis 1979, 1980).

We know that Asian *Schistosoma* and some *Paragonimus* species (e.g. *P. skrjabini, P. miyazakii* and members of the *P. ohirai* complex) are transmitted by two subfamilies of the Pomatiopsidae (Pomatiopsinae and Triculinae). In Asia, therefore, the trematodes transmitted by these snails are likely to have shared the same biogeographic history.

There is however no clearly stated biogeographic hypothesis for the Assimineidae, also of the rissoacean superfamily and which only transmit members of the *Paragonimus ohirai* species complex. Members of the *Paragonimus westermani* species complex are transmitted by a completely different superfamily of snails, the ceritheacean families Thiaridae and Pleuroceridae (Fig 2.5), which do not transmit any *Schistosoma* species. As is the case for the Assimineidae, there is no clearly stated biogeographic hypothesis for the Thiaridae and Pleuroceridae. This will be discussed further in Chapter 10, section 10.1.3).

According to Davis, non-specific *Schistosoma* (precursors to *S. japonicum* and *S. sinensium* species complexes) probably inhabited pomatiopsid snails at the time of their arrival in Southeast Asia approximately 16 MYA, as described above (section 2.2). The schistosome parasites must have tracked the snail host evolution as they moved down the evolving rivers, with specificity of the snail-parasite interaction increasing with the rapid speciation of the snails, so that the schistosome species became ecologically divided.

This "explosive stage of speciation" and cladogenesis of the snails started about 12 MYA (± 4 MY), with increasing morphological specialisations observed (Davis 1980), as the snails evolved downstream. Schistosomes had to coevolve (as defined by Janzen 1980) with their snail hosts or become extinct. This is supported by the observation that many more species of snails than schistosomes exist in Asia and that the schistosome transmitting snails are those with the most "conserved" body plan, i.e. the least derived morphology (Davis 1992). *Schistosoma japonicum* remained restricted to the *Oncomelania hupensis* subspecies (Pomatiopsinae), with the evolution of population specificity between host and parasite (Davis *et al.* 1995). This coevolved specificity is further supported by experimental studies, in which different strains of

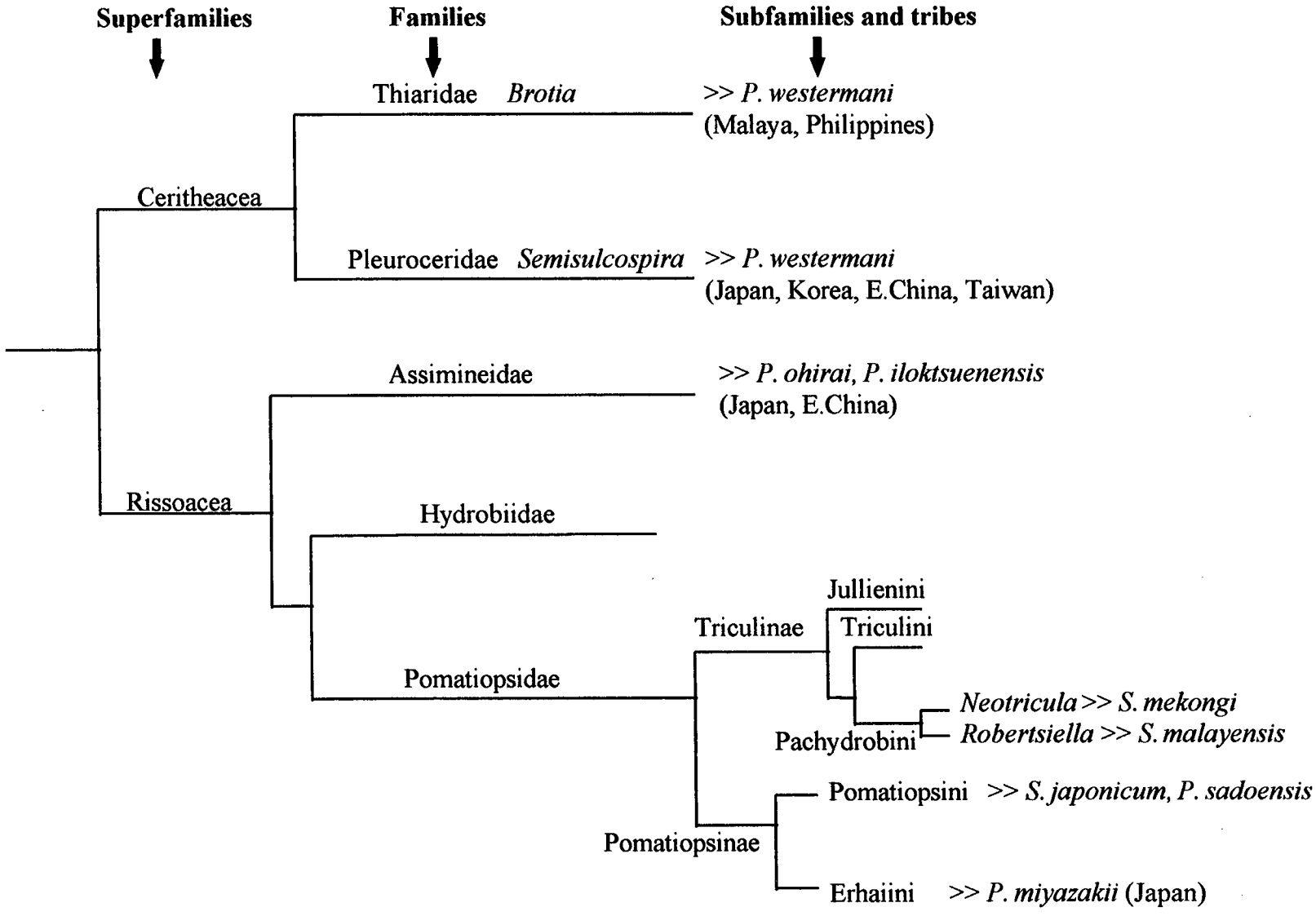


Fig. 2.5 Phylogeny of superfamilies, families, subfamilies and tribes of snail hosts transmitting *Schistosoma* and *Paragonimus* species, with parasite species mapped onto the tree (adapted from Davis, Spolsky & Zhang 1995).

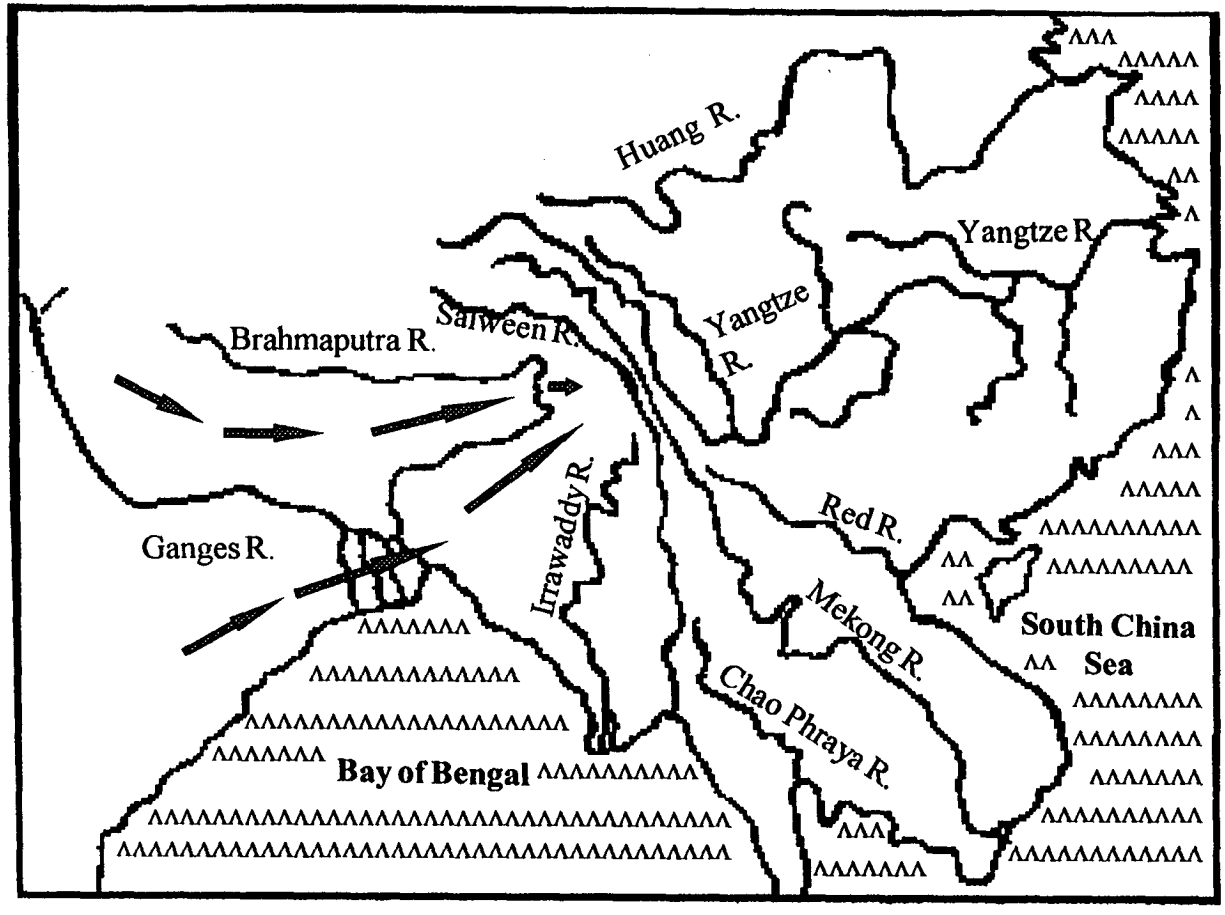


Figure 2.6 Map showing the major Asian rivers, with the drainages extending southward along the Salween River and to the west of it and eastward along the Yangtze and Red Rivers and to the northeast of these rivers.

Arrows indicate the possible route of introduction of Pomatiopsid snails from the Indian plate (according to Davis 1980).

Oncomelania hupensis may be infected by different *S. japonicum* strains, but at very much reduced levels, compared to infections of coevolved strains (De Witt in Davis 1980).

Another lineage of the Pomatiopsidae, the Triculinae diverged into three tribes, of which the Pachydrobini contains the genera *Neotricula* and *Robertsiella*, which respectively transmit *S. mekongi* and *S. malayensis*. These schistosomes diverged from each other 4 - 5 MYA, on the basis of a molecular genetic distance, Nei's D, of 0.6 (assuming a molecular clock) (Yong *et al.* 1985, Woodruff *et al.* 1987). *Neotricula aperta*, which transmits *S. mekongi*, lives only in the lower Mekong River, in quiet flowing or still water. Both *Robertsiella*, which transmits *S. malayensis* and *Neotricula* are thought to be derived from a *Tricula*-like ancestor. Extant *Tricula* species live in tiny mountain streams.

As the *S. sinensium* species complex has not been investigated in this study, I shall not discuss the distribution and co-evolution of this group and the snails transmitting it, in detail. However, for the sake of completeness, Table 2.1 and Fig. 2.7 indicate how both Asian schistosome species complexes are transmitted and distributed in Asia.

In contrast, the relationships of *Paragonimus* species in Asia with their snail hosts has not been adequately researched to date, as transmitting snails have not yet been studied in terms of modern systematics (Hotez *et al.* 1997). However, we know that *Paragonimus* has diversified extensively in China and that there are vast radiations of snails that transmit *Paragonimus* there. The Pomatiopsidae has two subfamilies that transmit trematodes: the Pomatiopsinae (with the genus *Oncomelania*), which transmits both members of the *Schistosoma japonicum* species complex and *P. sadoensis* (a member of the *Paragonimus ohirai* species complex) and the Triculinae, which are implicated in the transmission of both *Paragonimus* and *Schistosoma* species (Davis, Spolsky & Zhang 1995).

Table 2.1 *S.japonicum* and *S. sinensium* species complexes with their transmitting prosobranch snail hosts and geographic locations in Asia (Davis 1992).

Schistosome	snail host	location
S. japonicum complex		
S. japonicum	*Oncomelania hupensis*	China, Japan, Philippines, Sulawesi
S. mekongi	*Neotricula aperta*	Lower Mekong River
S. malayensis	*Robertsiella kaporensis*	Central Malaysia
S. sinensium complex		
S. sinensium ss.	*Tricula* or *Neotricula* sp.	Sichuan, China
S. sinensium-like "a"	*Jinhongia jinhongensis*	Yunnan, China
S. sinensium-like "b"	*Tricula bollingi*	Northwest Thailand

Some *Paragonimus* species, for example *P. ohirai* and *P. westermani*, appear to have strains that have coevolved with local strains of snail hosts, as has been observed for *S. japonicum* (Habe 1993 and Hamajima *et al.* 1989 in Blair, Xu & Agatsuma 1998). However, the ability to switch hosts also seems to be a property of some *Paragonimus* species which are capable of infecting many species of snails. For example both *P. ohirai* and *P. iloktsuenensis* of the *P. ohirai* species complex utilise 9 species of snails from two families (the Assimineidae and Pomatiopsidae) naturally and experimentally (Davis *et al.* 1994). In addition *P. sadoensis*, also of the *P. ohirai* species complex infects Pomatiopsidae. These species have even been shown to infect the American pomatiopsid *Pomatiopsis lapidaria* experimentally. *Paragonimus westermani*, which is transmitted by the cerithiacean family Pleuroceridae (*Semisulcospira libertina*) in Japan, is transmitted in Malaysia and the Philippines by species of Thiaridae (*Brotia* spp.) (Fig. 2.5). This issue will be revisited in Chapter 10 (section 10.1.3).

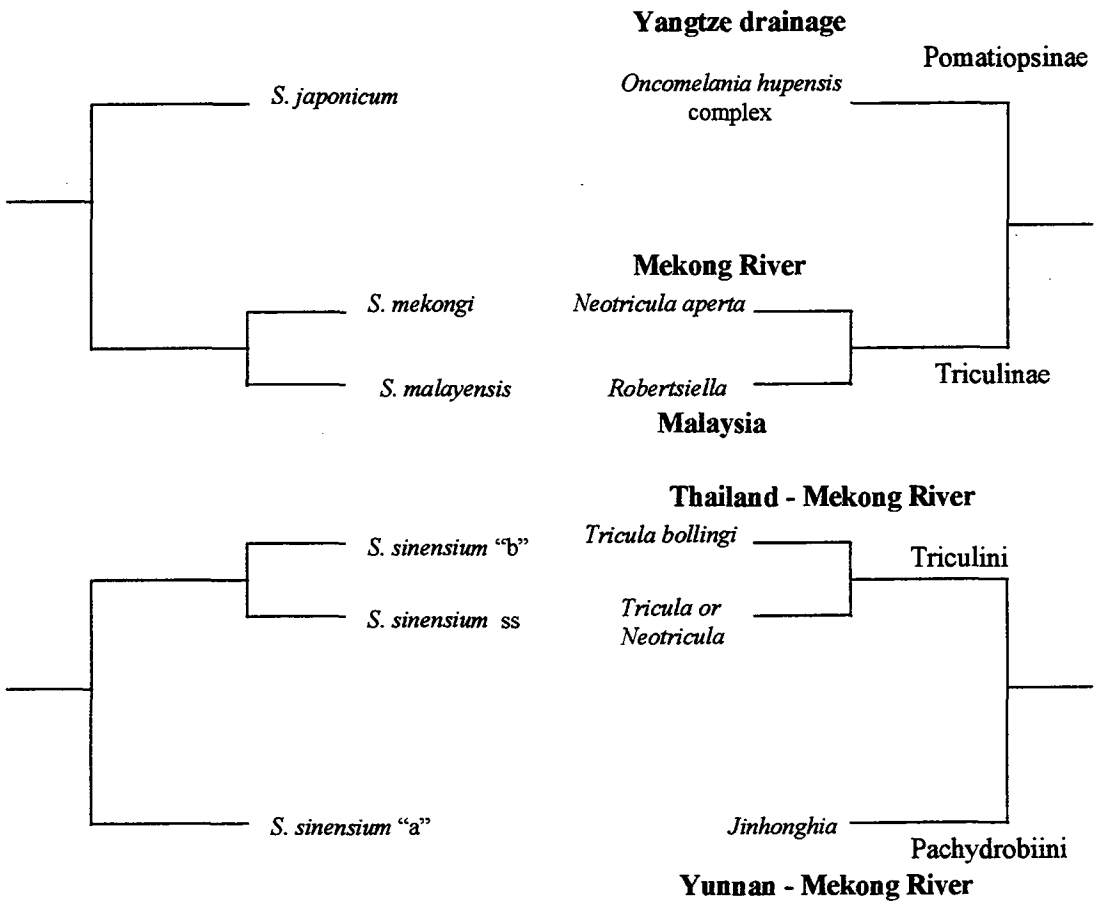


Fig. 2.7 Cladograms of Asian schistosomes, Asian rivers and schistosome transmitting snail families (redrawn from Davis 1992). The *S. japonicum* species complex phylogeny is based on data from allozymes, sequences and host-specificity, whereas the *S. sinensium* species complex phylogeny is assumed, based only on snail host distribution.

2. 3 Predictions of Davis' Hypothesis

Davis has predicted tight coevolution between snails and their trematode parasites, i.e. congruent phylogenetic trees should be obtained for the snails and their parasites. This does appear to be the case for members of the *S. japonicum* species complex (Fig. 2.7), but the pattern is less clear for the species of *Paragonimus* investigated. An explicit phylogeny has already been inferred for the schistosome transmitting snails (at the family level) based on allozymes, morphology (particularly the radula and sex organs of both males and females) and cytochrome b

sequences. (Davis 1979, 1980, 1992, 1994, Davis *et al.* 1994a, Davis *et al.* 1995). In this study I present additional phylogenetic data for the parasites.

Davis predicted that the allopatric, morphologically similar members within each of the nominal species of *S. japonicum* and *P. westermani* would be found to consist of at least two distinct species each, when molecular genetic data is obtained. This prediction was made on the basis of the different transmitting snail hosts for particular races of each of the two trematode species (Davis *et al.* 1994b). The prediction has been realised for the *S. japonicum* species complex (Bowles,Blair & McManus 1995a) and is supported by COI and ITS2 sequence data for the *P. westermani* species complex which have shown that there may be at least two species in the complex (Blair *et al.* 1997a). In summary, current data suggest that the biogeographic hypothesis of snail-parasite coevolution proposed by Davis is supported for some of the snails and their parasites, but not for others. This will be discussed further in Chapter 10.

2.4 Why use molecular techniques?

2.4.1 Molecules versus morphology

Soft-bodied organisms such as trematodes have a limited number of phylogenetically useful morphological characters, making it difficult to discriminate between closely related species or subspecies. In any case, morphological characters may be phenotypically plastic. For example extensive morphological homoplasy in the Digenea has confounded phylogenetic studies (Blair, Bray & Barker 1998), as has convergent evolution of the molluscan shell (Davis 1994, Davis *et al.* 1994a) and coral reef fish mouthparts (Streelman and Karl 1997). DNA sequences, however, are absolute and representative for a particular strain or species, irrespective of life-cycle stage characterised. Amplification of selected regions of the genome by PCR requires very small quantities of DNA, so that data can be obtained from small samples, such as individual worms, metacercariae or eggs (Gasser *et al.* 1993). Once a region of the genome has been amplified by PCR, it can be sequenced. DNA sequences produce large numbers of characters that can be compared between individuals.

The disadvantages of using DNA sequences as species markers are that the technology is expensive, though costs are diminishing while efficiency is improving. It is not adequate to sequence one gene region only, because gene trees do not always correspond to species trees (discussed below in section 2.6.3). Therefore, careful consideration of the most suitable regions to sequence is required. Base compositional and substitution / mutation rate heterogeneity are other potentially confounding factors, but these can be dealt with by using various approaches to analyse data. For example, this can be done by using a Markov model of nucleotide substitution whereby varying compositional biases between species are accommodated (Galtier & Gouy 1995 and Hillis *et al.* 1994). Reticulation and secondary contact, which is particularly acute in species complexes is another potentially confounding issue when using molecular sequences for the inference of phylogenetic relationships (Avise 1986, 1994, Arnold 1997). The choice of region depends on the depth of the phylogenetic relationship being investigated.

Ideally, a combined approach is favored, where the emphasis is on organismal evolution, which can be studied through molecular phylogenetic and comparative methods (Hedges & Maxson 1997). Indeed, the interplay between molecular, morphologic and other lines of evidence attain greater significance than does any individual data set by itself (Avise 1994).

2.4.2 Which genome - nuclear or mitochondrial?

DNA sequences of both nuclear and mitochondrial genomes are commonly used to obtain information about the relatedness of species and strains. The two genomes differ extensively in size, overall mutation rate, base composition, mode of inheritance and genetic code. A combination of sequence data from both genomes in a total-evidence approach is potentially a powerful source of information for phylogenetic inference (Huelsenbeck *et al.* 1996.). However, if a combination of data results in increased homoplasious signal, a total-evidence approach should not be used. The homoplasious signal is probably due to reticulation or secondary contact between members of a species complex (Avise 1986, Arnold 1997), particularly for organisms such as *Paragonimus westermani,* which are hermaphroditic, self-incompatible members of a species complex (Miyazaki 1991). Following is a discussion of each

genome as a source of phylogenetic information, indicating the advantages as well as the disadvantages associated with each.

2.4.2a Nuclear DNA

The nuclear genome varies enormously in size from about one million base pairs (1 Mb) in bacteria to approximately 10^5 Mb pairs in some eukaryotes (Alberts *et al.* 1989). In sexually reproducing species nuclear DNA is inherited from both parents and undergoes recombination.

The most commonly used part of the nuclear genome for systematic studies of the trematodes is the tandemly repeated ribosomal DNA (rDNA) (Blair *et al.* 1996). In the haploid *Schistosoma mansoni* genome rDNA consists of one hundred tandemly repeated units (Hirai, Spotila & LoVerde 1989) and all 100 copies occur at a single locus in the nuclear genome, the nucleolar organiser region (NOR). Each unit contains three coding (18S, 5.8S and 28S) regions and four non-coding (NTS, ETS, ITS1 and ITS2) spacer regions (Fig. 2.8). Three of the spacers are transcribed together with the ribosomal genes to produce a single precursor ribosomal RNA molecule (pre-rRNA). The spacers are subsequently removed from the pre-rRNA to produce mature rRNA (van der Sande *et al.* 1992). The degree of sequence conservation varies across the rDNA, with the spacers being most variable (Hillis and Dixon 1991).

Processes of concerted evolution and molecular drive maintain uniformity of sequences in tandemly repeated genes, such as the rDNA, within an individual and populations of a species (Dover 1986, Dover & Tautz 1986). Gene conversion, unequal crossing over and slipped strand mispairing are the processes of concerted evolution (Arnheim 1983). Molecular drive fixes variants that have arisen due to random genetic drift and natural selection throughout a species (Dover 1982, Dover & Tautz 1986). It is this property of fixation of variants within individuals and populations that has enabled the phylogeneticist to use rDNA sequences for comparative studies at the species level. Sample sizes can be small, because single individuals should be representative of a species (Blair *et al.* 1996).

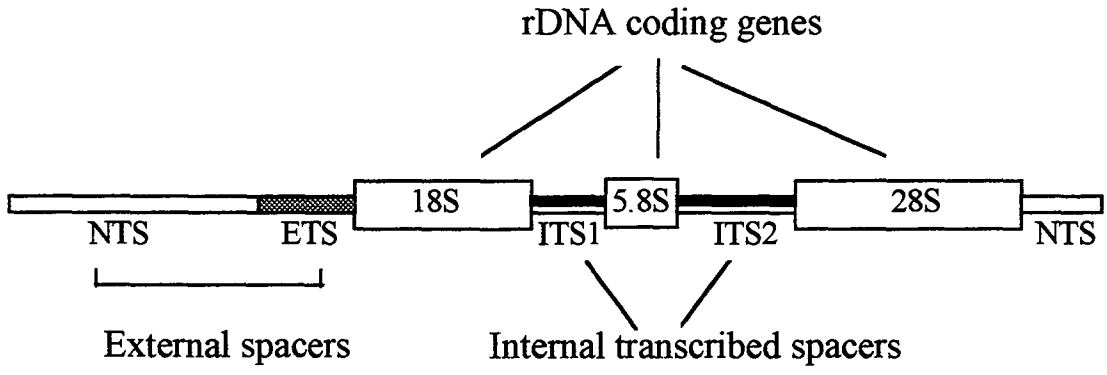


Fig. 2.8 The ribosomal DNA unit, consisting of three coding regions and associated spacers (external and internal).

ITS 1 and 2 sequences from the rDNA have been extensively used to investigate species level relationships in many plants and animals (Coleman and Mai 1997, Wesson *et al.* 1993). The first digenean trematode ITS 1 & 2 was sequenced for the genus *Dolichosaccus* by Luton, Walker and Blair (1992). Many interspecific differences were detected between two *Dolichosaccus* species. Subsequently, ITS2 sequences have been used successfully to discriminate among species of *Schistosoma* (see Després *et al.* 1992, Bowles, Blair and McManus 1995b), *Fasciola* (see Adlard *et al.* 1993) and *Echinostoma* (see Morgan & Blair 1995).

ITS1 sequences have been found to be more variable than ITS2 sequences in most of the trematodes studied. However, the ITS1 of some genera, specifically *Dolichosaccus* and the African *Schistosoma* may have internal repeats embedded therein (Luton, Walker & Blair 1992, Kane & Rollinson 1994). The presence of repeats may give rise to intra-individual variation of ITS1 sequences, though this was not the case in either the African *Schistosoma* or *Dolichosaccus* species investigated. Should there be intra-individual variation in the ITS1, it would detract from the usefulness of ITS1 as a phylogenetic tool, because of the problem of non-orthologous sequences (Patterson 1988). Single copy nuclear genes have been less well studied to date, but their use should circumvent the problems of paralogy that may be and have been encountered with multigene families such as the ribosomal genes in some taxa (Brower & De Salle 1994). Indeed, currently introns of single copy nuclear genes are increasingly being targeted for the development of primers for amplification and sequencing of variable regions, for use in phylogenetic studies (Lyons *et al.* 1996).

Nuclear genes as tools for phylogenetic study are often hampered by the presence of pseudogenes. During the course of this study, I developed primers to amplify the second intron of the triosephosphate isomerase (TPI) gene, as a good candidate variable region in a single copy nuclear gene. However, as multiple size products were amplified, I chose not to pursue this region. Pseudogenes for TPI have been identified in some organisms (Brown et al. 1985, Zhang & Chinnappa 1994). This work is not presented in this thesis.

2.4.2.b Mitochondrial DNA

The mitochondrial genome is a small, circular molecule of approximately 14 - 22 kb in animals. It is generally maternally inherited (Avise *et al.* 1987) and does not undergo recombination. Despite the presence of many mitochondria in each cell, all the copies within an individual are usually identical, as a result of random replication, random segregation and gene conversion (Birky 1996). In vertebrates mitochondrial DNA evolves on average 5 - 10 times more rapidly than typical, single copy nuclear DNA (Brown *et al.* 1979 in Avise 1986), however, the difference is not as great in the case of invertebrates. For the above-mentioned reasons mitochondrial DNA is frequently used to address phylogenetic questions at and below the species level. The first trematode mitochondrial sequences to be obtained were cytochrome *c* oxidase subunit 1 (CO1) and NADH dehydrogenase subunit 1 (ND1) from a gene library from *Fasciola hepatica* (see Garey and Wolstenholme 1989). Universal primers designed from the *F. hepatica* sequences have been successfully used to amplify and sequence either or both genes within the trematode genera *Schistosoma* (see Bowles, Blair and McManus 1995a, Blair *et. al* 1996, Blair *et al.* 1997b), *Echinostoma* (see Morgan & Blair 1995, 1998a & b), *Paragonimus* (COI only, Blair *et al.* 1997a) and the cestode genus *Echinococcus* (see Bowles, Blair and McManus 1992, 1994, 1995b, Bowles and McManus 1993). COI sequence data produced phylogenetic trees for all the above-mentioned genera that were congruent with trees based on nuclear rDNA ITS2 sequences (Bowles, Blair and McManus 1995b, Blair *et al.* 1996, Blair *et al.* 1997a). However, although ND1 data for the *Echinostoma* and *Echinococcus* genera produced the anticipated phylogenetic trees, there were some inconsistencies that cannot be adequately explained if conventional mitochondrial DNA genetics is assumed.

One of the possible reasons for inconsistencies in phylogenetic inferences using mitochondrial sequence data, is that PCR amplification of mitochondrial genes from total genomic DNA may amplify nuclear copies of mitochondrial genes (numts), if they are present . There is a growing body of literature that indicates that the integration of mitochondrial sequences into the nuclear genome (numt DNA) is indeed a significant problem for phylogenetic studies in many taxa, from vertebrates to insects and fungi (Lopez *et al.* 1994, Zhang & Hewitt 1996 & references therein, Sunnucks & Hales 1996, Zischler, Geisert & Castresana 1998, Sorensen & Quinn 1998).

Another possible reason for inconsistent phylogenetic inferences from mitochondrial sequence data is heteroplasmy, the presence of more than one mitochondrial genome type in an individual. Heteroplasmy may be caused by hybridisation of closely related species (reviewed by Avise 1986), paternal leakage or by biparental inheritance of mitochondria, as has been described for the *Mytilus edulis* species complex (Zouros *et al.* 1992, 1994, Stewart *et al.* 1995, Rawson & Hilbish 1995).

2.5 Sequence alignment

Once a number of sequences have been obtained for phylogenetic inference or other comparisons, they must be aligned.

When sequences are aligned an assumption is made that aligned characters are homologous (descended from a character present at a given site, in a common ancestor) in all taxa represented. It is generally neccessary to add gaps to maintain an alignment between sequences. Minor changes in an alignment may result in altered and therefore, possibly faulty, inferred trees (Morrison and Ellis 1997). A simple and reliable method of sequence alignment is by eye, when few taxa with similar sequences are being aligned (Blair *et al.* 1996). More complex sequences can be aligned by using one of several alignment programs, such as those present in the GCG database (Devereux, Haeberli & Smithies 1984).

A simple dot-matrix method of alignment places one sequence of a pair along the horizontal axis and the other along the vertical axis (dot plot). When the same base occurs on both axes, a dot appears in the matrix. A perfect diagonal results when the two sequences are identical. A discontinuous diagonal results when some base

differences occur and the diagonal shifts up or down in the presence of alignment gaps, such as may be caused by insertion-deletion (indel) events. The presence of repeats within a sequence results in additional diagonals, which are parallel to the main diagonal. It becomes difficult to obtain an alignment if the number of gaps and base changes increases sufficiently for the diagonal to become dissipated (Li 1997).

Sequences that have diverged substantially, so that alignment by eye or dot-matrix is not feasible, can be compared by similarity and distance methods. Both methods minimise the number of gaps. Similarity analysis maximises the number of identical pairs, while distance methods minimise the number of different pairs. Both methods use a "gap penalty", which takes into account the number and position of alignment gaps required (Li 1997).

The prediction of secondary structure (conformation) is another technique used to assist with sequence alignments of ribosomal spacers. This technique uses conserved folding motifs to align sequences with little primary (sequence) similarity, as structural motifs must be maintained for ribosomal processing (Michot *et al.* 1993, van Nues *et al.* 1995). It is therefore possible to compare secondary structures among genera that lack primary sequence similarity, for the inference of phylogenetic relationships. This has been done for trematodes, using ITS2 secondary structures (Michot *et al.* 1993, Morgan & Blair 1998c).

2.6 Genetic distance estimates

The degree of difference among aligned sequences is shown as a pairwise distance matrix. The p-distance, which measures the proportion of differences between two sequences, is the simplest measure of distance. The p-distance is an accurate measure of distance only if pairwise distances are not large (Kumar, Tamura and Nei 1993). However, with increased pairwise distances, multiple hits are more common, which results in the underestimation of the true distance, because the same site might be mutated more than once. Different models of the rate of substitution among bases (or amino acids for protein sequence) may be applied to estimate the pairwise distance matrix under these circumstances.

The most common mutations are point substitutions, of which transitions (A⇔G, T⇔C), occur more readily than transversions (A⇔C or T, G⇔ C or T),

because less energy is required for a purine ⇔ purine or a pyrimidine ⇔ pyrimidine change, than for a purine ⇔ pyrimidine change (Wakeley 1996). When sequences continue to diverge, multiple substitutions may occur at the same site (multiple hits), which results in transitions being masked by transversions. For a sequence of equal base frequency (0.25 : 0.25 : 0.25 : 0.25) for each of the four bases, saturation has been reached when two transversions occur per transition (Nei 1987). This ratio is a function of the base frequency.

Different distance models compensate for multiple substitutions with increasing complexity. Only the most commonly used, simpler models (Jukes-Cantor, Tajima-Nei, Kimura 2-parameter and Tamura-Nei) are discussed here. These models all assume that all positions in a non-coding sequence evolve at the same rate and exclude deletions and insertions from the comparisons. More complex models, for estimating evolutionary distances are reviewed by Li (1997). Additional parameters are estimated from the data as the models become more complex, so that the variance of the distance estimate increases. For this reason, the simpler model should always be used if similar distance measures are obtained by two models, however, the more complex models should be applied in order to compare the distance measures obtained by the two models (Li 1997).

The Jukes-Cantor model (JC) is the simplest substitution model and assumes that nucleotides are equally represented and that the rate of substitution for all nucleotides is the same. It should not be used if nucleotide frequencies deviate substantially from equal proportions. The Tajima-Nei model (TjN) is similar, but allows for different nucleotide frequencies. This estimate is robust only if multiple hits have not occurred per site. The Tamura-Nei substitution model (TrN) allows for G+C : A+T ratio bias and divides transitions into two components: A↔G and T↔C. The Kimura 2-parameter model (K2P) assumes different substitution rates for transitions and transversions, but like the JC model, assumes equal nucleotide frequencies. Studies of trematode systematics (Bowles, Blair and McManus 1995a and b, Barker and Blair 1996, Blair *et al.* 1997a & b) have most commonly usedthe Kimura 2-parameter model .

2.7 Phylogeny reconstruction procedures

Phylogenetic trees are graphical representations of evolutionary relationships among organisms. The operational taxonomic units (OTUs) of a phylogenetic tree may be alleles of a gene, individuals, well-isolated conspecific populations, species or higher taxa, depending on the study. Terminal nodes represent extant OTUs and internal nodes represent the ancestral states. Branch lengths indicate the number of changes along the path connecting ancestors to descendants. A tree is rooted when the common ancestor of all extant OTUs is specified, otherwise it is unrooted. The estimation of a phylogenetic tree is challenging, because astronomic numbers of possible trees exist that can connect a number of OTUs, of which only one can be the true tree (for example, for 10 OTUs the number of possible unrooted bifurcating trees = 2 027 025) (Avise 1994). Therefore, a variety of tree-building techniques needs to be evaluated in an attempt to ascertain which methods are best suited for phylogenetic reconstruction. Following is a brief description of some commonly used algorithms in molecular phylogenetic analyses, which are divided into quantitative and qualitative approaches (Avise 1994). Each of these approaches is used under different circumstances (Li 1997).

2.7.1 Quantitative, distance methods

Distance methods start with an OTU X OTU matrix, which consists of the estimated pairwise distances among all OTUs and uses all the variable sites. A tree is constructed from the distance matrix by comparing the relationships among these distances. Numerous tree-making methods which are based on different principles, are available. Here, I only discuss the two methods in common use.

The simplest method computationally is the UPGMA method (unweighted pair group method with arithmetic averages) (Sneath & Sokal 1973), which was originally developed for morphologic data. The method builds a phylogenetic tree by grouping (clustering) the two most similar units in a stepwise manner. Subsequent steps identify the next closest unit based on the next smallest distance and so on, in order of decreasing similarity (Li 1997). The major assumption of UPGMA is a constant rate of evolution along all branches, i.e. it assumes a universal clock. This assumption is often violated. Computer simulations have indicated that UPGMA produces reasonable tree

topologies (Nei *et al.* 1983 in Avise 1994), which may be an indication that the distance averaging of UPGMA reduces the large stochastic errors inherent in genetic-distance estimates.

The neighbour-joining method (N-J) (Saitou & Nei 1987) is conceptually related to UPGMA analysis, with the important difference that it allows for a variable rate of evolution along all branches, i.e. it assumes a local clock. Hillis *et al.* (1992) evaluated the performance of distance methods (including UPGMA & N-J) and found that all the algorithms tested produced the correct tree topology, but differed in their ability to obtain the correct branch lengths. Parsimony methods (see below, section 2.6.2) produced the best correlated value between predicted and observed branch lengths. The N-J method has been most frequently used in trematode studies for constructing phylogenetic trees (Bowles, Blair and McManus 1995a and b, Barker and Blair 1996, Blair *et al.* 1997b, Morgan & Blair 1998a & b).

Distance based methods are easily confounded by convergence, as the clustering of OTUs is done on the basis of overall similarities, regardless of whether the similarities are derived or primitive (ancestral). The second class of tree-building techniques is based on qualitative (character-state) approaches and will now be discussed.

2.7.2 Qualitative, character-state methods

Discrete character information, such as a particular nucleotide position in a DNA sequence or a particular amino acid position in a protein sequence, can be used to develop a data matrix, which consists of an assigned character state for each character for each OTU. A particular nucleotide position has five possible states (A,C,G or T) or - (gap), so it is a multi-state character. In the case of sequence data alternative character states are unordered, as it is difficult to make assumptions about the particular evolutionary pathway for the change from one nucleotide to another (Avise 1994). The most commonly used qualitative character-state methods for molecular phylogenetics are maximum parsimony and maximum likelihood which are briefly described below.

Maximum parsimony (MP) seeks to find the tree that requires the smallest number of character changes to account for the observed data (Felsenstein, 1989).

The algorithm was originally developed for phylogenetic analysis of morphological characters. If interior branches are very short or if the evolutionary rate varies extensively within evolutionary lineages, the method tends to be inconsistent (Kumar, Tamura and Nei 1993). An exhaustive search of all possible trees will find the shortest tree(s), but is very slow and should not be undertaken for more than 10 taxa due to time constraints (Li 1997). The branch-and-bound algorithm will also find the shortest tree(s) and can be used for up to 20 taxa using PAUP 3.1.1, but beyond that becomes too laborious and slow. The limit to feasibility depends on the amount of homoplasy in the data, the speed of the algorithm and the computer it is running on (Trueman 1997)). The branch-and-bound algorithm requires information on the length of some tree first. Such a tree length becomes the upper bound at the outset of the analysis. Trees are then built by the algorithm by adding one taxon at a time, until either all taxa are joined or the upper bound is reached. All partly built trees that exceed the upper bound are rejected as unparsimonious. If the bound is not reached, then the shorter tree obtained and its length becomes the new upper bound..This process continues until all trees have been evaluated or rejected (Trueman 1997). For analyses involving more than 20 taxa, it is essential to use heuristic tree-search algorithms (Li 1997). Heuristic searches are done by first building a starting tree and then doing branch swapping to find the shortest tree. Branch swapping continues until no shorter tree can be found. Swapping can be done in one of three ways, nearest neighbour interchange (NNI), subtree pruning regrafting (SPR) and tree bisection reconnection (TBR) (Li1997). TBR is the branch swapping method of choice, as it includes SPR as a subset, which includes NNI as a subset. The MP method performs only a little better (or at best as good as) the NJ method when the number of substitutions is small and when the sequence length is $\geq$ 1200 nucleotides (Li 1997). In all other cases the NJ method is superior to the MP method (Li 1997).

The Maximum likelihood algorithm (M-L) calculates the probability of a process of random change (based on a model) giving rise to an observed sequence (Li 1997). The probability of observing the original configuration of sequences is calculated for a given hypothetical tree. This probability is calculated for all possible trees and the preferred tree is the one that produces the largest (maximum) likelihood value (Li 1997). Maximum likelihood methods (M-L) are computationally very

demanding and cannot realistically be used for determining the phylogenetic relationship among more than 10 OTUs (Li 1997) It is however more robust than both MP and NJ methods, as shown by simulation studies.

2.7.3 Evaluating phylogenetic inferences

It is advisable to use multiple methods of data analysis, particularly using both types of approach described above (2.6.1 &2), for example N-J and M-P methods. Generally, real phylogenetic signal in a data set is reflected in the phylogenetic tree obtained, regardless of the algorithm used (Avise 1994). If different phylogenetic trees are obtained for the same data set when using different algorithms, particularly resulting in different topologies, additional data should be acquired.

All tree building methods will produce a tree (or trees) from a given data set. Even completely random sequences will do so (Nadler 1990). It is therefore neccessary to assess if the phylogenetic inference for a particular data set differs from random. This can be done by examining the distribution of tree lengths for all possible trees. Random sequences usually produce a symmetrical length frequency distribution of trees. Real data with hierarchical structure, however, produce a frequency distribution that is skewed to the left. The degree of asymmetry, indicated by by the g_1 statistic, can be tested for statistical significance, where

$$g_1 = \frac{\Sigma^{n}_{i=1}(T_i-T)^3}{ns^3}$$

where *s* is the standard deviation of the tree lengths. Table 2.2 shows what g_1 should be for four-state character data if there is significantly more hierarchic signal present in the data than random. (Hillis & Huelsenbeck 1992).

It is important to provide a measure of confidence in the lineages obtained from phylogenetic inference. The most commonly used method of confidence testing is bootstrapping (Felsenstein 1985a), which involves resampling the data (with replacement). This process of resampling indicates how well particular phylogenetic lineages are supported by the data. Ideally, a minimum of 1000 bootstrap replicates is recommended and clades with bootstrap confidence limits (BCL) of $\geq 70\%$ have been shown to be correct in $> 95\%$ of estimates for a known experimental phylogeny (Hillis & Bull 1993). It is important to note also that the bootstrap technique was developed

Table 2.2. Critical values of g_1 for sequence (four-state) data. Data for which the g_1 values are less (i.e. more negative) than those shown, contain significantly more structure than random, p = 0.05 (Hillis & Huelsenbeck 1992).

no. of bases	no. of taxa							
	5	6	7	8	9	10	15	25
10	-0.95	-0.7	-0.59	-0.51	-0.44	-0.34	-0.23	-0.16
50	-0.78	-0.58	-0.45	-0.37	-0.25	-0.28	-0.16	-0.12
100	-0.66	-0.56	-0.40	-0.31	-0.25	-0.30	-0.15	-0.10
250	-0.81	-0.43	-0.39	-0.26	-0.22	-0.20	-0.14	-0.08
500	-0.73	-0.43	-0.27	-0.29	-0.23	-0.16	-0.12	-0.07

under the assumption that sites in the DNA sequences are independent of one another and that all sites evolve at the same rate, conditions which are likely to be violated. Future research will focus on removing the assumptions of independence and site equivalence during bootstrapping procedures (Li 1997).

As well as assessing confidence in particular tree topologies, it is possible to test the reliability of branch lengths obtained by N-J methods. A confidence probability (CP) indicates if branch lengths differ significantly from zero by using a t-test, which is obtained by dividing the branch length by its standard error (Kumar, Tamura and Nei 1993).

Finally, it must be emphasised that an assumption is made that the characters in a data set are homologous and independent in all phylogenetic inferences of the data (Avise 1994). The assumption of independence is probably valid for most molecular characters (but not so for many morphological characters). The question of homology, however, requires some further discussion. If a gene has been duplicated, different copies of the gene may have been sequenced in different species (or individuals) inadvertently, so that the genes are not homologous. They are instead paralogous, and the phylogenetic tree obtained is a gene tree rather than a species tree. A gene tree may correspond to a species tree only if the gene is orthologous.

In this study, I have used both nuclear rDNA spacer sequences and mitochondrial gene sequences to infer phylogenetic relationships within species complexes in two trematode genera present in southeast Asia. I set out to use the phylogenetic data to compare with an existing biogeographic hypothesis. Furthermore, I wished to evaluate another hypothesis pertaining to the clonal origin of parthenogenetic triploid *P. westermani*. This was achieved, but only to a limited extent, as the two selected regions were variable within individuals, which was an unexpected observation. I wished to investigate the molecular evolution of these regions, which neccessitated the reduction in the emphasis on the original hypotheses. I therefore focused on the molecular evolution of the markers, as understanding such processes would help me to explain why they are not suitable for phylogenetic inference in this group of trematodes.

The work and the interesting implications of the findings are presented in the following seven chapters.

CHAPTER 3

General Materials and Methods

3.1 Sample material

Preserved sample material used in this study was kindly provided by various workers, listed in Table 3.1. These workers are respectively experts on the trematodes supplied and they have done extensive morphological and taxonomic studies on the organisms. As all material was obtained in preserved form, I was unable to isolate mitochondria, but instead obtained total DNA, comprised of both nuclear and mitochondrial DNA. Since the same sample material was used throughout this study, I present the details here once.

Adult trematodes were generally raised in experimental animals. *Paragonimus* species were raised by infecting domestic cats or dogs with metacercariae, except *P. macrorchis*, which was obtained from a naturally infected bandicoot (Table 3.1). All sequences were obtained from individual adults, when dealing with *Paragonimus* species. ITS1 and ND1 sequence data were obtained from DNA extracts from the same individuals in all cases. When Southern blots were prepared, total DNA from additional individuals was used. Unfortunately, it was not possible to obtain material from the western (India) or southern (Indonesia) edges of the *P. westermani* range, so that samples from the entire range of distribution could not be included in this study. Samples from Korea, Japan and NE China are from localities close to the E and NE limits of the species range.

S. malayensis adults had been raised in mice infected with cercariae from *Robertsiella* sp. snails. Snails were obtained from Baling, Malaysia in 1993 (Institute of Medical Research, Kuala Lumpur). *S. mekongi* from Khong Island, Laos were maintained in *Neotricula aperta* for 10 years at Mahidol University, Bangkok, Thailand. Strains of *S. japonicum* were collected from *Oncomelania* snails near Sorsogon, Philippines in 1992 and maintained at Queensland Institute of Medical Research (QIMR), Brisbane, Australia. Snails naturally infected with *S. japonicum* from Anhui, China were routinely sent to Brisbane, where released cercariae were collected and used to infect Balb/c mice. DNA from 20 - 100 pooled individuals of *S. mansoni* (Puerto Rico, Caribbean) was donated by Paul Brindley (Table 3.1). All

Asian *Schistosoma* species sequenced were from pooled samples (5-20 individuals), except where indicated in Figure and Table captions.

As can be seen from Table 3.1, several isolates of *P. westermani* from Japan, China and Malaysia were included in this study. Philippine and Korean isolates were also included, although multiple isolates were not available from these locations. In addition, two of the three members of the *P. ohirai* species complex (*P. ohirai* and *P. iloktsuenensis*) were also represented by two individuals each. No material from *P. sadoensis* was available for this study. Two additional species, *P. macrorchis* and *P. miyazakii* were included in this study, but only a single individual of each species was available. The availability of different isolates of *Schistosoma* species was also restricted to individual isolates of *S. mekongi* and *S. malayensis*, although I did obtain two strains of *S. japonicum* (Chinese and Philippine). *S. mansoni* material was only obtained from an individual isolate, as many other isolates of this species have already been studied and are available for comparison through the sequence databases (Genbank).

3.2 DNA extraction

Adults of *Schistosoma* and *Paragonimus* species had been stored in 80% ethanol at 4^0C. Prior to DNA extraction they were washed 3 times in extraction buffer (40mM Tris pH8.0, 10mM EDTA, 200mM NaCl) before being digested in digestion buffer (extraction buffer containing 1% SDS, 500μg/ml Proteinase K) at 37^0C for 2hrs. DNA was then extracted by standard phenol-chloroform techniques (Sambrook *et al.* 1989).

3.3 DNA amplification and PCR primers used

Primers used for PCR amplification of ITS1 and ND1 are presented in Table 3.2. These primers anneal to highly conserved sequences listed in Table 3.2. PCR amplification of ITS1 was done as follows, using primers BD1 and 4S: 10 - 100 ng template and 10 pmol of each primer was added to 1 x *Taq* polymerase buffer

Table 3.1 Sample material used in this study, abbreviated codes, source and geographic location of isolates. For *Paragonimus* every line represents an individual worm.

Species	Species code (ploidy)	Source	Geographic location
Paragonimus	PwJO(2n)	T. Agatsuma	Japan, Ohita
westermani	PwJH (2n)	"	Japan, Hyogo
species	PwJM(2n)	"	Japan, Mie
complex	PwP (2n)	"	Philippines, Sorsogon
	PwM (2n)	"	Malaysia, K.P.
	PwM (2n)	"	Malaysia, U.L.
	PwC(2n)	"	E China, Minchin
	PwC(?n)	Li Yong-long	China, ?
	PwJA(3n)	T. Agatsuma	Japan, Amakusa
	PwJT(3n)	"	Japan, Tsushima
	PwCL(3n)	"	NE China, Liaoning
	PwK(3n)	"	Korea, Bogil Island
P. ohirai [a]	PoT	"	Japan, Tanegashima
P. ohirai [a]	PoK	"	Japan, Kinosaki
P. iloktsuenensis [a]	Pilok 279	"	Japan, Amami
P. ikoktsuenensis[a]	Pilok K	"	Japan, Amami
P. macrorchis	Pmac	J. Waikagul	Thailand
P. miyazakii	Pmi	T. Agatsuma	Japan, Miyazaki
Schistosoma japonicum	Sjap	D. McManus	Philippines, Sorsogon
S. japonicum	Sjap	D. McManus	China, Anhui
S. malayensis	Smal	K. Lai	Malaysia, Baling
S.mekongi	Smek	S. Upatham	Thailand, Laos
S. mansoni	Sman	P. Brindley	Caribbean, Puerto Rico

[a] Members of the *P. ohirai* species complex.

(Promega), 40μM each dNTP, 1 - 3mM $MgCl_2$ and 0.25 Units of *Taq* Polymerase (Promega) in a final volume of 20μl. Amplification reactions were done after denaturation at 96°C, 1 min as follows: 30 cycles of 95°C, 1 min; 55°C, 1 min and 72°C 2 min. on a DNA Engine (MJ Research Inc., USA). USP and RSP were used as sequencing primers of cloned PCR products (Table 3.2 and section 3.5 below). Chapter specific PCR reaction conditions are also presented in the relevant chapters (Chapters 4, 5 and 8).

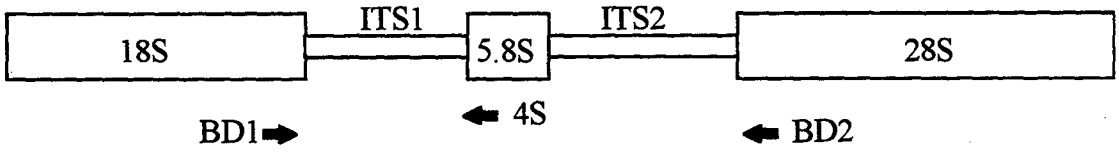


Fig. 3.1 Diagram of rDNA indicating position, direction and sequence of PCR primers used to amplify and sequence ITS1.

3.4 Cloning

Throughout this study, multiple variants were produced by PCR of both gene regions investigated (ITS1 and ND1). It was therefore neccessary to clone PCR products prior to sequencing. Cloning was done as per manufacturer's instructions (pMos Blue kit, Amersham, UK). Positive recombinant clones were picked, grown in 5ml Terrific broth (in the presence of 50μg/ml Ampicillin) overnight at 37°C and plasmids were extracted using standard alkaline lysis procedures as described in the Prism cycle sequencing manual (Perkin Elmer Cetus). However, the PEG precipitation was omitted at the end of the procedure. Plasmids containing inserts of the correct size were confirmed by *Eco* RI / *Hin*d III digestion of an aliquot at 37°C, 2 hrs. and examined by ethidium bromide stained 0.6% agarose gels before sequencing.

3.5 Sequencing

Recombinant clones were sequenced in both orientations as follows: 8μl terminator premix (Prism ready reaction dyedeoxy terminator cycle sequencing kit,

Table 3.2 PCR primers used in this study, their sequences and regions recognised by them.

Primer	Sequence (5' to 3')	Region recognised
BD1[1]	GTCGTAACAAGGTTTCCGTA	→ 18S rDNA
BD2[1]	TATGCTTAAATTCAGCGGGT	← 28S rDNA
4S[1]	TCTAGATGCGTTCGAA(G/A)TGTCGATG	← 5.8S rDNA
Parfor[2]	GGTGGAGCGTTTCTCCTCTGCCAT	→ *Paragonimus* "post-repeat" ITS1
Parrev[2]	GACCGGGTGCATGTCGGGCA	← " "
Pwfor[2]	TTATACTTGCAGCAGGGTGCC	→ *Paragonimus* ITS1 repeat region
Pwrev[2]	CGAACATCCAGTCACCCAGAC	← " "
Asitsfor[2]	TATTACCTGACCGGGGTATCT	→ schistosome "post-repeat" ITS1
Asitsrev[2]	CCTCTAAGCCAGCGTACATGC	← " "
JB11[3]	AGATTCGTAAGGGGCCTAATA	→ *Fasciola hepatica* partial ND1
JB12[3]	ACCACTAACTAATTCACTTTC	← " "
JB11par[2]	CAGAGGTTTGCTGATCTATTG	→ *Paragonimus*-specific partial ND1
JB12par[2]	CTTTCAGCCTCAGCATAATCC	← " "
USP[4]	GTTTTCCCAGTCACGAC	→ -40 Forward M13 phage vector
RSP[4]	AGCGGATAACAATTTCACACA	← Reverse M13 phage vector

Sources of primer sequences were [1]Bowles & McManus 1993, [2] this study, [3]Garey & Wolstenholme 1989, [4]Primers USP and RSP are Universal sequencing primers from bacteriophage M13 and are used to sequence cloned fragments from vector sequences flanking the cloning site. → identifies forward primers, ← identifies reverse primers.

Perkin Elmer) was added to 0.5µg of double stranded template and 3.2 pmoles of primer in a 20µl reaction volume. Universal and PCR primers were used for sequencing as follows: 96^{0}C, 1min followed by 25 cycles of 96^{0}C, 30 sec; 50^{0}C, 15 sec and 60^{0}C, 4 mins. Sequenced products were precipitated with 1/10 volume of 3M NaAcetate pH4.5 and 2.5 volumes of absolute ethanol for 30 mins at 0^{0}C, pelleted (20 mins. in a microfuge) and then washed in 70% ethanol before drying in a Speedvac. The pellets were sent to automated sequencing facilities at the University of

Queensland or Queensland Institute of Medical Research, where they were reconstituted in 4 µl of 8.3 mM EDTA, 0.83 volumes deionised formamide and denatured at 90^{0}C, 2 mins. before loading onto an ABI 373 DNA sequencer (as per instructions in the User's Manual).

At least 2 individual ITS1 clones were sequenced per size variant, per species (except for *P. macrorchis*, where only 1 clone was obtained) (Chapters 4, 5 &7). Any rare base differences found in the sequences of the otherwise identical clone pairs were ascribed to PCR misincorporations by *Taq* polymerase (estimated at 2 X 10^{-4} per nucleotide per cycle (Saiki *et al.* 1988). The situation for ND1 will be discussed in Chapter 8.

3.6 Sequence Alignment

Sequences were aligned in the sequence editing program ESEE (Version 1.09) (Cabot & Beckenbach 1989) or version 3.1s (Cabot 1997). DNAsis version 6 (Pharmacia) was used to construct Harr plots (dot plots) for identification of repeat sequences.

3.7 Phylogenetic analysis

Generally, ITS1 and ND1 data were subjected to the same treatment for phylogenetic analysis. This was done by parsimony in PAUP 3.11 (Swofford 1993), MP heuristic searches with the tree-bisection-reconnection (TBR) branch swapping algorithm in force, the addition sequence was simple. Distance methods (Neighbour Joining, Kimura 2-parameter model for ITS1 data or number of differences for ND1 amino acid sequences) were used in MEGA (Kumar, Tamura & Nei, 1993). The sequence data were resampled 1000 times by bootstrapping (with replacement). Generally bootstrap confidence limits (BCL) ≥ 70% only are reported (Hillis & Bull 1993). Alignment gaps were treated as missing data. Phylogenetic trees were generally rooted by midpoint rooting, without specifying an outgroup.

CHAPTER 4

Intra- and inter-specific variation in nuclear ribosomal internal transcribed spacer 1 of the *Schistosoma japonicum* species complex.

4.1 Abstract

The first internal transcribed spacer (ITS1) of the nuclear ribosomal DNA repeat was sequenced for members of the *Schistosoma japonicum* species complex (*S. malayensis*, *S. mekongi* and two geographical isolates of *S. japonicum*). The ITS1 is composed of three distinct regions: the 5' end (23 nucleotides); a tract of approximately 90 -140 nucleotides, which is repeated up to seven times in tandem, the number varying even within an individual in all species investigated in this study; the "post-repeat" region (378 nucleotides), which lacks repeats and is 3' to the repeat region. There is size and sequence variation among copies of the ITS1 repeat within a single individual. The relative abundances of size variants of ITS1 in *S. japonicum* have been ascertained by hybridizing genomic digests with an ITS1 probe. Multiple repeats and intra-individual variation in numbers and abundance of these is a feature of the Asian schistosomes, but not generally of African schistosomes, such as *S. mansoni*. The ITS1 repeat sequences described for African schistosomes are different to, and cannot be aligned with, those from the Asian species described here, whereas the remainder of the ITS1 can be aligned quite easily.

4.2 Introduction

As discussed in Chapter 2, nineteen species of the genus *Schistosoma* have been recognised to date, of which seven infect man: *S. mansoni* (*S. mansoni* group); *S. haematobium*, *S. intercalatum* and *S. mattheei* (*S. haematobium* group); *S. japonicum*, *S. mekongi* and *S. malayensis* (*S. japonicum* group) (McManus & Hope 1993). The most important species infecting man are *S. mansoni*, *S. haematobium* and *S. japonicum*. Studies on phylogeny and / or intraspecific variation in *Schistosoma*

species have recently been done using nuclear rDNA and mitochondrial DNA sequences (Després *et al.* 1992, Bowles, Blair & McManus 1995a, Littlewood & Johnston 1995, Barker & Blair 1996, Blair *et al.* 1997). Ribosomal DNA sequences are present as tandemly repeated clusters of highly conserved genes for 18S, 5.8S and 28S rRNA separated by variable spacer sequences (Long & Dawid 1980). There are about 100 copies of rDNA per haploid schistosome genome, each approximately 10 kb in size, of which about 4kb per copy is spacer (Simpson *et al.* 1984, Walker, Rollinson & Simpson 1986).

In Asia, the most important schistosome medically, is *S. japonicum,* with the Chinese and Philippine strains being regarded by some workers as very distinct. There has been some controversy concerning differences between these strains. No differences were found in randomly amplified polymorphic DNA (RAPDs) or ribosomal ITS2 sequences and only 0.5% difference was found in the mitochondrial CO1 sequence of these strains (Bowles, Blair & McManus 1995a). Likewise, triose phosphoisomerase (TPI) cDNA, paramyosin and NADH Dehydrogenase subunit 1 (ND1) sequences differed little between the strains (Hooker & Brindley 1996, Hooker *et al.* 1995, Bøgh *et al.* 1998). In addition a tandemly repeated *S. japonicum* specific genomic fragment demonstrated both inter- and intra-strain variation in the fragment sizes present (Drew *et al.* 1998). However, isoenzyme studies indicated a high level of genetic divergence (Woodruff *et al.* 1987, Merelender *et al.* 1987) between the strains. There are biological differences between the strains. In particular, infectivity and pathogenicity to the definitive host differs between the different geographic strains (McManus & Hope, 1993 and references therein), as does protective immunity after heterologous vaccination (Hope, Duke & McManus 1996).

In this Chapter, I investigate the primary structure of the ribosomal first internal transcribed spacer (ITS1) in the *S. japonicum* species complex. Both the 5' and 3' regions of ITS1, which are separated by a region of internal repeats, differ significantly (>5%) only between species. Here I present the interesting variation of ITS1 within individuals and between strains and species, which is due principally to the presence of the internal repeats, and the implications of such variation, with particular reference to the Chinese and Philippine strains of *S. japonicum*. A possible functional role for the

repeats is investigated in Chapter 6, whilst the phylogenetic signal of the "post-repeat" fragment is investigated in Chapter 7.

4.3 Materials and Methods

4.3.1 DNA extraction, PCR amplification, cloning, sequencing and alignment

These were done as described in Chapter 3, General Materials and Methods. In addition, internal ITS1 primers were designed to amplify the schistosome "post-repeat" tract only. These primers are AsitsF and AsitsR (see Table 3.2). PCR was done using these primers, in a volume of 20µl, as follows: 10 - 100 ng template and 10 pmol of each primer was added to 1 x *Taq* polymerase buffer (Promega), 400µM of each dNTP, 2.5 mM $MgCl_2$ and 0.25 Units of *Taq* Polymerase (Promega). After denaturation at 94^{0}C, 2 min., amplification reactions were as follows: 30 cycles of 94^{0}C, 1 min.; 50^{0}C, 30 sec. and 72^{0}C, 30 sec. on a Corbett Thermal cycler.

At least 2 individual clones were sequenced per cloned ITS1 size variant, per species. The few random point differences observed in clone pairs were considered to be due to gel reading problems and misincorporation during PCR.

Genbank numbers for the Philippine *S. japonicum* strain ITS1 sequences are U82282 (clone *SjP_a*) and U89868 (clone *SjP_(ab)a*). Chinese strain ITS1 sequences are U97533 (*SjC_a*), U97531 (*SjC_(ab)aL* and U97532 (*SjC_(ab)a*). *S. malayensis* ITS1 clone sequences are U82283 (*Sma_a*), U89869 (*Sma_(ab)a*) and U89870 (*Sma_(ab)$_2$a*). *S. mekongi* ITS1 clone sequences are U82284 (*Sme_(ab)a*) and U89871 (*Sme_(ab)$_2$a*). *S. mansoni* ITS1 sequence is AF029309.

4.3.2 Genomic digests and Southern blots

S. japonicum genomic DNA from 5 pooled individuals was digested by standard techniques (Sambrook, Fritsch & Maniatis 1989) with *Hind* III and *Xba* I. *Hind* III cuts 200 bp upstream of the 3' end of 18S rDNA and *Xba* I cuts 133 bp downstream of the 5' end of ITS2, to remove complete ITS1 from the rest of the genome (with flanking 18S and 5.8S rDNA - partial ITS2 fragments, accounting for 483 bp). Digestion should excise a fragment expected to be approximately 1000 bp, of which 500 bp is ITS1 only, in the smallest variant. Similarly, *S. mansoni* genomic DNA

was digested using *Alu* I and *Bln* I, which have respectively, a single cut site towards the 5' and 3' ends of ITS1, to produce a 300 bp ITS1 fragment. Genomic digests of *S. japonicum* and *S. mansoni* were run on an 0.8% agarose gel. DNA was transferred to Hybond N+ membranes (Amersham) as described by Southern (1975). The filters were hybridised to an α-P^{32} dATP random-primed labelled (Boehringer Mannheim) 280 bp ITS1 fragment, which had been generated by PCR amplification of genomic template with internal primers (AsitsF & AsitsR). The probe did not contain any of the ITS1 repeat sequence, as the primers only amplify the "post-repeat" fragment, which is 3' of the repeats. Both membranes were hybridised and washed at 50 °C. Washes were done to a stringency of 0.5 X SSC, 0.1% SDS before exposure to a Phosphor Imager screen (Molecular Dynamics) and scanning. Relative band intensities were determined using ImageQuant (Molecular Dynamics).

4.4 Results and Discussion

Internal repeats are located only near the 5' end of ITS1, in the *S. japonicum* species complex. The repeats are preceded by 23 bp of sequence at the 5' end and followed by 378 bp at the 3' end of ITS1 (the latter discussed in Blair *et al.* 1997). No repeats have been found elsewhere within the ITS1 of *S. japonicum* and related species, unlike the situation in *S. mansoni* and other African schistosomes (Kane *et al.* 1996), where an isolated copy of the internal repeat is sometimes present near the 3' end of ITS1. Although the strain of *S. mansoni* used in this study was from the Caribbean, there was only a single base difference in the ITS1 sequence of the Puerto Rican sample, compared to other isolates sequenced. Repeats within the *S. japonicum* species complex align well, as do such repeats among the African schistosomes studied to date. However, alignment is not possible between repeats of the African schistosomes and those of the *S. japonicum* species complex. Neither is alignment possible between these groups at the 5' end of ITS1 (there is only 62.5% similarity), although *S. japonicum*, *S. malayensis* and *S. mekongi* differ by <10% in this region (data not shown). There is however good similarity between the 3' tract (post -repeat) of ITS1 in both African and Asian groups (94% similarity), except for the 30 bp immediately prior to the 5.8S rRNA coding region.

Different numbers of the repeat element (in two parts: *a* ~90 nt in all 3 species, followed by *b* ~15 nt in *S. mekongi* and *S. malayensis* and 39 or 53 nt in *S. japonicum*) occur, commencing 24 nucleotides from the 5' end of the ITS1. Length variants in the repeat region in a single individual can be demonstrated by PCR amplification (Fig. 4.1). ITS1 clones sequenced (from the DNA of pooled worms) were arbitrarily termed small (smallest sequenced clone), medium (intermediate size sequenced clone) or large (largest clone sequenced). These did not represent all the size variants present, as detected from genomic digests of DNA from pooled worms. In fact, the most abundant ITS1 variants present in the genome, were larger ($(ab)_2\ a$ and $(ab)_4\ a$ in the Philippine strain and $(ab)_4a$, $(ab)_5a$ in the Chinese strain (Fig. 4.2). Furthermore, the smaller variants sequenced were barely detectable by Southern blot analysis, so they are much less abundant in the genome. PCR and cloning procedures are known to favour smaller fragments (Sambrook, Fritsch & Maniatis 1989, Saiki *et al.* 1988), which would explain why these were over-represented among the clones sequenced.Furthermore, clones were routinely screened prior to sequencing by *Eco RI – Hin* dIII digestion , which removes the insert (plus flanking sequence) from the multiple cloning site of the vector. During the screening process of 5 - 10 randomly selected ligated vectors large inserts were not detected (not shown).

Variation among repeat sequences, both within and between clones of the same species, was 0 - 7.5%, with the greatest intra-individual variation occurring between a_1 and the fourth copy a_4 of the largest clone of *S. malayensis* sequenced (Fig. 4.3 & Table 4.1). A similar situation occurs in *S. mekongi.* Copies of *a* at the same position share more similarity across clones than they do within. Thus the first *a* repeat in the largest clone is more similar to the first *a* repeat in smaller clones from the same species than to the second third or fourth *a* repeats in the largest clone (Table 4.1). The *S. japonicum a* region is somewhat divergent relative to the other two Asian species (Table 4.1).

Table 4. 1

% Pairwise differences of repeat *a* sequences in "large" ITS1 clones in three Asian schistosomes.

	S.mal a_1	S.mek a_1	SjC a_1	SjP a_1
S.mal a_1	-	2.5	7.5	8.75
S.mal a_2	1.25	2.5	7.5	10.0
S.mal a_3	2.5	3.75	10.0	11.25
S.mal a_4	7.5	8.75	15	13.75
S.mek a_1	2.5	-	8.75	10.0
S.mek a_2	1.25	1.25	6.25	8.75
S.mek a_3	2.5	2.5	8.75	10.0
S.mek a_4	5.0	3.75	10.0	11.25
SjC a_1	7.5	8.75	-	1.25
SjC a_2	7.5	8.75	1.25	2.5
SjP a_1	8.75	10.0	1.25	-
SjP a_2	8.75	10.0	2.5	0

S. mal = *S. malayensis*, S. mek = *S. mekongi*, SjP = *S. japonicum* (Philippine), SjC = *S. japonicum* (Chinese).

Repeat element *b* is smaller and much less variable than *a*, such that no variation is found either among *S. malayensis* clones or *S. mekongi* clones. The *b* repeat is identical in *S. malayensis* and *S. mekongi*. The b_1 sequence in both strains of *S. japonicum* differs dramatically from the other two species, due to the presence of a 24 nucleotide insert containing a $(TA)_8$ microsatellite (*SjP_(ab)$_2$a*-1 in Fig. 4.3) and 4/11 base differences prior to the insert mentioned above, (ie. a 29% variation from the b_1 sequence of the other two species). The Chinese strain had the same $(TA)_8$ microsatellite insert (*SjC_(ab)$_2$a*-1 in Fig. 4.3) and an additional b_1 variant detected by direct sequencing (of a band excised from a gel), not observed in the Philippine strain, which contained an expanded $(TA)_{18}$ microsatellite (*SjC_(ab)$_2$aL*1 in Fig. 4.3).

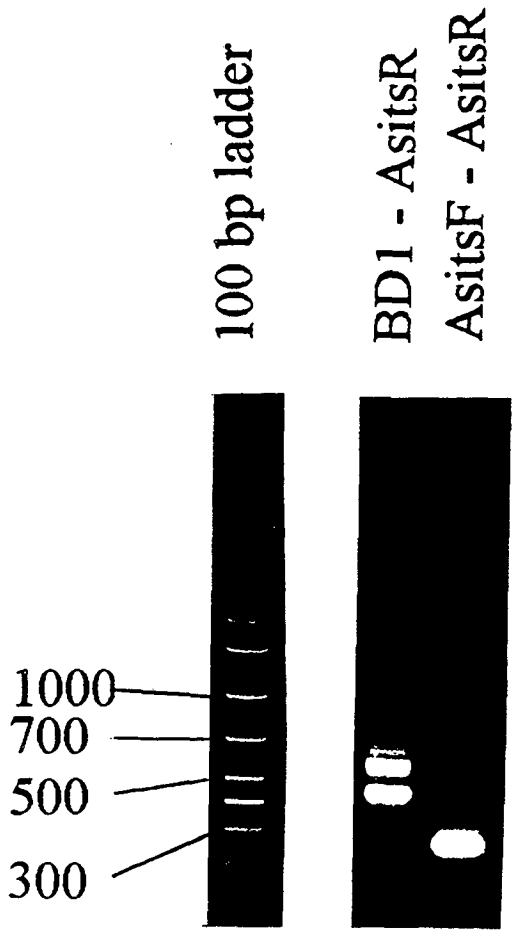


Fig. 4.1. ITS1 PCR products of a single adult *S. japonicum* (Chinese isolate). Lane 1: 100 bp ladder (Biorad). Lane 2: ITS1 spanning both repeated and "post-repeat" components (primers BD1 - AsitsR), bandsizes most strongly amplified correspond to 490 bp (*SjC_a*)and 620 bp (*SjC(ab)a*), with weaker products corresponding to 640 bp (*SjC(ab)aL*) and 800 bp (*SjC(ab)$_2$aL*). Lane 3: "Post - repeat" component of ITS1 only (primers AsitsF - AsitsR), which corresponds to 280 bp of the 378 bp fragment. *SjC* is the Chinese strain of *S. japonicum*, *(ab)$_n$a* describes the nature of the repeat and *L* indicates that it is the large ITS1 variant.

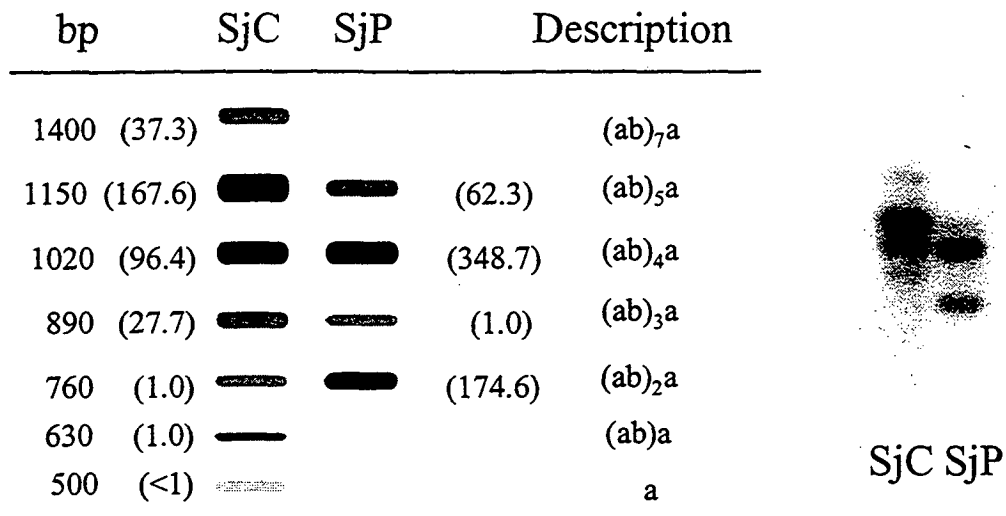


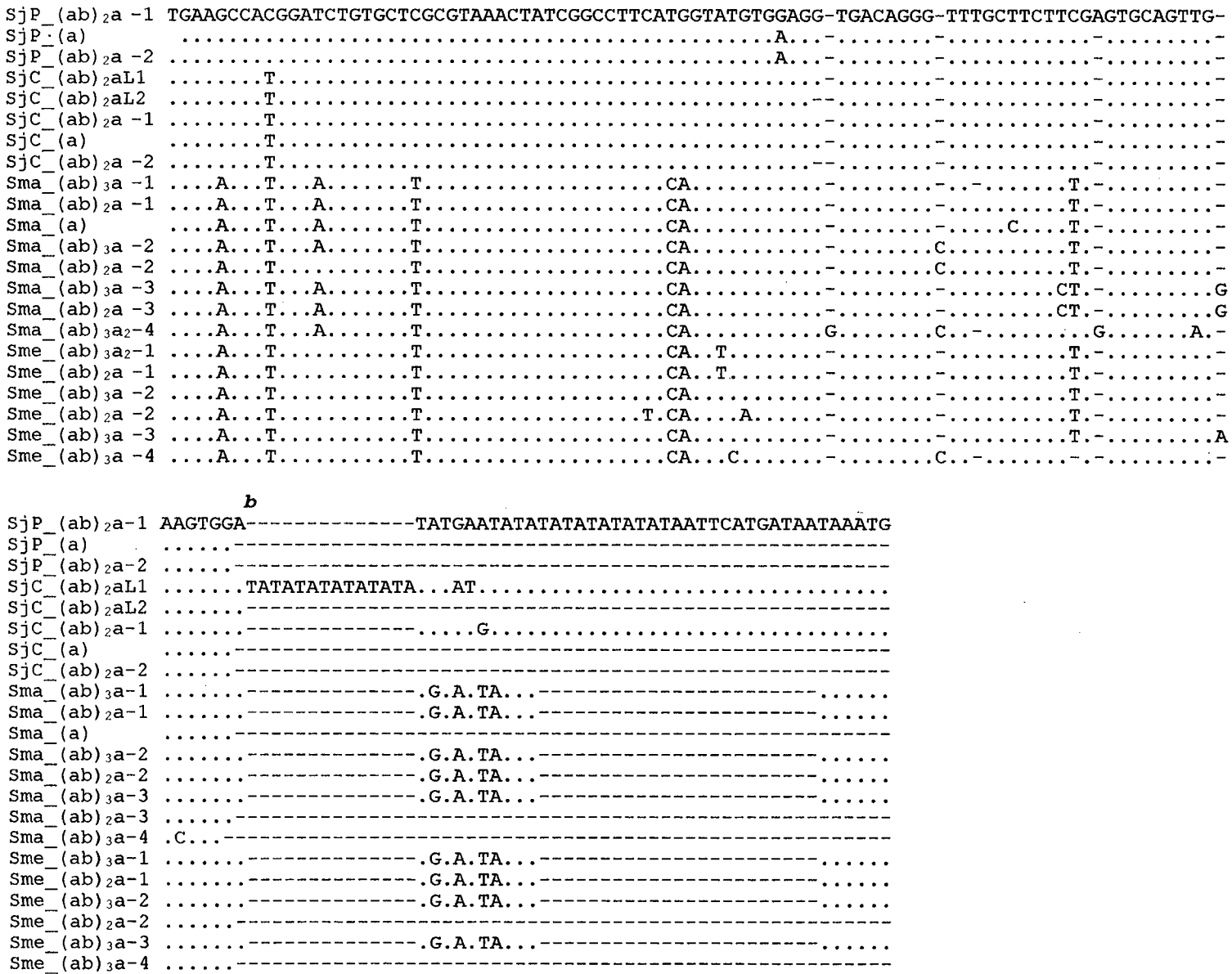
Fig. 4.2 a) Diagram based on Southern blots (b) of *Hind* III and *Xba* I restriction digested Chinese *S. japonicum* (*SjC*) and Philippine *S. japonicum* (*SjP*) genomic DNA (from 5 pooled individuals in each case), probed with P^{32} - labelled, AsitsF-AsitsR amplified PCR fragment of the ITS1 "post-repeat"from *S. japonicum.* Relative abundance, as determined by ImageQuant area curves, is given in brackets for ITS1 size variants. Sizes indicated are for the actual size of the ITS1 and do not include the flanking 18S and 5.8S rDNA - ITS2 fragments (not to scale).

The relative abundances of size variants of ITS1 were determined by Southern blots of genomic digests of pooled DNA from adult worms of *S. japonicum* (Chinese and Philippine strains) (Fig. 4.2) and of *S. mansoni* (data not shown). ITS1 was isolated from the rest of the genome by restriction digestion, as described in materials and methods. The most abundant ITS1 size variants in the Chinese strain were 1150 bp and 1020 bp, consistent with them representing *SjC_(ab)$_5$a* and *SjC_(ab)$_4$a* respectively. The relative abundance of each was 50.8 and 29.2 % respectively, with a reduction in relative abundance as the number of repeats increases or decreases (Fig.

4.2). In a Philippine strain from Sorsogon only 4 variants were detectable, of which 1020 bp and 760 bp were the most abundant, consistent with *SjP_(ab)$_4$a* and *SjP_(ab)$_2$a* respectively, (59.3 and 29.7 % relative abundances respectively). *S. mansoni* from Puerto Rico was digested and probed in a similar manner as described above for the Asian schistosomes. As expected from published *S. mansoni* ITS1 data (Kane *et al.* 1996), there was only one ITS1 length class present in the genome. This appears to be the case for all African species for which ITS1 sequences have been published, as they lack intra-individual variation in the ITS1 (Kane & .Rollinson 1994, Littlewood & Johnston 1995, Kane *et al.* 1996).

Many variant forms of ITS1 have been described, within and among individuals across a range of organisms, eg. the dipterans: *Simulium damnosum* species complex (Tang *et al.* 1996), *Drosophila* (Schlötterer *et al.* 1994) and *Aedes* species (Wesson, Porter & Collins 1992), the orthopteran *Melanopline* species (Kuperus & Chapco 1994), the coleopteran *Cicindela dorsalis* (Vogler & De Salle 1994), the cestode *Echinococcus* (Bowles, Blair & McManus 1995b), the trematodes *Dolichosaccus* species (Luton, Walker & Blair 1992) and *Schistosoma* species (Kane & Rollinson 1994, Kane *et al.* 1996) and the mammal *Homo sapiens* (Gonzalez *et al.* 1990). The degree of sequence variation is a balance between processes generating variation and homogenising processes (Schlötterer *et al.* 1994, Fritz *et al.* 1994). Variation can be generated by many factors, such as sexual reproduction (due to recombination during the formation of bivalents and chiasmata), dispersal of rDNA loci on various chromosomes (>1 nucleolar organiser region (NOR) (Vogler & De Salle 1994), interbreeding with sibling species and environmental pressure, which may select for minor alelles. Known homogenising processes are those involved in concerted evolution (Dover 1982, Hillis & Davis 1988, Hillis *et al.* 1991, Linares, Bowen & Dover 1994, Schlötterer *et al.* 1994, Tang *et al.* 1996).

Fig. 4.3 (overleaf). Alignment of ITS1 repeats from Asian schistosomes. A "." indicates the same base as shown on upper line; "-" indicates alignment gaps; *SjP_(a)* from small *S. japonicum* (Philippines) clones, *SjP_(ab)a* from larger *S. japonicum* (Philippines) clones, *SjC_(a)* from small *S. japonicum* (Chinese) clones, *SjC_(ab)a* from large *S. japonicum* (Chinese), *SjC_(ab) aL* from large *S. japonicum* (Chinese) with additional $(TA)_7$, *Sma_(a)* is the "small" *S. malayensis* clone, *Sma_(ab)$_2$a* (1-3) are from "medium" *S. malayensis* clones, copies numbered according to positional order within clone, *Sma_(ab)$_3$a* (1 - 4) are from "large" *S. malayensis* clones, copies numbered according to positional order within clone. *Sme_(ab)a* (1-2) are from "small" *S. mekongi* clones, copies numbered according to positional order within clone, *Sme_(ab)$_3$a* (1 - 4) are from "large" *S.mekongi* clones, copies numbered according to positional order within clone. An "***a***" indicates the start of *a* repeat and "***b***" indicates start of *b* repeat in the alignment.


a
SjP_(ab)2a -1 TGAAGCCACGGATCTGTGCTCGCGTAAACTATCGGCCTTCATGGTATGTGGAGG-TGACAGGG-TTTGCTTCTTCGAGTGCAGTTG-
SjP_(a)
SjP_(ab)2a -2
SjC_(ab)2aL1
SjC_(ab)2aL2
SjC_(ab)2a -1
SjC_(a)
SjC_(ab)2a -2
Sma_(ab)3a -1
Sma_(ab)2a -1
Sma_(a)
Sma_(ab)3a -2
Sma_(ab)2a -2
Sma_(ab)3a -3
Sma_(ab)2a -3
Sma_(ab)3a2-4
Sme_(ab)3a2-1
Sme_(ab)2a -1
Sme_(ab)3a -2
Sme_(ab)2a -2
Sme_(ab)3a -3
Sme_(ab)3a -4
b
SjP_(ab)2a-1 AAGTGGA----------TATGAATATATATATATATAATTCATGATAATAAATG
SjP_(a)
SjP_(ab)2a-2
SjC_(ab)2aL1 TATATATATATATA...AT
SjC_(ab)2aL2
SjC_(ab)2a-1 .....G.
SjC_(a)
SjC_(ab)2a-2
Sma_(ab)3a-1 .G.A.TA...
Sma_(ab)2a-1 .G.A.TA...
Sma_(a)
Sma_(ab)3a-2 .G.A.TA...
Sma_(ab)2a-2 .G.A.TA...
Sma_(ab)3a-3 .G.A.TA...
Sma_(ab)2a-3
Sma_(ab)3a-4 .C...
Sme_(ab)3a-1 .G.A.TA...
Sme_(ab)2a-1 .G.A.TA...
Sme_(ab)3a-2 .G.A.TA...
Sme_(ab)2a-2
Sme_(ab)3a-3 .G.A.TA...
Sme_(ab)3a-4

The presence of intra-individual variation in ITS1 of the Asian schistosomes and absence of such variation in the African schistosomes may be explained by two factors, both of which may contribute to a reduction in concerted evolution processes in the Asian species. It is concerted evolution that generally eliminates variation in the multiple copies of ribosomal genes found within individuals and across populations of a species. Firstly, putative "hot spots" for recombination (Chi-like sites), present in the internal repeats of ITS1 of the African schistosomes (Kane & Rollinson 1994, Kane *et al.* 1996), are absent from the internal repeats of the ITS1 of the *S. japonicum* species complex. Secondly, the chiasma frequency in *S. japonicum* is extremely low (FXi = 3.0), whereas in *S. mansoni* FXi = 15.3 (Hirai *et al.* 1996). Similar levels of chiasma formation have been determined in mammals (FXi = 15.3 for Chinese hamsters and FXi = 17.3 for mice) as in *S. mansoni* (Hirai *et al.* 1996 and refs. therein). This suggests that genetic recombination due to crossing-over at chiasmata is reduced in *S. japonicum*. Both absence of recombination at "hot spots" in the ITS1 and reduced genome-wide chiasma formation in the *S. japonicum* species complex would explain the apparent reduction in concerted evolution processes, which would in turn explain why there is increased intra-individual variation in the ITS1 of the *S. japonicum* species complex.

There appears to be a difference in the rates / patterns of concerted evolution of repetitive ribosomal genes within the genus *Schistosoma*. Such differences presumably occurred subsequent to the speciation event which produced the *S. japonicum* lineage, but prior to the speciation within the *S. japonicum* lineage. It would be most interesting to investigate the ITS1 sequence of another Asian schistosome species complex, the *S. sinensium* complex and identify whether internal repeats occur and if they do whether they are *S. mansoni* or *S. japonicum* - like, as the *S. sinensium* complex may be a sister group to the *S. japonicum* species complex.

Chapter 5

Intra - and inter- individual variation in ITS1 of *Paragonimus westermani* (Trematoda: Digenea) and related species.

5.1 Abstract

Numerous clones containing ITS1 PCR products were sequenced for *P. macrorchis* and members of the *P. ohirai* and *P. westermani* species complexes. In *Paragonimus*, as in the *Schistosoma* species the ITS1 is composed of three distinct regions: the short 5' end, followed by a tract of approximately 120 nucleotides which occurs a variable number of times in tandem, and the "post-repeat" region 3' to the repeats which lacks repeats. Sequences from all three regions can be aligned among the species studied. Intra-individual sequence variation in *P. westermani* was sometimes greater than between individuals of the species complex. In the *P. ohirai* species complex, however, sequence variation within individuals was minimal. Possible reasons for these observations are discussed in the light of what has already been observed in the Asian *Schistosoma* species. I also wished to determine which of the length variants detected were the most abundant variants present in *Paragonimus* species. This was done by probing Southern blots of genomic digests with an ITS1 fragment which lacks repeat sequences, as was done for the *Schistosoma* species (Chapter 4). As in *Schistosoma,* there is generally greater abundance of large variants, with much lower abundance of the small variants, such as those sequenced. Differences in ITS1 lengths are largely attributed to differing numbers of repeats, though some exceptions (which are discussed) were found.

5.2 Introduction

In this chapter I describe the intra-individual variation detected in ITS1 sequences of *Paragonimus* species. The utility of the ribosomal first internal transcribed spacer (ITS1) for phylogenetic studies of both *Paragonimus* and *Schistosoma* will be further investigated in Chapter 7. In addition, the possible functional implications of variation in ITS1 detected in *Schistosoma* (Chapter 4) and *Paragonimus* (this Chapter), is further investigated in Chapter 6.

Studies on trematodes indicate that sequences in the ITS1 might be less conserved than ITS2 as has been shown for the Asian *S. japonicum* species complex (Chapter 4) and for another trematode genus, *Dolichosaccus* (Luton, Walker & Blair 1992) and thus provide greater phylogenetic resolution among members of a species complex such as *P. westermani*. I therefore characterized the ITS1 for a number of species and strains of *Paragonimus*, with emphasis on members of the *P. westermani* complex. The results of this work are presented here.

5.3 Materials and Methods:

Details about samples used in this study are given in Chapter 3 (General Materials and Methods), as is methodology relating to DNA extraction, PCR, cloning, sequencing and sequence alignment. Additional primers, used specifically for the amplification of *Paragonimus* ITS1, were Pwfor and Pwrev (see Table 3.2), which were developed to sequence through the repeat region of large ITS1 variants of *Paragonimus* species. Another set of primers was designed to amplify only the "post-repeat" region of ITS1. These were parfor and parrev (see Table 3.2), which were used firstly, to look for length variants beyond the repeat region within ITS1, secondly, as sequencing primers and thirdly, to amplify the "post-repeat" fragment of ITS1 only for use as a probe during Southern blotting investigations of the ITS1. Genbank numbers for ITS1 sequences used in this study are presented in Table 5.1 and ITS1 primary structure, throughout the ITS1 is shown diagrammatically in Fig. 5.1, with repeat sequence alignment only presented in Fig. 5.2.

Genomic digests and Southern blots: Genomic DNA from *Paragonimus* species was double digested with *Eco*R I and *Taq* I by standard techniques (Sambrook *et al.* 1989) to cut almost intact ITS1, from the genome. *Eco* R I cuts ITS1 only once, near the 5' end, whilst *Taq* I cuts only once near the 3' end. Digestion should excise a fragment consisting of ITS1 only, which is expected to be approximately 450 bp in the smallest variants. Genomic digests were separated on 0.8% agarose gels, DNA was transferred to Hybond N+ membranes (Amersham, as per Southern, 1975) and fixed by UV cross-linking for 10 minutes. The probe used was an α-P^{32} dATP random-primed labelled PCR product amplified from the ITS1 of *P. westermani* using primers

Table 5.1: *Paragonimus* species investigated here, indicating geographic origin of isolate, source, whether complete or [3' only]* ITS1 sequences have been obtained for particular samples (number of clones sequenced) - applies only to complete ITS1 sequences from cloned material / whether Southern blot (Sb) data is available for samples and Genbank numbers for variants. Each line represents a single worm (see Table 3.1)

	Species & code (ploidy)	Geographic location	complete (# clones)/ partial* ITS1 sequence	Southern blot	Genbank number
P. westermani	PwJO(2n)	Japan, Ohita	complete (7)	no	AF040925 - 040927
species	PwJH (2n)	Japan, Hyogo	partial	yes	AF040941
complex	PwJM(2n)	Japan, Mie	partial	yes	AF040942
	PwPhil(2n)	Philippines, Sorsogon	partial	yes	AF040943
	PwMal(2n)	Malaysia, K.P.	no	yes	-
	PwMal (2n)	Malaysia, U.L.	complete (6)	no	AF040934 - 040935
	PwC(2n)	China, Minchin	partial	no	AF040940
	PwC(?n)	China, ?	complete (6)	yes	AF040928 - 040929
	PwJ(3n)	Japan, Amakusa	complete (6)	no	AF040933
	PwC(3n)	China, Xigutai	complete (6)	no	AF040932
P. ohirai	Pohirai FF	Japan, Tanegashima	complete (3)	no	AF040930
species	Pohirai J	Japan, Kinosaki	complete (3)	no	AF040931
complex	Piloktsuenensis	Japan, Amami (279)	partial	yes	AF040939
	Pikoktsuenensis	Japan, Amami (K)	partial	yes	AF040938
other	Pmacrorchis	Thailand	complete (1)	no	AF040936
species	Pmiyazakii	Japan, Miyazaki	partial	yes	AF040937

Key: ? When relevant information was not available, K.P. = Kualah Pilah, U.L. = Ulu Langat, . 3' Fragments were amplified by primers Parfor & Parrev & sequenced directly from PCR products, as described in Materials & Methods.*

parfor and parrev. These amplify only the "post-repeat" region 3' to the repeats. Hybridisation was done at 50°C. The membranes were washed at 50°C to 0.5 X SSC, 0.1% SDS, exposed to a Phosphor Imager screen (Molecular Dynamics, California) overnight and scanned.

5.4 Results and Discussion

In the genus *Paragonimus,* as in the schistosomes (Chapter 4) the ITS1 is composed of three distinct regions: (i) the 5' end (80 nucleotides), which is 75% similar among the species complexes of *P.ohirai* and *P. westermani*; (ii) a tract of approximately 120 nucleotides, which occurs a variable number of times within individuals and is composed of 2 subrepeats (*a* = 82nt and *b* = 38 nt), which occur in tandem and *b* (if it is present) is always flanked on both sides by *a*; (iii) the "post-repeat" region, (346 nucleotides), which is 3' to the repeats and lacks repeats. This region is 90% similar among the species (see diagram, Fig. 5.1).

5.4.1Variation in repeat sequences:

PCR amplification of ITS1 from individual worms produced multiple bands ranging in size from approximately 400 - 2000 bp (not shown), similar to what was observed for *S. japonicum* (Fig. 4.1). Analysis of cloned PCR products from a single individual showed that multiple bands were due to the presence of varying numbers of repeats within variant fragments. Among the sequenced variants there were between 1 and 3 copies of the repeat element which occurred in tandem (minimum a, maximum sequenced $(ab)_2a$) (Figs.5.1 &5.2). All *P. ohirai* clones sequenced (5 clones from 2 isolates) had $(ab)_2a$ type repeats only (748 nt). No other variants occurred in sequenced clones. Variation among repeat sequences from an individual was greater from one copy to the next within a clone (eg. a_1 vs a_2 vs a_3), than between copies at the same position in different clones from the same individual, indicating positional stratification of repeats (Fig. 5.2).

5.4.2 Variation in "post-repeat' regions:

Alignments of 26 "post-repeat" ITS1 sequences were used to construct phylogenetic trees. In most cases, alignment was straightforward. Differences among the post-repeat sequences are further characterised for phylogenetic signal in Chapter 7.

Fig. 5.1 Diagrammatic presentation of sequenced ITS1 variants, indicating intra-individual differences in *P. westermani* and comparing these to clones from *P. ohirai* and *P. macrorchis*. Isolate, variant description, identity code (as in Table 5.1) and individual clone identity, inferred size (in bp) and number of clones represented of intact ITS1 fragments are shown. In all isolates shown, only one individual was sequenced, except for *P. ohirai*, for which 2 individuals were sequenced from different geographic localities.

species & code	variant	clone id	Schematic ITS1 variants	Size (bp)	#clones
P. westermani (2n) Japan, Ohita PwJO(2n)			18S ITS1 5.8S		
	a	7c		508	3
	a, del3'-5.8S	107,108,110		408	2
	del a	7e, 102		408	2
P. westermani (3n) Japan and China PwJA & CL(3n)	a	eg. U7		508	2
	aba	eg. U6		628	2
	$(ab)_2a$	eg. U9		748	2
P. ohirai	$(ab)_2a$	PoT, PoK		748	5
P. macrorchis	$(ab)_2a$	Pmac		748	1

* One individual each from Japan & China, each possessing the 3 variants indicated here.

```
          a
PoT- 1    TAGCCTCGGCTGCCTAAGGTGGAGCGTTTCTCCTCTGCCATTCGGTCCGGCTGTGCCTGGCGCATTCGTGTGCCTGCGCACAG
PoT- 2    C..A....T.....................................................T.....G..............
PoT- 3    C..A....T.....................................................T.....G..............
PoK- 1    ..............................................................T.....G..............
PoK- 2    C..A....T.....................................................T.....G..............
PoK- 3    C..A....T.....................................................T.....G.A............
Pmac-1    C.......T......C...........C.............T.A.T..A.......A......C.T........A..T....
Pmac-2    C...T.T.T......C...........C.............T.A.TT.A.............C.T........A..T....
Pmac-3    C...T.T.T......C...........C...............ACTGT..............C.T........A.......
PwJ3nlg 1 C..A....T......CG..........-............GT.T.T..............A..C..........A..T....
PwJ3nlg 2 C.....T.T......CG..........-............GT.T.T..............A..C..........A..T....
PwJ3nlg 3 C.....T.T......CG..........-............GT.T.T..............A..G..........A..T....
PwJ3nsm 1 C..A....T......CG..........-............GT.T.T..................C..........A..T....
PwC3nlg 1 C..A....T......CG..........-............GT.T.T..............A..C..........A..T....
PwC3nlg 2 C.....T.T......CG..........-............GT.T.T..............A..C..........A..T....
PwC3nsm 1 C..A....T......CG..........-............GT.T.T..............A..C..........A..T....
PwCn    1 C..A....T......CG..........-............GT.T.T..............A..C..........A..T....
PwJ-110 1 C..A....T......CG..........-............G..T.AT.......T.G...A..C..........A..T....
PwJ-7e  1 ------------------------------------------------------------------------------------
PwM2nlg 1 C.......T......CG..........-............GT.T.T.....A..........C..........A..T....
PwM2nlg 2 C.......T......CG..........-............GT.T.T.....................A..T....
PwM2nsm 1 C.......T......CG..........-............GT.T.T..................C..........A..T....

          b
PoT-1     TCGGATCTTTTCTGATGAGCCTTCGGGTTTGTTGGATGC
PoT-2     ....................................
PoK-1     ....................................
PoK-2     ................A...................
Pmac 1    .....G.......C......................
Pmac 2    ....TC......C.......................
PwJ3nlg 1 .T.......C..C.......................
PwJ3nlg 2 .T.......C..C.......................
PwC3nlg 1 .........C..C.......................
PwM2nlg 1 .........C..C.......................
```

Figure 5.2: ITS1 repeat sequences of 12 example clones from *Paragonimus* species aligned showing positional stratification of repeats. Codes for clones are as in Table I. Numbers at the end of clone names indicate position in the repeat array, 1 being at the 5' end, 2 following 1, etc. A "." indicates the same base as shown on upper line; "-" indicates alignment gaps; >***a*** = start of repeat sequence *a*, >***b*** = start of repeat sequence *b*.

5.4.3 Relative abundance of ITS1 size variants:

The relative abundances of size variants of ITS1 were determined by Southern blots of genomic digests of standardised amounts of DNA from individual adult worms of *P. miyazakii* and members of the *P. westermani* and *P. ohirai* complexes (Table 5.1 & Fig. 5.3). ITS1 was isolated from the rest of the genome by restriction digestion (see materials & methods). The ITS1 can vary from 550 nucleotides to greater than 2000 nucleotides in size (Fig. 5.3). However, only the smaller fragments were represented among the clones sequenced. This may be an artefact of cloning, as plasmid vectors containing smaller, rather than larger inserts have higher transformation efficiencies (Sambrook, Fritsch & Maniatis 1989). These results demonstrate that the ITS1 clones sequenced did not represent all the size variants present. In addition, the small variants are quantitatively in the minority among the length variants present in the genome.

It is not known that the large size variants are entirely due to multiple copies of repeat sequence, as they were not sequenced. However when genomic templates were amplified with primers "parfor" and "parrev", which amplify the entire 3' region of ITS1 post - repeats, a single 350 nt fragment resulted (not shown), which suggests that large variants generally do not have inserts beyond the repeat region. This was done for all isolates used in the phylogenetic study. *P. macrorchis* did however have an additional smaller PCR band of 200 bp, indicating that it does have an ITS1 variant with a deletion in the region 3' to the internal repeats, such as was present in the diploid Japanese *P. westermani* isolate from Ohita (Figs. 5.1 and 5.2).

5.4.4 ITS1 variants: implications for processes of concerted evolution

I have demonstrated that considerable intra-individual differences occur in the ITS1 of members of the *Paragonimus westermani* complex (and in other species of the genus). These differences are due both to varying numbers of repeats and to sequence differences. Furthermore, the clones which were sequenced may represent only a small minority of the ITS1 variants present in the genome.

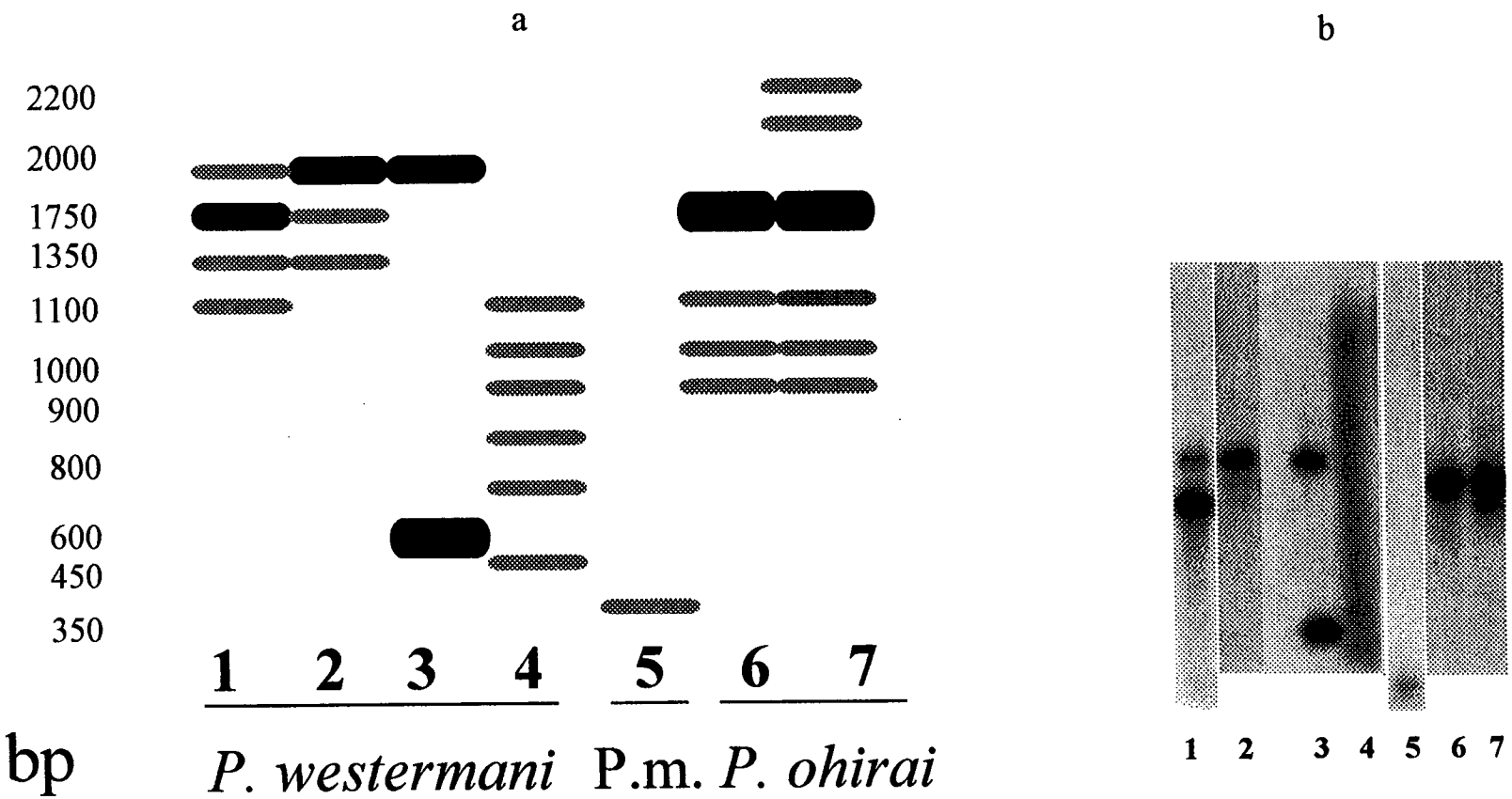


Figure 5.3 :Diagram (a) and autoradiograph (b) of Southern blot of restriction digested (*Eco* RI & *Taq* I) individuals of the *P. westermani* species complex, the *P. ohirai* species complex (*P. iloktsuenensis*) & *P. miyazakii*, probed with P^{32} labelled "post-repeat" ITS1 fragment. Sizes indicated are for the actual sizes of intact ITS1 rather than for the fragments excised by digestion. The specimens shown are: *P. westermani* (1 - 4) from Hyogo, Japan (2n) =1; Mie, Japan (2n) = 2; Sorsogon, Philippines (2n) = 3; China (?n) = 4; *P. miyazakii* = 5; *P. ohirai* complex (6 - 7) - *P. iloktsuenensis* (individual 279) (6) & *P. iloktsuenensis (individual K)* (7).

Among trematodes, members of the genera *Dolichosaccus* and *Schistosoma* have repeats, similar to those reported here, embedded in the ITS1 (Luton, Walker & Blair 1992, Kane & Rollinson 1994, van Herwerden, Blair & Agatsuma 1998). Such repeats are absent from the genus *Echinostoma* (Morgan & Blair 1995). The repeats and possible functional significance of them, are the subject of the next chapter (Chapter 6).

The homogenising processes involved in concerted evolution of tandemly repeated genes, such as the ribosomal DNA genes (Dover 1982, Hillis & Davis 1988, Hillis *et al.* 1991, Linares, Bowen & Dover 1994, Schlötterer *et al.* 1994, Tang *et al.* 1996), are not operating uniformly within each of the trematode genera. Sequence variation may exceed homogenisation due to factors such as dispersal of rDNA loci on various chromosomes (>1 NOR) (Vogler & De Salle 1994) or interbreeding with sibling species. Alternatively, reduced genetic recombination due to a reduction of chiasmata may cause a reduction in the rate of concerted evolution. Indeed, the frequency of chiasma formation (FX) can vary enormously within a genus, as has been observed for *Schistosoma japonicum* and *S. mansoni,* (FX = 3 and 15 respectively) (Hirai *et al.* 1996). The frequency of chiasma formation (FX) in *P. ohirai* is similar to that found in *S. mansoni* (Hirai *et al.* 1996). However, FX has not yet been determined for *P. westermani*. The schistosome species differ with respect to intra-individual variation in ITS1 sequences (van Herwerden, Blair & Agatsuma 1998), as do the *Paragonimus* species described in this study. The ITS1 data presented here suggest that *P. westermani* may indeed have a reduced FX relative to *P. ohirai*, as has been observed for *S. japonicum* relative to *S. mansoni.* This has not yet been determined for *P. westermani*.

CHAPTER 6

Multiple repeats in abundant first internal transcribed spacer (ITS1) variants of some trematodes (Digenea) contain regulatory motifs.

6.1 Abstract

Sequence motifs in transcriptional promoters can initiate and regulate transcription and have been identified in the external transcribed spacer (ETS) of the ribosomal DNA (rDNA) of many taxa. In addition, transcription of ribosomal genes can be increased in the presence of repeats within the ETS that contain promoter motifs. If these same motifs occur elsewhere in the genome, they may also act as transcription regulators. I identified multiple repeats, varying in number within individuals, within the first internal transcribed spacer (ITS1) of some species in three genera of trematodes (*Paragonimus, Schistosoma* and *Dolichosaccus*). In all three genera the repeats have two segments where, if one (segment "b") is present, it is always flanked on both sides by the other (segment "a"). In species of *Paragonimus* and *Schistosoma,* ITS1 variants that have more copies of repeat sequences were more abundant than variants with fewer repeats (Chapters 4 & 5). I report the presence of enhancer-like motifs in the "b" segments of ITS1 in two of three trematode genera that have multiple repeats. If these variants were evolving neutrally, the occurrence of short motifs, such as known promoters, would not be more frequent than expected by chance. In addition, I searched for similar motifs in published ITS1 sequences of other flatworms, which lack multiple repeats in ITS1. If these motifs observed in ITS1 repeat sequences also function as enhancers, multiple repeats in the ITS1 may have a regulatory function. I speculate that these abundant ITS1 variants may regulate tissue- or stage-specific transcription of ribosomal genes. A possible functional role for these ITS1 variants is supported further by the observation that sequence conservation of repeated genes within individuals (due to processes of concerted evolution), is not eliminating intra-individual variation of the ITS1 in these trematodes, which would be expected if ITS1 sequences were evolving neutrally.

6.2 Introduction

In this chapter I investigate a possible functional role for the repeats embedded in ITS1 of the *Paragonimus* and *Schistosoma* species investigated in this thesis (Chapters 4 & 5). I extend my observations to consider additional trematode genera and a cestode (data obtained from GenBank and the literature).

Specific sequence motifs are required to initiate or enhance the transcription of both protein coding and ribosomal genes (Latchmann 1995, Li 1997). These motifs occur in the promoter regions of genes, where they initiate transcription or outside the promoter region, where they may regulate transcription.

Promoter regions occur upstream of transcription start sites of genes (in the 5' flanking region). Generally they contain several sequence motifs that bind transcription factors (TFs) which, when bound, form a complex neccessary to initiate transcription of a gene. Sequence motifs (and their variants) that bind TFs include TATA boxes, which occur upstream of most eukaryotic transcription start sites. The TATAAT box controls the starting point of transcription when it is present. Another motif, the CCAAT box occurs further upstream in a wide variety of eukaryotic promoters and is essential for transcription to occur from such promoters. Multiple CCAAT box binding proteins have been described, suggesting the importance of this motif, not only in activation, but also in gene regulation (Latchman 1995). Thirdly, the Sp1 motif is a GC-rich region (with motif sequence GGGCGG, its complement CCCGCC and its reverse complement CCGCCC), the GC box, which activates transcription of protein coding genes, in any orientation, in the presence of Sp1 protein. Sp1 protein is a DNA-binding zinc finger protein which is present in all cell types and activates promoters containing it (Rhodes & Klug 1993, Latchman 1995). Both the GC-and CCAAT boxes control initial binding to the DNA of the enzyme RNA Polymerase II, which transcribes protein coding genes in eukaryotes (Li 1997). Promoter motifs have recently been described from repeats in the external transcribed spacer (ETS) of the rDNA of the cestode, *Echinococcus granulosus* (Picon *et al.* 1996).

Enhancer motifs, in comparison to promoter motifs, can activate promoters from a considerable distance, regardless of their orientation relative to the promoter, or their

position relative to the transcription unit (Latchman 1995). Enhancers have been found in both the 5' and 3' flanking regions of genes and in introns (non-coding regions within genes), however, they have not previously been described from the first or second Internal Transcribed Spacers (ITS1 and 2). These motifs may be targets for the stage- or tissue-specific regulation of transcription of genes (Li 1997).

Trematode ITS1 can contain a number of repeats near its 5' end, as described in Chapters 4 and 5 (van Herwerden, Blair & Agatsuma 1998*a*, 1998*b*, Luton, Walker & Blair 1992, Kane *et al.* 1996). Where repeats occur, they can vary in number, even within individuals. Species that have variable numbers of repeats in ITS1, from three trematode genera (*Paragonimus*, *Schistosoma* and *Dolichosaccus*), have two segments referred to as "*a*" and "*b*". If "*b*" is present, it is always flanked on both sides by "*a*". Furthermore, ITS1 variants containing multiple repeats are much more abundant in the genome than ITS1 variants lacking multiple repeats, as shown in Chapters 4 and 5 (van Herwerden, Blair & Agatsuma 1998*a* & 1998*b*).

Such intra-individual variation is not expected for repeated genes, such as the rDNA multigene family. Ribosomal DNA sequences occur as repeated clusters (in tandem) of highly conserved genes for 18S, 5.8S and 28S rRNA, separated by variable spacer sequences (ITS1 & 2) (Long & Dawid 1980). Sequence homogeneity of tandemly repeated genes, such as the rDNA is generally maintained by processes of concerted evolution such as gene conversion, unequal crossing-over and slipped-strand mispairing (Dover 1982, Hillis & Davis 1988, Hillis *et al.* 1991, Linares, Bowen & Dover 1994, Schlötterer *et al.* 1994, Tang *et al.* 1996). Despite the expectation of homogeneity, many variant forms of ITS1 have been described, within and among individuals across a range of organisms other than the trematodes and the cestode described in this study (see Chapter 4).

The degree of sequence variation in tandemly repeated genes therefore, is a balance between variation generating and homogenising processes (Schlötterer *et al.* 1994, Fritz *et al.* 1994). The rate of sequence homogenisation may be affected by factors such as sexual reproduction through the formation of bivalents and chiasmata, dispersal of rDNA loci on various chromosomes (i.e. the presence of multiple Nucleolar

Organiser Regions, NOR) (Vogler & De Salle 1994), hybridisation with sibling species and natural selection, which may favour certain variants.

Intra-individual variation in ITS1 of the Asian schistosomes and absence of such variation in the African schistosomes (van Herwerden, Blair & Agatsuma 1998*a*) may be due to a reduction of homogenisation of ITS1 in the Asian species as discussed in Chapter 4. The absence of recombination at "hot spots" in the ITS1 coupled with a genome-wide reduction of chiasma formation in *S. japonicum* may result in the reduction of rDNA sequence homogenisation, and consequently may be responsible for the increased intra-individual variation detected in the ITS1 of *S. japonicum* (van Herwerden, Blair & Agatsuma 1998*a*). Given that homogeneity of ITS1 sequences in these flatworms is not being maintained, there may be functional significance for the observed intra-individual variation. In this study I determined whether the abundance in the genome of ITS1 variants that contain multiple repeats is related to the presence in those repeats of sequence motifs for which functional significance has been demonstrated in other gene regions. I also sought to demonstrate whether, given the base composition of the repeats, such motifs occurred more frequently than might be expected by chance alone.

6.3 Methods

All known trematode ITS1 sequences were examined in this study, including *Schistosoma* and *Paragonimus* species respectively (van Herwerden, Blair & Agatsuma 1998*a*, 1998*b*)as listed in Table 6.1. Sequence data for the cestode genus *Echinococcus*, derived from the literature, was also included in this study (Table 6.1), because a promoter motif has recently been described from the ETS of this genus. ITS1 repeat sequences were used to screen the promoter database in GenBank, using Blastn (Altschul *et al.* 1990). As sequences of great similarity would have a greater number of bases in common, including short sequence motifs, than sequences of reduced similarity, I assessed the overall level of similarity in the ITS1 of the trematode genera *Dolichosaccus, Schistosoma, Paragonimus* and *Echinostoma* by sequence alignment in DNAsis (version 6, Pharmacia).

Table 6.1. Platyhelminth ITS sequences examined for the presence of enhancer motifs. The source of sequence data and genbank accession numbers for ITS1 sequences are reported where available.

Genus	species	Source	Genbank accession number
Schistosoma	*japonicum* (Philippines)	1	U82282, U89868
Schistosoma	*japonicum* (China)	1	U97531-33.
Schistosoma	*malayensis*	1	U82283, U89869-70.
Schistosoma	*mekongi*	1	U82284, U89871.
Schistosoma	*mansoni*	1	AF029309
Schistosoma	*margrebowiei*	2	Z50118
Schistosoma	*mattheei*	2	Z21718
Paragonimus	*westermani*	3	AF040925-29, AF040932-35, AF040940-43.
Paragonimus	*ohirai*	3	AF040930-31.
Paragonimus	*macrorchis*	3	AF040936.
Paragonimus	*miyazakii*	3	AF040937.
Dolichosaccus	*symmetrus*	4	L01631
Dolichosaccus	*helocirrus**	4	L01630
Echinostoma	*trivolvis*	5	ETU58097
Echinostoma	*paraensi*	5	ETU58098
Echinostoma	*caproni*	5	ETU58098
Echinostoma	*revolutum*	5	ETU58102
Echinococcus	*granulosus*	6	not available
Echinococcus	*multilocularis*	7	not available

* Described by Barton (1994). 1 van Herwerden *et al.* 1998*a*, 2 Kane *et al.* 1996, 3 van Herwerden *et al.* 1998*b*, 4 Luton, Walker & Blair 1992, 5 Morgan & Blair 1995, 6 Bowles *et al.* 1995*b*, 7 Rinder *et al.* 1997.

I then determined whether the rate of occurrence of enhancer motifs in the ITS1 of these trematodes was greater than expected by chance. This was done by estimating the probability of the random occurrence of the previously described motifs TATAAT, GGGCGG, CCCGCC, CCGCCC and CCAAT within the ITS1 sequences of each species listed in Table 6.1. These probabilities were estimated for the repeated fragments and for the unique,"post-repeat" fragments separately, by randomising each ITS1 sequence and scanning the randomised sequence for the presence of each putative promoter / enhancer motif. Each ITS1 sequence was randomised 10^5 times and scanned for the presence of each motif. Probabilities of detecting any given motif between 1 and 6 times were obtained (program written in BASIC by Julian Caley).

6.4 Results and Discussion

Overall sequence similarity of ITS1 among trematode genera was < 62%, thus the sequence conservation of particular short motifs among many of these genera is not likely due to phylogenetic conservation of primary sequence. Visual inspection of the ITS1 of the cestode *Echinococcus* spp. and the trematodes failed to find either the promoter GCCGTG$(T)_{11}$ CACGCC or the repeat element $(TG)_4CT$ described from the ETS of *Echinococcus granulosus* (Picon *et al.* 1996). In contrast, motifs that are identical to common eukaryote enhancer motifs were present in the ITS1 of many of the trematode species examined here. Therefore, it is plausible that these motifs may similarly act to bind transcription factors (TFs) and thereby regulate transcription of the ribosomal genes (Lewin 1990, Latchman 1995) (Table 6.2).

One of the sequence motifs that was identified within the ITS1 was the TATAAT motif and its reverse, which occurs within the repeats of members of the *S. japonicum* species complex as TAATAATATAAAT. This TA-rich sequence is absolutely conserved in the repeats of all species, except for a single A to T transversion in the short ITS1 variant of *S. mekongi* (which only has one copy of the repeat). In *S. malayensis* and

Table 6.2 The occurrence of promoter motifs in ITS1 of trematode species and a cestode. Values in the Table indicate the probability of chance occurrence of the sequence motifs in ITS1.

Organism	TATAAT box		GC box	CCAAT box
	repeat	[a] post-repeat	[a] post-repeat	
Paragonimus westermani	-	-	0.1	0.25
Paragonimus ohirai.	-	-	0.1	0.23
Schistosoma japonicum	0.08	-	0.03	0.2
Schistosoma malayensis,	0.05	-	0.04	0.02*
Schistosoma mekongi	0.05	-	0.04	0.19
Schistosoma mansoni	-	-	0.05	0.18
Echinostoma revolutum	na	0.06	0.1	0.03*
Echinostoma paraensi	na	0.06	0.1	0.03*
Echinostoma caproni	na	0.06	0.1	0.03*
Echinostoma trivolvis	na	0.06	0.1	0.03*
Dolichosaccus helocirrus.	0.02	-	0.1	0.33
Dolichosaccus symmetrus	na	-	0.09	0.26
Echinococcus granulosus	na	0.07	0.23	0.3
Echinococcus multilocularis	na	0.08	0.19	0.27

Unless otherwise indicated, motifs occurred once in the sequences searched. * CCAAT occurs twice in *S. malayensis* and in the echinostomes. [a] The post-repeat compartment of ITS1 sequences is the remainder of the ITS1, excluding multiple copies of repeat sequence, when multiple copies occur. na, not applicable, as repeats are absent.

S. mekongi the TAATAT motif occurs significantly more often than expected by chance (Table 6.2). In *S. japonicum* the value is not significant, but only by a small margin (Table 6.2). This is an extremely conservative estimate, because the ITS1 repeat region in *S. japonicum* (both Philippine & Chinese strains) contains an expanded TATAAT motif in the form of a microsatellite $(AT)_{13}$ AAT (both strains) and $(AT)_{20}$ AAT detected in the Chinese strain only). This motif is comparable to the promoter motif $(AT)_8$ AAT

which occurs in the gene promoter region of another eukaryote (the tomato plant, *Lycopersicon esculentum* (accession number U63117). Had I screened the randomised ITS1 repeat sequence for the expanded promoter motif $(AT)_8$ AAT, it would have been significantly more abundant than expected by chance given the marginal significance of the "truncated" TATAAT motif in both strains of *S. japonicum*. ITS1 repeats from African schistosomes (*S. mansoni, S. margrebowiei* and *S. mattheei*) could not be aligned with repeats from the ITS1 of Asian schistosomes (*S. japonicum, S. malayensis* & *S. mekongi*) and neither did they have a TATAAT motif. Additionally, the African, unlike the Asian schistosomes, do not display any intra-individual variation of ITS1 despite the presence of repeats (Kane & Rollinson 1994). This suggests that the TATAAT motif may have functional significance, which is affecting the structure and abundance of the ITS1 variants in which they occur. The hypothesis that the repeats and motifs present in them are the result of random variation is rejected for the *S. japonicum* species complex.

The other trematodes that did have repeats in the ITS1 varied with respect to the presence of the TATAAT motif. ITS1 repeats of *Paragonimus* species lacked the TATAAT motif, whereas *Dolichosaccus helocirrus* (*Dolichosaccus* sp.in Luton, Walker & Blair 1992) contained repeats, which also had the TATAAT motif, more frequently than expected by chance (Table 6.2). In *Dolichosaccus symmetrus,* however, the ITS1 lacked both repeats and the TATAAT motif. Likewise, *Echinostoma* species had no repeats in the ITS1 (Morgan & Blair 1995), however, the TATAAT motif was present, and the probability that this motif occurred by chance was approximately 6% (Table 6.2). In *Echinococcus* spp. (Cestoda) the ITS1 sequences also varied within individuals (Bowles, Blair & McManus 1995b, and Rinder *et al.* 1997), due to the occurrence of insertion-deletions (indels), rather than variable numbers of repeats. Nevertheless, both *E. granulosus* and *E. multilocularis* had a TATAAT motif which would be expected to occur with a probability of 7 and 8 % respectively. It appears that the TATAAT motif may be important in ITS1 of all trematodes examined here that contain it, even in the absence of repeats, but particularly when it occurs in the repeats (*S. japonicum* species complex and *D. helocirrus*).

A variant of another putative enhancer motif, the Sp1 or GC box (CCCGCC), occurred in the ITS1 of all schistosomes investigated here, more frequently than expected by chance (Table 6.2). This motif however occurred external to the ITS1 repeats. None of the other trematode genera or the cestodes examined had the Sp1 box (or variants of it) present more frequently than expected by chance (Table 6.2). It is already known that GC-rich sequences are functionally important for the maintenance of ITS secondary structure (Michot *et al.* 1993) and I do not make any further inferences in this study about its abundance.

Although the CCAAT motif was present in the ITS1 of all the trematode species investigated here, as well as the cestode, it never occurred within the ITS1 repeats. Furthermore, it only occurred more frequently than expected by chance in those species where it occurred twice (*S. malayensis* and all *Echinostoma* species investigated here) (Table 6.2). As this motif is not part of the repeat sequences, I do not make any inferences about its possible functional significance here.

Embedded within the ITS1 repeat sequences of members of the Asian *S. japonicum* species complex and *D. helocirrus* is the motif TATAAT. Promoter motifs have been demonstrated to increase transcription when more copies of internal repeat sequence containing such motifs occur in the ETS of *Drosophila* (Tautz *et al.* 1987, Schlötterer *et al.* 1994), *Xenopus* (Long & Dawid 1980, Moss 1983, Busby & Reeder 1993), *Echinococcus* species (Picon *et al.* 1996) and *Trypanosoma cruzi* (Dietrich *et al.* 1993). Given this, it is plausible that they may also regulate transcription when present elsewhere in the genome, specifically in the ITS1. None of the described motifs were detected in the ITS1 repeats of *Paragonimus* species, even though ITS1 variants containing multiple repeats are much more abundant in the genomes of all members of *S. japonicum, P. westermani* and *P. ohirai* species complexes than variants with few repeats (van Herwerden, Blair & Agatsuma 1998a & b). This suggests that there may be other, as yet unidentified motifs present in the *Paragonimus* repeats.

Although GC and CCAAT boxes are present in the ITS1 of some trematodes more frequently than expected by chance, they are unlikely to act as transcription regulators of the ribosomal genes, as they affect binding of RNA Polymerase II rather

than RNA Polymerase I (Pol I) and it is Pol I which is involved in the transcription of ribosomal genes (Li 1997). Furthermore, GC-rich sequences play an important role in maintaining secondary structure in the ribosomal spacers (Michot *et al.* 1993).

I postulate a functional role for the TATAAT motif in the ITS1 repeats, based on three observations made in this study. Firstly, the statistical improbability of them occurring and secondly, their demonstrated role as promoters elsewhere in the genome. Thirdly, processes of concerted evolution which normally maintain sequence homogeneity in tandemly repeated genes within individuals and species, are not eliminating intra-individual variation of the ITS1 in these trematode species. Indeed, it is possible that intra-individual variation is maintained in the ITS1 by reduced genomic recombination and cross-overs at "hot-spots" in *S. japonicum* (as discussed in Chapter 4) and possibly in *P. westermani* as well, as discussed in Chapter 5). The intra-individual variation is generally a direct consequence of the presence of the repeats. The validity of this hypothesis remains to be tested. In vivo rDNA transcription and quantitation of the transcripts from rRNA variants of schistosomes that lack multiple repeats and those that have multiple repeats embedded in the ITS1, should indicate if transcription is indeed regulated by the presence of additional repeats and by implication, additional enhancer motifs. Doing so for the three species of the *S. japonicum* species complex, and some *Paragonimus* species would determine firstly, if the TATAAT motif enhances transcription when more abundant (*Schistosoma*) and secondly, if the absence of known enhancers in the repeats of ITS1 results in a lack of transcriptional enhancement (*Paragonimus*) or if other, as yet undetected enhancers may be present.

I have described the occurrence of enhancer motifs in the ITS1 of several genera of trematode and a cestode. Only one of these motifs, the TATAAT box (or variants of it), occurred within the repeats which occur in the ITS1 of some of the trematodes. When it did occur in the repeats, it occurred more frequently than expected by chance. This is of interest, as the ITS1 variants that have multiple copies of repeats have been shown to be in greater abundance in the genome. The other enhancer motifs, a GC- and a CCAAT box, occurred in the ITS1 but not in the repeats. Additionally, these motifs are

functionally associated with transcription of protein coding, rather than ribosomal genes, as they control the binding of RNA Pol II (Li 1997).

CHAPTER 7

Intra - and inter- individual variation in ITS1 of two trematode (Digenea) genera, *Paragonimus* and *Schistosoma*: phylogenetic inferences.

7.1 Abstract

In this Chapter I investigate the utility of the ribosomal first internal transcribed spacer (ITS1) for phylogenetic studies of Asian trematodes of the genera *Paragonimus* and *Schistosoma*. Numerous clones containing ITS1 PCR products were sequenced for *P. miyazakii*, *P. macrorchis* and members of the *P. ohirai* and *P. westermani* species complexes. In addition, clones containing ITS1 variants of the Asian *Schistosoma japonicum* species complex, which includes *S. malayensis* and *S. mekongi* were also sequenced for possible use as a phylogenetic marker in this genus and for comparison, given the supposed common biogeographic history of *Schistosoma* and *Paragonimus* in East Asia. The ITS1 of both trematodes is composed of three distinct regions which have been described and characterised in Chapters 4, 5 and 6. Sequences from all three regions can be aligned among those species within the genera investigated here. The initial hypothesis, that the "post-repeat" region would be valuable for phylogenetic studies within the species complexes, was proved correct for the *P. ohirai* and *S. japonicum* species complexes, however it was proved wrong for the *P. westermani* species complex. Intra-individual sequence variation in *P. westermani* was sometimes greater than among individuals from different species complexes.

7.2 Introduction

In this Chapter I consider only the "post-repeat" region of ITS1 from both *Paragonimus* and *Schistosoma* species. Additional characterisation of the ITS1 variants in these trematodes has already been extensively discussed in Chapters 4, 5 and 6.

Studies on phylogeny and / or intraspecific variation in *Paragonimus* species have recently been done using nuclear ribosomal DNA (second internal transcribed spacer (ITS2)) and mitochondrial DNA (partial COI) sequences (Blair *et al.* 1997a, Blair, Agatsuma & Watanobe 1997). The ribosomal spacers are expected to be particularly useful for distinguishing among species. Nuclear ribosomal sequences are often assumed to be homogenised within individuals and populations of a species by concerted evolution (Dover 1982, Hillis & Davis 1988), though see Chapters 4, 5 and 6 (van Herwerden, Blair & Agatsuma 1998a, 1998b). Consequently, both inter- and intra-individual variation of these sequences should be reduced or eliminated. Homogenisation is most likely to occur if the ribosomal genes occur in a single tandem array on one chromosome, as seems to be the case with trematodes (e.g. Hirai, 1988, Hirai *et al.* 1989). The ITS2 has been shown to be a sensitive marker at the species level in trematodes (Morgan and Blair, 1995; Bowles, Blair & McManus 1995a, 1995b, Luton, Walker & Blair 1992). In the previous study on species of *Paragonimus*, Blair *et al.* (1997a) noted some geographical variation in the ITS2 among isolates of *P. westermani*, consistent with the notion that *P. westermani* is a complex of cryptic species. In contrast, little or no variation has been detected in the ITS2 of the genus *Schistosoma* (Bowles *et al.* 1993, Després *et al.* 1992). Other studies on trematodes indicate that sequences in the ITS1 might be less conserved than ITS2 (e.g. Luton, Walker & Blair 1992) and thus provide greater phylogenetic resolution among members of species complexes such as *P. westermani, P. ohirai* and *S. japonicum*. ITS2 sequences could not resolve the phylogenetic relationships within the NE Asian *P. westermani* cluster, as there was not sufficient variation. I have thus evaluated ITS1 sequences in species of both *Paragonimus* and *Schistosoma*, in an attempt to obtain better phylogenetic resolution below the species level. I therefore characterized the ITS1 for a number of Asian species and strains of both *Paragonimus*

and *Schistosoma*, with emphasis on members of these species complexes (Chapters 4, 5 & 6). The results of this phylogenetic component, and conclusions regarding the utility of the "post-repeat" region of ITS1 for phylogenetic inference, are presented here.

7.3 Materials and Methods

7.3.1 DNA extraction and related methods

Adults of Asian *Paragonimus* and *Schistosoma* species were obtained as shown in Table 3.1. DNA extraction from specimens, ITS1 amplification, cloning, sequencing and alignment were done as described in the General Methods (Chapter 3). At least 2 individual clones per ITS1 variant, per species, were identified from all sequenced clones (except for *P. macrorchis*, where only 1 clone was obtained) (Table 7.1). No differences were found in the sequences of the clone pairs, other than the occasional unique base change. Such base changes, which occurred at a rate of 1-2 bases per thousand were considered due to inherent *Taq* polymerase misincorporation, as they occured at approximately the expected frequency (Saiki *et al.* 1988). These occassional base changes occured in single clones at otherwise invariant sites and had no influence on the phylogenetic inferences from the sequence data. This indicates that the distinct differences observed are not due to PCR artefacts. An exception was the individual diploid *P. westermani* from Japan, Ohita (PwJO(2n)) where all 7 clones were unique, but fell into 3 variant types (see Fig. 5.1) Recombinant clones of complete ITS1 were sequenced completely in both directions. "Post-repeat" ITS1 fragments amplified by primers Parfor / Parrev (see Table 3.2 and section 5.2) were sequenced directly from PCR products in both directions, using an ABI dideoxy terminator sequencing kit, as per manufacturer's instructions (Perkin Elmer Cetus).

7.3.2 Phylogenetic Analysis

Genbank numbers for ITS1 sequences produced in this study are presented in Table 7.1 (*Paragonimus* species) and Table 7.2 (*Schistosoma* species). Sequence alignments for the "post-repeat" ITS1 fragments, used for the phylogenetic analysis are

presented in Fig. 7.1 (*Paragonimus* spp.) and Fig. 7.3 (*Schistosoma* spp.). Phylogenetic analysis was done as described in Chapter 3 (section 3.7). The data matrices included only the "post-repeat" regions which could be amplified by primers parfor and parrev (259 sites) for the *Paragonimus* species and by the primers AsitsF - AsitsR for the *Schistosoma* species (375 sites) (see Table 3.2). Hillis and Bull (1993) showed that the correct topology in an "experimental" phylogeny could be supported by bootstrap values of as low as 70%. I therefore collapsed branches with bootstrap support <70% in the NJ tree and only show BCL ≥ 70% in all phylogenetic trees.

7.4 Results and Discussion

In both genera the ITS1 is composed of three distinct regions as previously described in Chapter 5 (van Herwerden, Blair & Agatsuma 1998a & b). The focus of this chapter is the "post-repeat" region (259 or 375 nucleotides) for *Paragonimus* and *Schistosoma* respectively, as it lacks repeats and is 90% similar among the species within a genus. The ITS1 repeat regions are not suitable for use in phylogenetic studies and have already been described in detail in previous chapters (4, 5 and 6).

7.4.1 Variation in "post-repeat' regions and phylogenetic analysis in *Paragonimus*

An alignment of 26 "post-repeat" ITS1 sequences was used to construct phylogenetic trees. In most cases, alignment was straightforward, however, two of the three variant groups from the Japanese diploid isolate from Ohita had deletions involving different parts of the ITS1 region (Chapter 5). Clones PwJO(2n) 7e and 102 had the entire repeat region deleted, including the preceding 13 bases, whereas clones PwJO(2n) 07, 08 and 10 had 125 bases deleted from the 3'end of "post-repeat" ITS1. The PwJO(2n) 7e and 102 variants differed extensively throughout the post-repeat sequence from clones (from this individual and others) which do not contain such a deletion (Fig. 7.1). Members of the *P. westermani* group display much intra-individual variation, whereas members of the *P. ohirai* group display little or no intra-individual variation in the "post-repeat" fragment. This difference between members of the two *Paragonimus* species complexes is further supported by the presence of multiple

restriction fragment length polymorphisms (RFLPs) of the rDNA, flanking the ITS of all *P. westermani* individuals examined, but only a single band in the rDNA from the lineage containing members of the *P. ohirai* species complex as well as *P. miyazakii* (lanes 7-10, Fig. 9.1). Differences among the post-repeat sequences are displayed as a phylogenetic tree (Fig. 7.2). Neighbour-joining and Parsimony trees were constructed as described in Materials & Methods (Chapter 3). In a total of 259 characters, 123 were variable and 86 of those were informative for parsimony. Transitions outnumbered transversions 3:1.

Twelve thousand six hundred most parsimonious trees (MPT) of length 243 were obtained with a consistency index (CI) of 0.720, retention index (RI) of 0.766 and a rescaled consistency index (RC) of 0.552 for the strict consensus tree. Both strict and 70% majority rule (not shown) consensus trees (Fig. 7.2) had topologies that were in agreement with the NJ tree topology (not shown). The only difference between the 70% majority rule tree and the NJ and strict consensus trees, was the placement of *P. miyazakii*. In the 70% majority rule tree *P. miyazakii* is placed within the *P.ohirai* - *P. macrorchis* clade (asterisked in Fig. 7.2). Strong hierarchical structure was present in the phylogenetic signal obtained, as $g_1 = -1.116$ (see Table 2.2).

7.4.2 Variation in "post-repeat" regions and phylogenetic analysis in *Schistosoma*

Sequences of "post-repeat" ITS1 amplified PCR products were aligned for 12 clones from the *S. japonicum* species complex (Fig. 7.3) and used to construct phylogenetic trees. Sequences of the ITS1 of the African schistosomes, *S. mansoni*, *S.mattheei*, *S. haematobium* and *S. intercalatum* (taken from GenBank, see Table 7.2) were included as additional taxa. Alignment was straightforward. Differences among the "post-repeat" sequences are displayed as a phylogenetic tree (Fig. 7.4). Parsimony and neighbour-joining trees were constructed as described in Materials & Methods (Chapter 3). In a total of 375 characters, 110 were variable and 90 of those were informative for parsimony, which produced 324 most parsimonious trees of length 102. The consistency index (CI) was 0.951, the retention index (RI) was 0.981 and the rescaled consistency index (RC) was 0.933 for the strict consensus tree. Within the *S. japonicum* species complex, there was strong support for the monophyly of the 2

strains of *S. japonicum* and the other two species of the complex (BCL = 77). In addition, hierarchical structure was obtained from this data, as g_1 = -0.898 (see Table 2.2).

Table 7.1: *Paragonimus* species investigated here, indicating geographic origin, number of clones sequenced (applies only to complete ITS1 sequences from cloned material) and Genbank numbers for variants. Direct sequences were obtained only for the 3' "post-repeat" region of ITS1. Each line represents data from a single individual (see Table 3.1).

Species	Id code (ploidy)	Geographic location	#clones sequenced or direct sequence	Genbank number
P. westermani	PwJO(2n)	Japan, Ohita	7	AF040925 - 040927
species	PwJH (2n)	Japan, Hyogo	direct *	AF040941
complex	PwJM(2n)	Japan, Mie	direct *	AF040942
	PwP (2n)	Philippines, Sorsogon	direct *	AF040943
	PwM (2n)	Malaysia, U.L.	6	AF040934 - 040935
	PwCM(2n)	E China, Minchin	direct *	AF040940
	PwC(?n)	China, ?	6	AF040928 - 040929
	PwJA(3n)	Japan, Amakusa	6	AF040933
	PwCL(3n)	NE China, Liaoning	6	AF040932
P. ohirai	PoT	Japan, Tanegashima	3	AF040930
P. ohirai	PoK	Japan, Kinosaki	3	AF040931
P. iloktsuenensis	279	Japan, Amami	direct *	AF040939
P. ikoktsuenensis	K	Japan, Amami	direct *	AF040938
P. macrorchis	Pmac	Thailand	1	AF040936
P. miyazakii	Pmi	Japan, Miyazaki	direct *	AF040937

Key: ? When relevant information was not available, K.P. = Kualah Pilah, U.L. = Ulu Langat, . 3' Fragments were amplified by primers Parfor & Parrev & sequenced directly from PCR products, as described in Materials & Methods.*

Table 7.2: Asian *Schistosoma* species investigated here, indicating geographic origin of isolate, source, number of clones sequenced (applies only to complete ITS1 sequences from cloned material) and Genbank numbers for variants. Each line of the *S. japonicum* complex members represents data from the DNA of 5 pooled individuals.

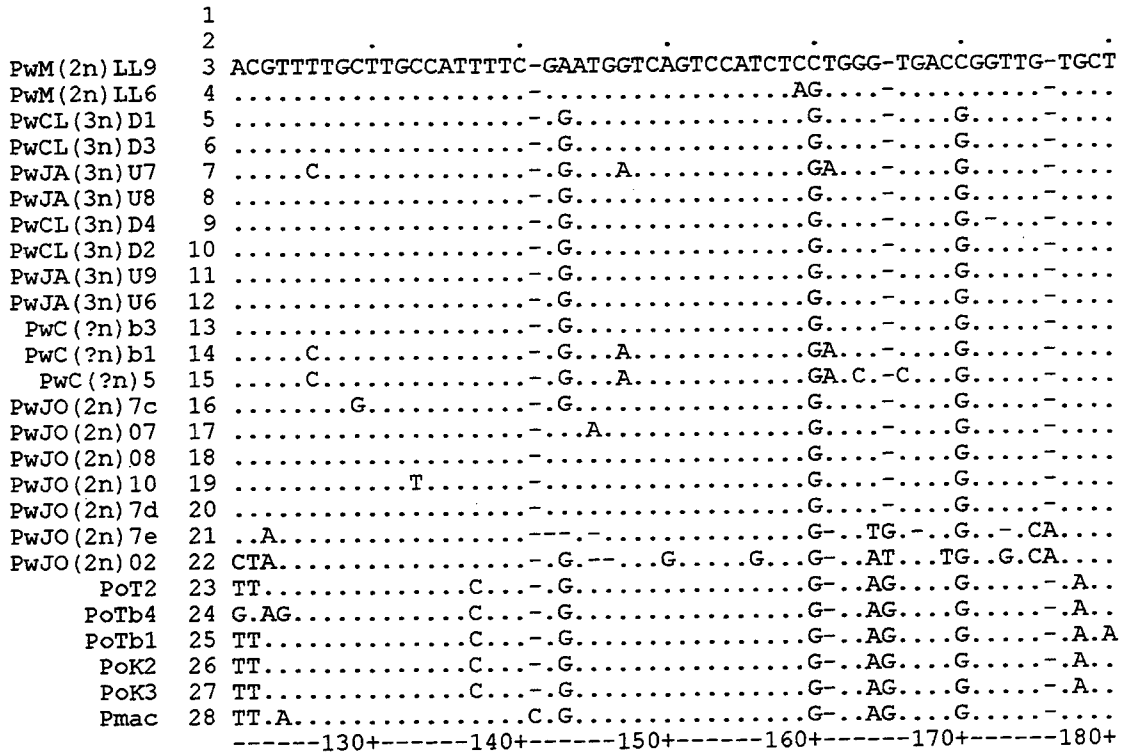
	Species	**id code**	**Geographic location**	**Source**	**#clones sequenced**	**Genbank number**
Schistosoma	*S. japonicum* P	SjP	Philippines, Sorsogon	D. McManus	3	U82282
japonicum	*S. japonicum* C	SjC	China, Anhui	D. McManus	3	U97531
species	*S. mekongi*	Smek	Thailand, Laos	Upatham	3	U82284
complex	*S. malayensis*	Smal	Malaysia, Baling	K. Lai	3	U82283
African	*S. mansoni*	-	Caribbean, Puerto Rico	P. Brindley	2	AF029309
schistosomes	*S. mattheei*	-	Africa,	Genbank	a	Z21718
	S. haematobium	-	Africa, Kenya	Genbank	a	Z21716
	S. intercalatum	-	Africa	Genbank	a	Z21717

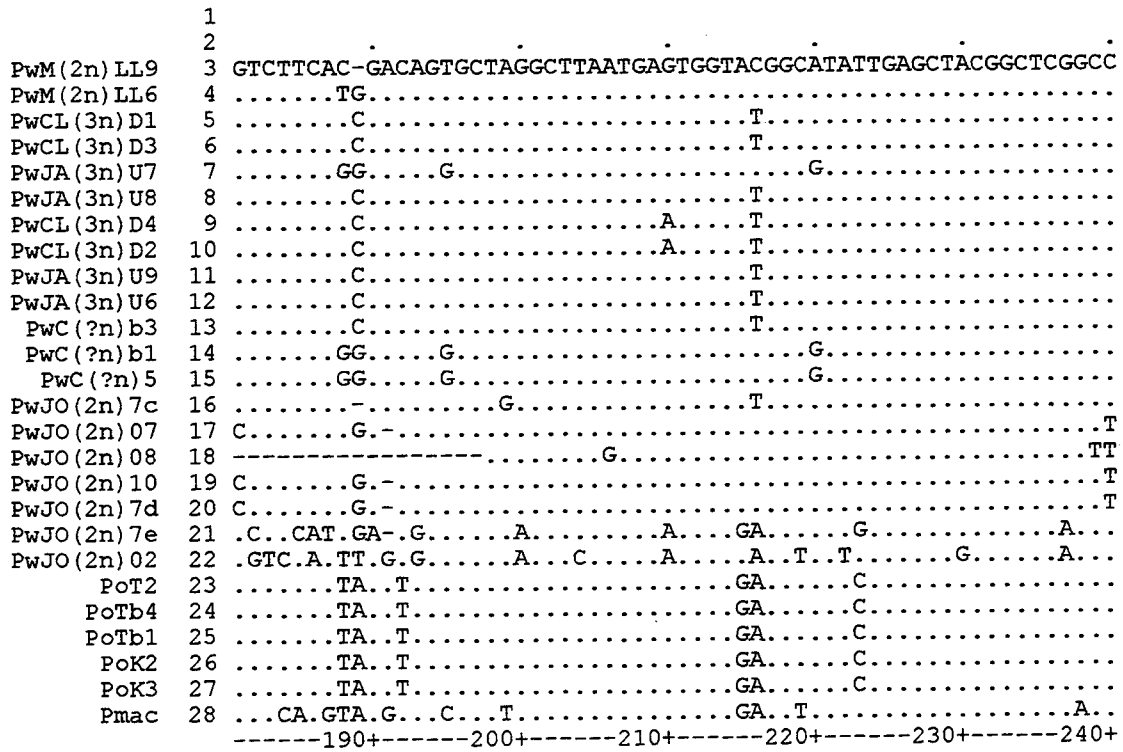
a From Kane & Rollinson 1994.

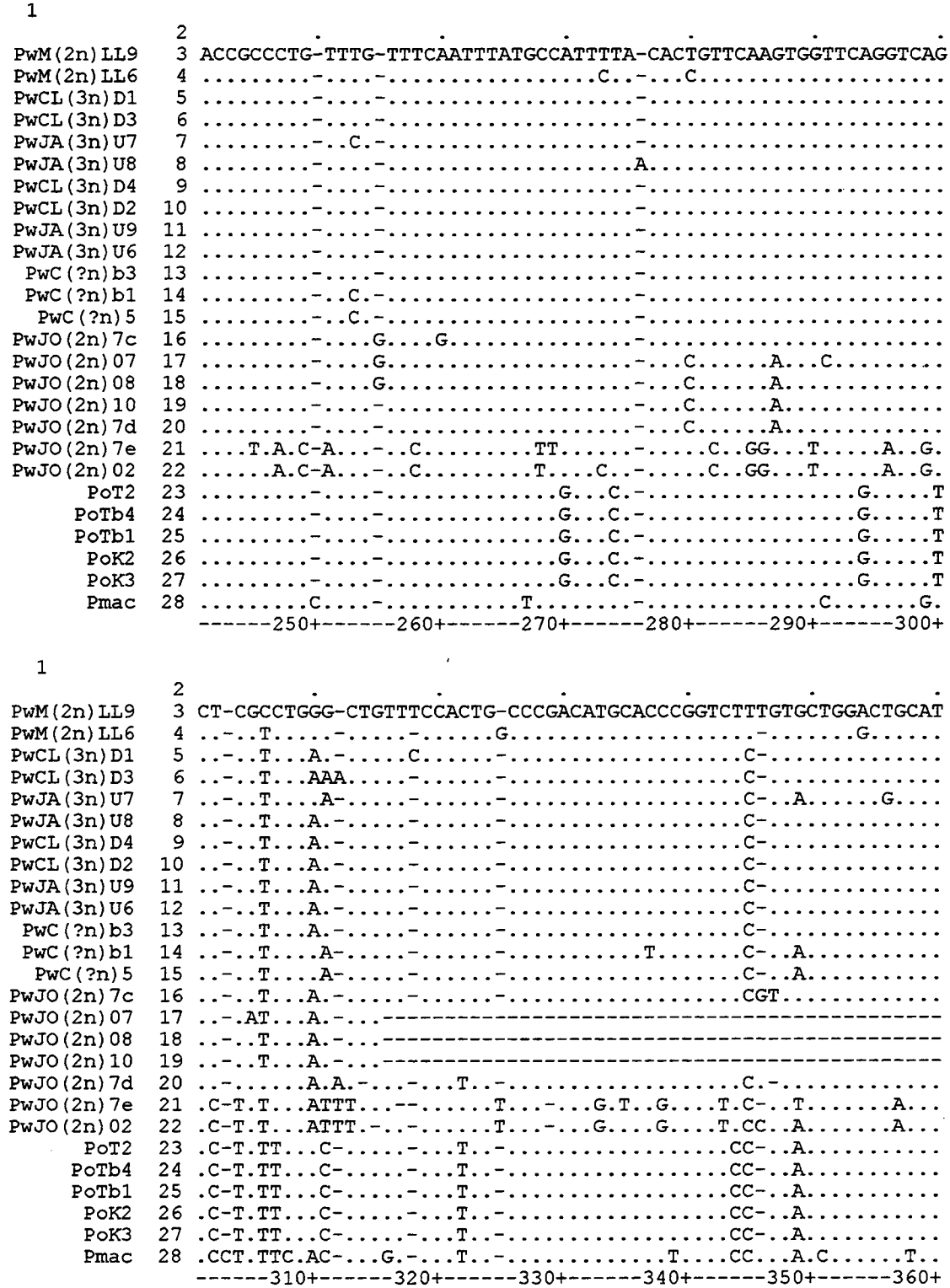
Fig. 7.1 Alignment of ITS1 sequences from *Paragonimus* species, showing only the "pre- and post-repeat" sequences. Repeat sequences from ITS1 are presented in Chapter 5. Identification codes are as listed in Table 7.1. The * indicates where the ITS1 repeat sequences occur.

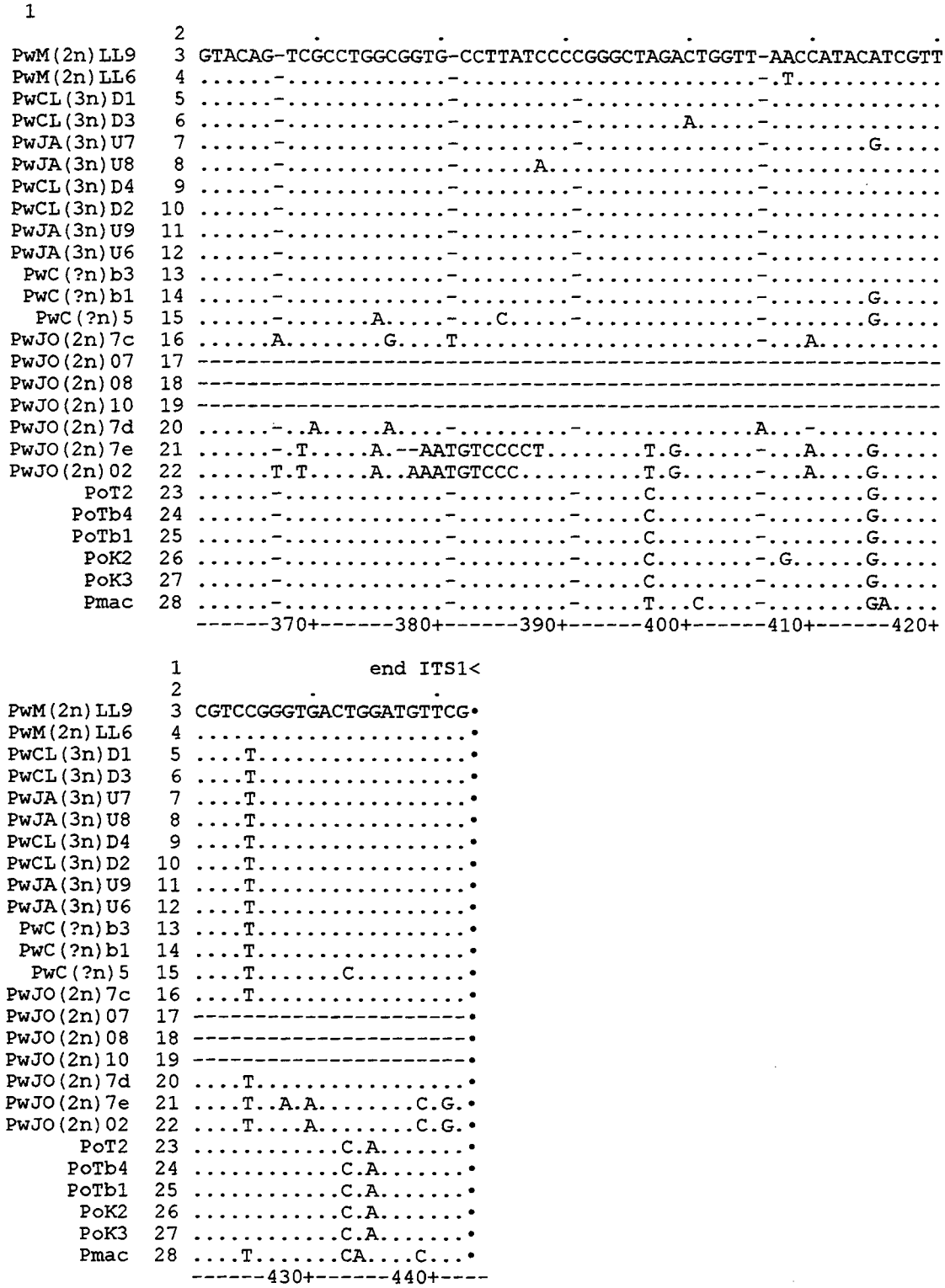
```
                >Pre-repeat ITS1
              2         .         .         .         .         .
PwM(2n)LL9    3 ACAGAATTCCCAAATCTGAAAGTGCCTCAATGAGCACAAAGGCGTGT--TCTAAACGGG-
PwM(2n)LL6    4 ..........................................................-
PwCL(3n)D1    5 ............T.......C.....................--........A.-
PwCL(3n)D3    6 ............T.......T.....................--..........-
PwJA(3n)U7    7 ..........................C.T......C.....--...G......-
PwJA(3n)U8    8 ............T.......C.....................--........A.-
PwCL(3n)D4    9 ............T.......C.....................--........A.-
PwCL(3n)D2   10 ............T.......C.....................--........A.-
PwJA(3n)U9   11 ............T.......C.....................--........A.-
PwJA(3n)U6   12 ............T.......C.....................--........A.-
 PwC(?n)b3   13 ..........................................--........A.-
 PwC(?n)b1   14 ..........................C.T......C.....--...G......-
  PwC(?n)5   15 ............T.............C.T......C.....--...G......-
PwJO(2n)7c   16 ............T.......T.....................--..........-
PwJO(2n)07   17 ............G.............C.............A.....--..C..T....-
PwJO(2n)08   18 ............G.........................A.....--..C.......-
PwJO(2n)10   19 ............G.........................A.....--..C.......-
PwJO(2n)7d   20 ......A.....G.........................A.....--..C.......-
PwJO(2n)7e   21 ..........T....T........T........A..A.TG....A..GTCT.C.....-
PwJO(2n)02   22 ..........T....T.....A..T.C......A..A.TG...AA..GTCT.C.....-
      PoT2   23 ...........T..AT.A.........AG...G...A.TG...AA..GT.T.CT.A.C.A
     PoTb4   24 ...........T..AT.A.........AG...G...A.TG...AA..GT.T.CT.A.C.A
     PoTb1   25 ...........T..AT.A.........AG...G...A.TG...AA..GT.T.CT.A.C.A
      PoK2   26 ...........T..AT.A.........AG...G...A.TG...AA..GT.T.CT.A.C.A
      PoK3   27 ...........T..AT.A.........AG...G...A.TG...AA..GT.T.CT.A.C.A
      Pmac   28 .............AT.C...A.............T.TG...AA..AC.T.CC...-.-
                -------10+-------20+-------30+-------40+-------50+-------60+

              1                     <*>Post-repeat ITS1
              2         .         .         .         .         .         .
PwM(2n)LL9    3 -CAGTCATGCTCGTCAACTG*CTATAATTG-CAGCACGGTGCCTACCTGTCTGATGCCCT
PwM(2n)LL6    4 -................A..*T....C...-..........A...........C....C
PwCL(3n)D1    5 -................AC.*T....C...-.....G.....................T..
PwCL(3n)D3    6 -................A..*T....C...-.....G.....................T..
PwJA(3n)U7    7 -............T..A..*T.........-.....G.....................
PwJA(3n)U8    8 -................AC.*T....C...-.....G.....................T..
PwCL(3n)D4    9 -................AC.*T....C...-.....G.....................T..
PwCL(3n)D2   10 -................AC.*T....C...-.....G.....................T..
PwJA(3n)U9   11 -................AC.*T....C...-.....G.....................T..
PwJA(3n)U6   12 -................AC.*T....C...-....GG.....................T..
 PwC(?n)b3   13 -................A..*T....C...-.....G.....................T..
 PwC(?n)b1   14 -............T..A..*T.........-.....G.....................
  PwC(?n)5   15 -............T..A..*T.........-.....G.....................
PwJO(2n)7c   16 -................A..*T....C...-.....G.....................T..
PwJO(2n)07   17 -.....G..G.......A..*T....C...-.....G.....................T..
PwJO(2n)08   18 -.....G..G.......A..*T....C...-.....G.....................T..
PwJO(2n)10   19 -.....G..G.......A..*T....C...-.....G.....................T..
PwJO(2n)7d   20 -.....G..G.......A..*T....C...-.....G.....................T..
PwJO(2n)7e   21 -..A..G-------------*T...GC..A-.....GA....T...TC------------
PwJO(2n)02   22 -..A..G-------------*T...GCG.A-.....GA....T....CACA.AGC.TT..
      PoT2   23 GT......A....CT..A..*TC..GC..T-.....G........T.C.........T..
     PoTb4   24 GT......A....CT..A..*TCGG.TC.T-.....G.......-T.C.GT.TG.TGGA.
     PoTb1   25 GT......A....CT..A..*TCGG.TC.T-.....G........T.C.........T..
      PoK2   26 GT......A....CT..A..*TC..GC..T-.....G........T.C.........T..
      PoK3   27 GT......A....CT..A..*TC..GC..T-.....G........T.C.........T..
      Pmac   28 GT......A.....T..A..*T...G-..TG.....G.......G..C.........T..
                -------70+-------80+-------90+------100+------110+------120+
```


PwM(2n)LL9 3 ACGTTTTGCTTGCCATTTTC-GAATGGTCAGTCCATCTCCTGGG-TGACCGGTTG-TGCT
PwM(2n)LL6 4 .....-.....AG....-.....-....
PwCL(3n)D1 5 .....-.G.....G....-....G.....-....
PwCL(3n)D3 6 .....-.G.....G....-....G.....-....
PwJA(3n)U7 7 .....C.....-.G...A.....GA...-....G.....-....
PwJA(3n)U8 8 .....-.G.....G....-....G.....-....
PwCL(3n)D4 9 .....-.G.....G....-....G.-...-....
PwCL(3n)D2 10 .....-.G.....G....-....G.....-....
PwJA(3n)U9 11 .....-.G.....G....-....G.....-....
PwJA(3n)U6 12 .....-.G.....G....-....G.....-....
PwC(?n)b3 13 .....-.G.....G....-....G.....-....
PwC(?n)b1 14 .....C.....-.G...A.....GA...-....G.....-....
PwC(?n)5 15 .....C.....-.G...A.....GA.C.-C...G.....-....
PwJO(2n)7c 16 .....G.....-.G.....G....-....G.....-....
PwJO(2n)07 17 .....-...A.....G....-....G.....-....
PwJO(2n)08 18 .....-.....G....-....G.....-....
PwJO(2n)10 19 .....T.....-.....G....-....G.....-....
PwJO(2n)7d 20 .....-.....G....-....G.....-....
PwJO(2n)7e 21 ..A.....---.-.....G-..TG.-..G..-.CA....
PwJO(2n)02 22 CTA.....-.G.--...G.....G...G-..AT...TG..G.CA....
PoT2 23 TT.....C...-.G.....G-..AG....G.....-.A..
PoTb4 24 G.AG.....C...-.G.....G-..AG....G.....-.A..
PoTb1 25 TT.....C...-.G.....G-..AG....G.....-.A.A
PoK2 26 TT.....C...-.G.....G-..AG....G.....-.A..
PoK3 27 TT.....C...-.G.....G-..AG....G.....-.A..
Pmac 28 TT.A.....C.G.....G-..AG....G.....-....
------130+------140+------150+------160+------170+------180+


PwM(2n)LL9 3 GTCTTCAC-GACAGTGCTAGGCTTAATGAGTGGTACGGCATATTGAGCTACGGCTCGGCC
PwM(2n)LL6 4 .....TG.....
PwCL(3n)D1 5 .....C.....T.....
PwCL(3n)D3 6 .....C.....T.....
PwJA(3n)U7 7 .....GG.....G.....G.....
PwJA(3n)U8 8 .....C.....T.....
PwCL(3n)D4 9 .....C.....A.....T.....
PwCL(3n)D2 10 .....C.....A.....T.....
PwJA(3n)U9 11 .....C.....T.....
PwJA(3n)U6 12 .....C.....T.....
PwC(?n)b3 13 .....C.....T.....
PwC(?n)b1 14 .....GG.....G.....G.....
PwC(?n)5 15 .....GG.....G.....G.....
PwJO(2n)7c 16 .....-.....G.....T.....
PwJO(2n)07 17 C.....G.-.....T
PwJO(2n)08 18 -----.....G.....TT
PwJO(2n)10 19 C.....G.-.....T
PwJO(2n)7d 20 C.....G.-.....T
PwJO(2n)7e 21 .C..CAT.GA-.G.....A.....A....GA.....G.....A...
PwJO(2n)02 22 .GTC.A.TT.G.G.....A...C.....A.....A..T..T.....G.....A...
PoT2 23 .....TA..T.....GA.....C.....
PoTb4 24 .....TA..T.....GA.....C.....
PoTb1 25 .....TA..T.....GA.....C.....
PoK2 26 .....TA..T.....GA.....C.....
PoK3 27 .....TA..T.....GA.....C.....
Pmac 28 ...CA.GTA.G...C...T.....GA..T.....A..
------190+------200+------210+------220+------230+------240+


PwM(2n)LL9 ACCGCCCTG-TTTG-TTTCAATTTATGCCATTTTA-CACTGTTCAAGTGGTTCAGGTCAG
PwM(2n)LL6
PwCL(3n)D1
PwCL(3n)D3
PwJA(3n)U7
PwJA(3n)U8
PwCL(3n)D4
PwCL(3n)D2
PwJA(3n)U9
PwJA(3n)U6
PwC(?n)b3
PwC(?n)b1
PwC(?n)5
PwJO(2n)7c
PwJO(2n)07
PwJO(2n)08
PwJO(2n)10
PwJO(2n)7d
PwJO(2n)7e
PwJO(2n)02
PoT2
PoTb4
PoTb1
PoK2
PoK3
Pmac
PwM(2n)LL9 CT-CGCCTGGG-CTGTTTCCACTG-CCCGACATGCACCCGGTCTTTGTGCTGGACTGCAT
PwM(2n)LL6
PwCL(3n)D1
PwCL(3n)D3
PwJA(3n)U7
PwJA(3n)U8
PwCL(3n)D4
PwCL(3n)D2
PwJA(3n)U9
PwJA(3n)U6
PwC(?n)b3
PwC(?n)b1
PwC(?n)5
PwJO(2n)7c
PwJO(2n)07
PwJO(2n)08
PwJO(2n)10
PwJO(2n)7d
PwJO(2n)7e
PwJO(2n)02
PoT2
PoTb4
PoTb1
PoK2
PoK3
Pmac


PwM(2n)LL9
PwM(2n)LL6
PwCL(3n)D1
PwCL(3n)D3
PwJA(3n)U7
PwJA(3n)U8
PwCL(3n)D4
PwCL(3n)D2
PwJA(3n)U9
PwJA(3n)U6
PwC(?n)b3
PwC(?n)b1
PwC(?n)5
PwJO(2n)7c
PwJO(2n)07
PwJO(2n)08
PwJO(2n)10
PwJO(2n)7d
PwJO(2n)7e
PwJO(2n)02
PoT2
PoTb4
PoTb1
PoK2
PoK3
Pmac
GTACAG-TCGCCTGGCGGTG-CCTTATCCCCGGGCTAGACTGGTT-AACCATACATCGTT
------370+------380+------390+------400+------410+------420+
end ITS1<
CGTCCGGGTGACTGGATGTTCG•
------430+------440+----

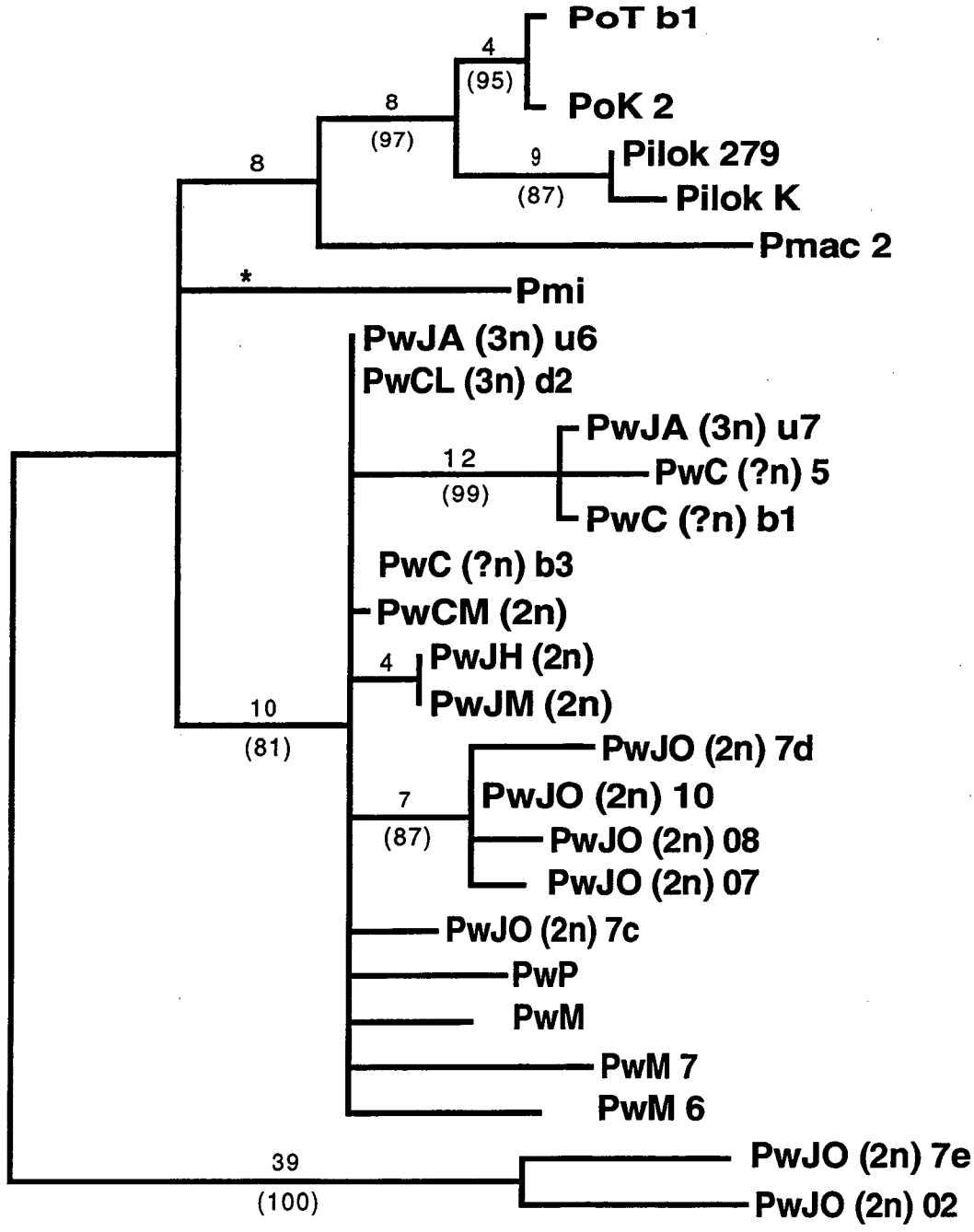


Figure 7.2 Strict consensus tree from 12600 MPTs (found by PAUP) of 259 bp "post-repeat" ITS1 fragment for all 26 unique *Paragonimus* clones from 16 individuals. Bootstrap values (only if >70%) derived from the NJ tree (1000 replicates) shown in parentheses. Isolates are identified by codes as described in Table 7.1. Individual clones are identified by the number following the species code and ploidy (Table 7.1). The absence of a clone number indicates that sequences were obtained by direct sequencing. Assigned branch lengths (from PAUP) are shown above the lines. * Indicates where there is a difference between the strict and 70% majority rule consensus trees.

Fig. 7.3 Alignment of ITS1 sequences from *Schistosoma* species, showing only the "post-repeat" region. Repeat sequences from ITS1 are presented in Chapter 4. Identification codes are as listed in Table 7.1

```
             1 >post-repeat ITS1
      sjPp1  2 ATTGATTATTATGAATTCTATTACCTGACC-GGGGTATCTGGCCTATC---TTATGCTCT
      SjPp5  3 ..............................-..................---.........
      SjPs2  4 ..............................-..................---.........
      SjCsS  5 ..............................-..................---.........
      SjCsL  6 ..............................-..................---.........
      SjCsL  7 ..............................-..................---.........
     Smek1.4 8 G......G...CA....T............C..........T...G..--C.........
     Smek1.1 9 .......A.....................................--..........
     Smek1.6 10 G..........CA....T............C..........T...G..--C.........
     Smalsml 11 GA.T.....C.CA....T.......C....C..........T...G..---.........
     Smalmed 12 GA.T.....C.CA....T.......C....C..........T...G..---.........
     Smallge 13 GA.T.....C.CA....T.......C....C..........T...G..---.........
   Smansoni  14 -------------------....TT..TGA.C......C..A....G.---CG.....C..
Shaematob.   15 -------------------....TT..TGA.C......C..A....G.---CG.....C..
   Sinterc.  16 -------------------....TT..TGA.C......C..A....G.---CG.....C..
Smattheeii   17 -------------------....TT..TGA.C......C..AT..C..GATCG.....C..
                -------10+-------20+-------30+-------40+-------50+-------60+

   1
      sjPp1  2 GATGGTGTTCTCGTG-ACTTTCGGGTTGCCTGATCTGCCAAGGGTGATGGGATAGTACGT
      SjPp5  3 ..............-.............................................
      SjPs2  4 ..............-.............................................A.
      SjCsS  5 ..............-.............................................R.
      SjCsL  6 ..............-.............................................A.
      SjCsL  7 ..............-.............................................
     Smek1.4 8 ..............-.............................................A.
     Smek1.1 9 ..............-.............................................
     Smek1.6 10 ..............-.............................................A.
     Smalsml 11 ..............C.A...........................................A.
     Smalmed 12 ..............C.A...........................................A.
     Smallge 13 ..............C.A...........................................A.
   Smansoni  14 ..............-.........................C.......C...G.A.
Shaematob.   15 .............-A.........................C.......C...G.A.
   Sinterc.  16 .............-A................W........C.......C...G.A.
Smattheeii   17 .............-A.........................C.......C...G.A.
                -------70+-------80+-------90+------100+------110+------120+

N  1
      sjPp1  2 AGCATAATTGCTATGTGCAAGGTTCAAA-GAGGGTTGCATGCGATGTTATGCATGTTACC
      SjPp5  3 ..........................-.............................
      SjPs2  4 ..........................-.............................
      SjCsS  5 ..........................-..................-........-...
      SjCsL  6 ..........................-..................-........-...
      SjCsL  7 ..........................-..................-........-...
     Smek1.4 8 GA.G.G....T.G.............A........T...TT.C..-G......C-...
     Smek1.1 9 ..................................................-........-...
     Smek1.6 10 GA.G.G....T.G.............A........T...TT.C..-G......C-...
     Smalsml 11 GA.G.G....T.G.............-........T...TT.C..-G......C-...
     Smalmed 12 GA.G.G....T.G.............-........T...TT.C..-G......C-...
     Smallge 13 GA.G.G....T.G.............-........T...TT.C..-G......C-...
   Smansoni  14 GA..AT....T.G.....T.........-...AA...TT...T.---.........CA.GT
Shaematob.   15 GA.GCT....T.G.....T.........-...AA...T....T.---.........CA.AT
   Sinterc.  16 GA.GCT....T.G.....T.........-...AA...T....T.---.........CA.AT
Smattheeii   17 GA.GCT....T.G.....T.........-...AA...T....T.---.........CA.AT
                ------130+------140+------150+------160+------170+------180+
```

```
                1
     sjPp1      2 CC-GCCCCGTTTATGTTCC-TTTCCAAACATTTTA-CACTGTTGAAGCGATCCGGTTTGG
     SjPp5      3 ..-..............T-............-.....................
     SjPs2      4 ..-..............-............-.....................
     SjCsS      5 ..-.............Y-............-.....................
     SjCsL      6 ..-.............T-............-.....................
     SjCsL      7 ..-..............-............-.....................
   Smek1.4      8 ..C.............C............A.....................
   Smek1.1      9 ..............................C.....................
   Smek1.6     10 ..C.............C...........C.A.....................
   Smalsml     11 ..C.............-............-.....................
   Smalmed     12 ..C.............-............-.....................
   Smallge     13 ..C.............-............-.....................
  Smansoni     14 ..-....T...AT.....TA-.TTC..ACC...TA.................A....
Shaematob.     15 ..-........AT.....TAC.TTC..AC-...TA.....................
  Sinterc.     16 ..-........AT.....TA-.TTC..AC-...TA.....................
Smattheeii     17 ..-........AT.....TA-.TTC..AC-...TA.....................
                  ------190+------200+------210+------220+------230+------240+

                1
     sjPp1      2 CTTGCCAATCACGGGTTTGCTGCCTGGCATGCACCTGGCCTTGTGCTGGACTGCATTTTA
     SjPp5      3 ..............................................................G...
     SjPs2      4 ..............................................................G...
     SjCsS      5 ..............................................................G.-.
     SjCsL      6 ..............................................................G.-.
     SjCsL      7 ..............................................................G.-.
   Smek1.4      8 ..............................................................G.-.
   Smek1.1      9 ................................................................-.
   Smek1.6     10 ..............................................................G.-.
   Smalsml     11 ..............................................................G.-.
   Smalmed     12 ..............................................................G.-.
   Smallge     13 ..............................................................G.-.
  Smansoni     14 .......T.....................................T.C.............G.-.
Shaematob.     15 .......T.....................................T.C.............G.-.
  Sinterc.     16 .......T..........N..........................T.C.............G.-.
Smattheeii     17 .......T.....................................T.C.............G.-.
                  ------250+------260+------270+------280+------290+------300+

  1
     sjPp1      2 CCCTGGCTTAACGGTAAATATCCTAGGC-TGCAGCGATAACCATTAGTTCTATGCACTTT
     SjPp5      3 .G........G.................-..................................
     SjPs2      4 .G........G.................-..................................
     SjCsS      5 .G........G.................-..................................
     SjCsL      6 .G........G.................-..................................
     SjCsL      7 .G........G.................-..................................
   Smek1.4      8 .G........G.................C..................................AC
   Smek1.1      9 G.G.............................................................
   Smek1.6     10 .G........G.................C..................................AC
   Smalsml     11 .G........G.................-..................................AC
   Smalmed     12 .G........G.................-..................................AC
   Smallge     13 .G........R.................C..................................AC
  Smansoni     14 .G........G.............GA..-........T.....................T..G
Shaematob.     15 .G........G.................-........T.....................T..G
  Sinterc.     16 .G........G...............N.-........T.....................T..G
Smattheeii     17 .G........G.................-........T.....................T..G
                  ------310+------320+------330+------340+------350+------360+
```

```
1                                        end ITS1•
    sjPp1    2 G---TGTT--ATA-GAG-TTATTGGCGTTT•
    SjPp5    3 .---....--...-...-...........•
    SjPs2    4 .---....--...-...-...........•
    SjCsS    5 .---...---...-...-.........A-•
    SjCsL    6 .---...---...-...-.........A-•
    SjCsL    7 .---...---...-...-.........A-•
  Smek1.4    8 ACTG...---...-.G.-.......T..A-•
  Smek1.1    9 .......---...-...-...........-•
  Smek1.6   10 ACTG...---...-.G.-.......T..A-•
  Smalsml   11 ACTG...---...-.G.-.A.....T..A-•
  Smalmed   12 ACTG...---...-.G.-.A.....T..A-•
  Smallge   13 ACTG...---...-.G.-.A.....T..A-•
 Smansoni   14 .GTAACCGATG..T.G.A.........A-•
Shaematob.  15 .GAAACCAATG..T.G.A.........A-•
 Sinterc.   16 .GAAACCAATG..T.G.A.........A-•
Smattheeii  17 .GAAACCATTG..T.G.A.........A-•
           ------370+------380+------390+
```

However, variation within the *S. japonicum* lineage (BCL= 100) prevented me from distinguishing the variants within each strain from each other (both 70% majority rule and strict consensus trees). The *S. mekongi* and *S. malayensis* species formed a well-supported cluster (BCL = 99), in which the species were distinguished by both 70% majority rule and strict consensus trees. There was some intra-individual variation in ITS1 "post-repeat" sequences from *S. mekongi*, but it did not disrupt the phylogenetic signal (clone Smek1.4). The African schistosomes formed a monophyletic clade (BCL = 100) in which *S. mansoni* is most different from the other three species (BCL = 77), as expected from previous studies (McManus & Hope 1993, Kane & Rollinson 1994, Bowles, Blair & McManus 1995, Blair *et al.* 1997a).

7.5 Conclusions

I have demonstrated that considerable intra-individual differences occur in the ITS1 of members of the *Paragonimus westermani* species complex. These differences are due both to varying numbers of repeats and to sequence differences. Even within the "post-repeat" region, 3' of the repeats, there is sufficient intra-individual sequence variation to confound phylogenetic studies within the *P. westermani* complex, but not within the *P. ohirai* species complex. Such variation may be due to the presence of multiple NORs in *P. westermani*, as compared to the presence of a single NOR in *P. ohirai* (Hirai 1988). However, this remains to be demonstrated. Other factors, such as hybridisation between sibling *P. westermani* strains or a reduced frequency of chiasmata (FX) may also account for the intra-individual sequence variation detected, as discussed

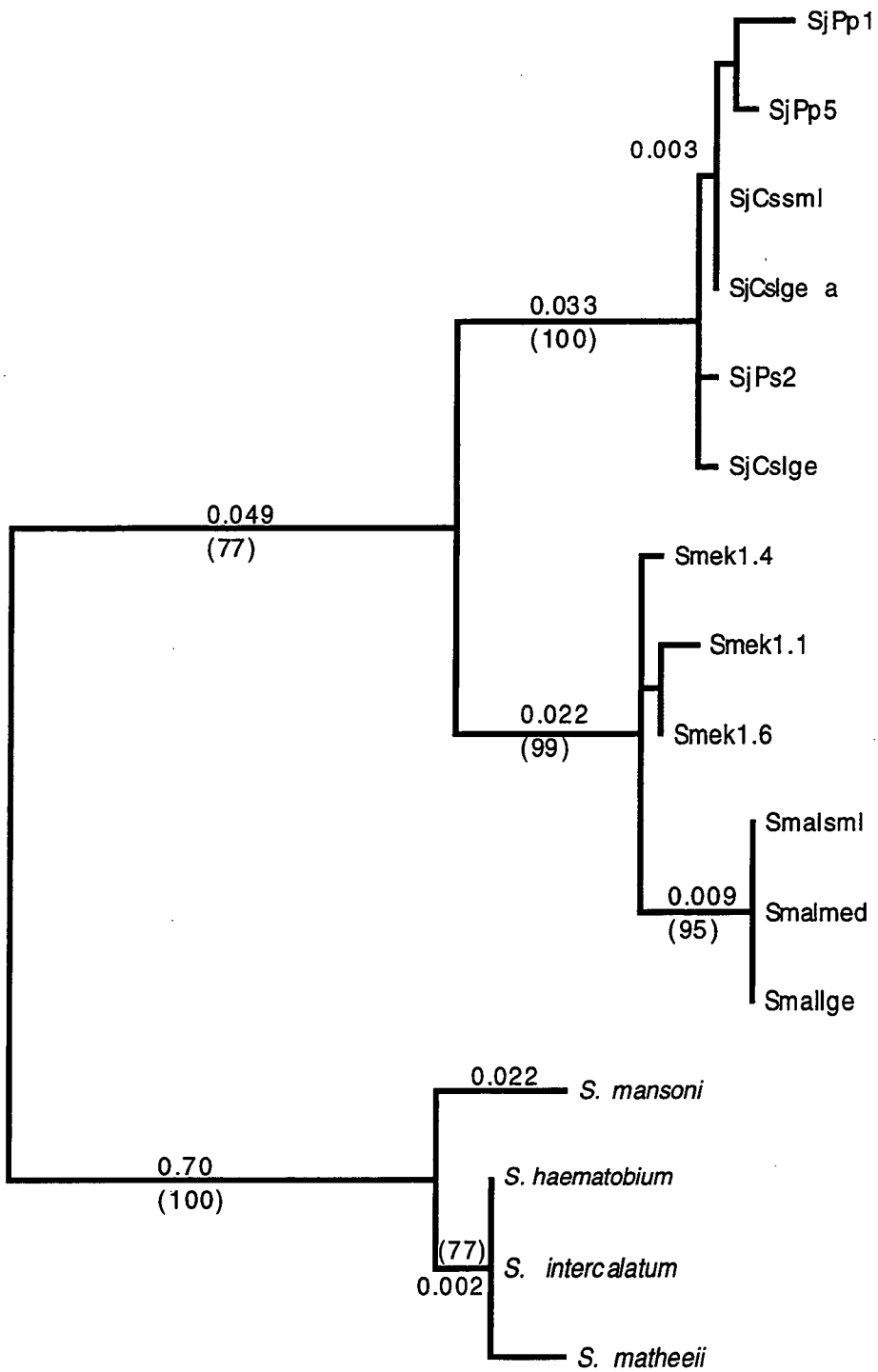


Figure 7.4 70% Majority rule consensus tree of 324 MPTs (found by PAUP 3.11) of 375 bp "post-repeat" ITS1 fragment from *Schistosoma* spp.. Assigned branch lengths and bootstrap support for the nodes (in parentheses) are shown. Bootstrap Confidence Limit (BCL) values > 70% only are indicated. Identity codes are as listed in Table 7.2. In the clone identity codes "p" is for pooled worms, "s" is for single worms, lge indicates sequence obtained from the "large", med indicates sequence obtained from the "intermediate" and sml indicates sequence obtained from the "small" ITS1 variants sequenced respectively. Numbers at the end of the id codes indicate clone identity numbers.

in Chapters 4 and 5. The situation for the *S. japonicum* species complex is intermediate. *S. japonicum* could be distinguished from *S. mekongi* and *S. malayensis* and the latter two species could be distinguished from each other by ITS1 "post-repeat" sequences, but intra-individual variation was also evident (for example clone Smek1.4) (Fig. 7.4), though not nearly as extreme as in individuals of the *P. westermani* group. As discussed in Chapters 4 and 6, ITS1 of members of the *S. japonicum* species complex appear to undergo reduced sequence homogenisation compared to the African schistosomes, which lack intra-individual variation. Such differences were ascribed to the absence of "hot spots" for recombination within the ITS1 of Asian schistosomes, which are present in the African schistosomes (Kane *et al.* 1996) and also the genome-wide reduction in the rate of chiasma formation (Hirai *et al.* 1996).

In addition we know from the work presented in Chapters 4 and 5, that the clones which were sequenced may represent only a small minority of the ITS1 variants present in the genomes of both *Schistosoma* and *Paragonimus* (van Herwerden, Blair & Agatsuma 1998a & b). I therefore conclude that ITS1 sequences may be used with caution to infer phylogenetic relationships among members of the *P. ohirai* or *S. japonicum* species complexes, but should not be used for this purpose in the *P. westermani* species group. In contrast ITS2 and COI sequence data produced unambiguous, congruent phylogenetic signals for the *P. westermani* and *S. japonicum* species groups, but failed to resolve phylogenetic relationships between members of the *P. ohirai* species complex (Blair, Agatsuma & Watanobe 1997, Blair *et al.* 1997a). The intra-individual variation of ITS1 sequences investigated here indicates that greater resolution of phylogenetic relationships between strains and relatively recently diverged species may not be easily obtained. This is particularly important if no other phylogenetic information is available for the species under investigation or if the only other phylogenetic information is also from a region that has multiple variants.

An additional, variable region of the genome needs to be investigated in an attempt to obtain better resolution within the *P. westermani* species complex. This is the subject of the following chapter (Chapter 8), in which the mitochondrial gene NADH Dehydrogense subunit 1 (ND1) is partially sequenced for the purpose of obtaining a well-resolved phylogenetic signal.

CHAPTER 8

Multiple lineages of the mitochondrial gene NADH Dehydrogenase subunit 1 (ND1) in some trematodes.

8.1 Abstract

The use of partial ND1 sequences to infer a phylogeny for species of the genus *Paragonimus* (Trematoda: Digenea) is complicated by the discovery of at least two ND1 lineages within individual worms. The divergence of the ND1 lineages is shown by phylogenetic analysis not only to predate the divergence of the three *Paragonimus* species or species groups investigated, but also the divergence of some trematode families. Some sequences appear to be pseudogenes as they contain single base deletions and/or premature termination codons. The possible presence of both nuclear pseudogenes and mitochondrial heteroplasmy are invoked to explain the presence of multiple and divergent ND1 lineages. The implications for phylogenetic studies of trematodes using ND1 sequence data are discussed.

8.2 Introduction

I aimed to use ND1 sequences to obtain high resolution phylogenetic trees of the trematode genus *Paragonimus*, as has been done for the trematode genus, *Echinostoma* (see Morgan & Blair 1998). However, preliminary studies demonstrated that more than one sequence type occurred in each individual. I therefore investigated ND1 sequence variation present in *Paragonimus* individuals and in other trematode genera in detail and present my findings here.

It is generally considered that mitochondria are inherited through the female line and although mitochondrial genes are present in multiple copies in each cell, gene conversion, random replication and random segregation of mitochondrial genomes should maintain identical sequences within individuals (Birky 1996). However, if mitochondrial polymorphisms exist within species, the mitochondrial lineages may not neccessarily correspond to the species lineages (Avise 1986) due to possible introgression, asymmetrical hybridisation and lineage sorting. Here I present an analysis of sequence data for 70 clones of a partial ND1 gene fragment, obtained from 14 individuals of 3 species of *Paragonimus*. At least two distinct

ND1 sequence types were determined from nine individuals, representing all of the *Paragonimus* species in this study. In addition, I have included all available ND1 sequences from another three genera of trematodes (sources of data acknowledged in Table 8.1). The implications for phylogenetic studies are discussed. Furthermore, I explore the possible location of mitochondrial pseudogenes in trematodes of the genera *Echinostoma* and *Paragonimus*.

8.3 Materials and Methods

Nine geographic isolates of the lung fluke, *P. westermani* (from various locations in north east and south east Asia), representing both diploid and triploid forms, as well as four isolates of the *P. ohirai* species complex and a *P. miyazakii* isolate, were used to characterise partial ND1 sequences (the 5' half of the coding region) and the translated protein products (see Table 8.1 & Fig. 8.2). In addition, partial ND1 sequences of the liver fluke, *Fasciola hepatica* (see Garey & Wolstenholme 1989), eight species of the intestinal fluke genus, *Echinostoma* (see Morgan & Blair, 1998a & b) and two strains of the bloodfluke *Schistosoma japonicum* (from E. Sørensen, pers. comm.) were incorporated with the current data set for phylogenetic analysis of ND1 evolution in trematodes. DNA extractions were done as described in General Materials and Methods (Chapter 3).

8.3.1 PCR and sequencing

Total DNA was used as template for PCR amplification of a 500 bp fragment of ND1 using primers JB11 & JB12 (as described in Bowles & McManus 1993). In addition, specific PCR primers were designed for *Paragonimus* species, JB11par & JB12par (see Table 3.2). Amplification of a 500 bp fragment was performed as follows, using these primers: 30 cycles at 95 °C for 45 seconds, 40 °C for 30 seconds and 72 °C for 1 minute. PCR products were initially directly sequenced, but due to the presence of "mixed template" in purified PCR products (excised and purified from 1% LMP agarose gel slices), all subsequent PCR products were cloned prior to sequencing (2 - 12 clones per individual), using ABI Dye Terminator kits (as instructed by manufacturer). Contamination was excluded by i) including negative controls alongside positive controls, using both primer sets and ii) independent PCR and sequencing, using wholly independent reagents and templates, in a different laboratory (T. Agatsuma pers.

comm.), where "mixed template" sequences were also obtained when PCR products were sequenced directly. *Taq* polymerase errors are unlikely to affect the interpretation of the data, as they are random, non-informative changes that occur at a low rate and do not interfere with the phylogenetic inference of the sequence data. Genbank numbers for sequences obtained are presented in Table 8.1. Sequences were aligned and translated using the platyhelminth mitochondrial code (Garey & Wolstenholme 1989, Bowles, Blair & McManus 1992 and Ohama *et al.* 1990) in ESEE (Cabot 1997).

An alignment of partial ND1 sequences of representative clones for each lineage is shown (Fig. 8.1) to illustrate the extent of sequence variation between the lineages. Phylogenetic analysis of nucleotide and amino acid sequences was performed as described in General Materials and Methods (Chapter 3). Phylogenetic trees were rooted by mid-point rooting, without specifying an outgroup. Bootstrap support (1000 resamplings) was calculated and only bootstrap confidence limits (BCL) of ND1 lineages ≥ 70 were shown (Hillis & Bull 1993).

8.4 Results and Discussion

8.4.1 Multiple forms of ND1

Seventy partial ND1 sequences, representing 3 species of *Paragonimus*, were aligned and then translated. Fifty of the *Paragonimus* clones had unique sequences, which were used for phylogenetic analyses. The ND1 sequences of the liver fluke, *Fasciola hepatica* (from Garey & Wolstenholme 1989 : 1 sequence), the blood fluke, *Schistosoma japonicum* (2 sequences) and eight known species of intestinal fluke, *Echinostoma* spp. (from Morgan & Blair 1998 : 10 sequences) were included for comparative purposes. Two or more forms of ND1 were obtained from each individual of *Paragonimus*, though sequences were not obtained for all types in all individuals. A preliminary trial using single strand conformation polymorphism (SSCP) analysis of PCR products (Robin Gasser, pers. comm.) in all *Paragonimus* individuals confirmed the presence of multiple forms of ND1 sequence. However, the SSCP work is not reported further here.

8.4.2 Phylogenetic analysis

The comparison of distant taxa, such as those investigated here, indicated that saturation at the third codon position is likely to be complete (Morgan & Blair 1998a). In an effort to reduce this effect, only transversions were considered in the analyses of nucleotide data (473 characters in total). Most of the discussion is, however, based on analyses of amino acid sequences (157 characters). Nucleotides (transversion analysis) and amino acid sequences produced trees of identical topology for the *Paragonimus* isolates. Distance trees were

Table 8.1: Trematode ND1 sequences utilised in this study. Each line for data obtained in this study represents a single lungfluke individual, indicating species, geographical origin of isolate, number of clones sequenced, of which the number in parentheses were unique and used in the phylogenetic analysis. ND1 sequences were obtained from the same individuals that ITS1 sequences were obtained from (see Table 3.1). Data obtained from other sources is also identified, not neccessarily representing a single individual per line. Genbank numbers for representative ND1 variants are shown.

Species (ploidy) / id	Geographic location	# clones (sequenced)	Genbank number or source
P. westermani (2n) / PwJH2n	Japan, Hyogo	10 (8)	AF051844, AF063798-9
P. westermani (2n) / PwCM2n	E China, Minchin	10 (6)	AF051840, AF063789-90
P. westermani (2n) / PwJM2n	Japan, Mie	5 (5)	AF063783
P. westermani (2n) / PwP2n	Philippines, Sorsogon	9 (8)	AF052242, AF063788, AF063791
P. westermani (2n) / PwM2n	Malaysia	2	AF052243
P. westermani (3n) / PwJT3n*	Japan, Tsushima	9 (5)	AF063784-5
P. westermani (3n) / PwK3n	Korea, Bogil Isl.	6 (4)	AF051845, AF051843
P. westermani (3n) / PwCL3n	NE China, Liaoning	2 (2)	AF051842
P. westermani (3n) / PwJA3n	Japan, Amakusa	3 (2)	AF051841
P. ohirai / PoT	Japan, Tanegashima	2	AF063786
P. ohirai / PoK	Japan, Kinosaki	2	AF063787
P. iloktsuenensis / Pilok279	Japan, Amami	4 (2)	AF063792, AF063795
P. iloktsuenensis / PilokK	Japan, Amami	3	AF063793, AF063796
P. miyazakii / Pmi	Japan, Miyazaki	3	AF063794,AF063797
S. japonicum / Sj [+]	Hong Kong,	2	E. Sørenson
Echinostoma spp.	World-wide, Australia	8	Morgan & Blair 1998a & b
F. hepatica / Fhep	Utah, USA	1	M93388

* Occurs sympatrically with *P. westermani* (2n). [+] Represents 2 strains from Hong Kong and Anhui, China.

identical, regardless of the distance model used. Neither were there any differences in the tree topologies, whether obtained by distance or parsimony methods.

When Maximum Parsimony analysis of amino acid sequences was done, 4104 most parsimonious trees (MPT's) of length 418, were obtained. The 70% majority rule consensus tree of the MPT's (Figure 8.2a) had a Consistency Index (CI) of 0.758 and a Retention Index (RI) of 0.928. The Rescaled Consistency Index (RC) was 0.704.

Figure 8.1: Partial ND1 sequence alignment of representative clones for each ND1 type (as defined in section 8.4.1 below). Codes used are as described in Table I, with the ND1 "type" indicated by I, II, III or IV. Primer binding sites for JB11par and JB12par are indicated.. JB11 anneals 40 bp 5' to JB11par and JB12 anneals near the 3' end of JB12par, as shown and an additional 16 bp downstream. A . indicates that there is no base change relative to bases presented on the top line of the set. - indicates a deletion.

```
                JB11par-------------------/
PwP(2n)I     ATT CAG AGG TTT GCT GAT CTA TTG AAG TTG GTG ATA AAG TTC AAG
PwM(2n)I     T.G ... .C. ..C ... ... ..T ..A ... C.. ... ... ... ..T ...
PwCM(2n)I    ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ...
PwCL(3n)I    ... ... .A. ... ... ... ... ... ... ... ... ... ... ..T ...
PoT I        ..A ... .A. ... ..A ... ..T ... ... ... ..T ... ... ..T ...
Pmi I        ... ... ... ... ... ... ... ... ... ... ..T ... ... ..T ...
FHEP I       T.G ... ... ... ..G ... T.. A.. ... ..A ..T ... ... ..T ...
PwP(2n)II    T.G ..A ... ... ..G ... ..T ..A ... ..A A.C ... ... .AT ...
PwCM(2n)III  ... ... ... ... ... ... ... ... ... ... A.T ..T ... ... ...
Pmi III      ..A ... ... ... ... ... T.. ... ... ... A.T ..T ... ... ...
LCERC1 III   T.A ... ... ... ... ... T.. ... ... ... A.T ..T ... ... ...
PMETA3 III   T.A ..A ... ..C ..G ... T.G ..A ... ... A.T ..T ... ..T ...
ESP1 IV      T.G ... ... ... ..A ... ..C ..A ... ... ..T ... ... ..T ...
ETRIV IV     T.A ... ... ... ... ... ..T ..A ... C.. A.T ... ... ..T ...

PwP(2n)I     GTT GCC CTG TTT CAG GGG CGT AGT TGG CTC TCT TGA TGG GGT GTG
PwM(2n)I     ... .TT ... ... ... ..A ... ... ... ... ... ... ..A ... ...
PwCM(2n)I    ... ..T ... ... ... ... ... ... ... GG. ..G ..G ..A ..C ...
PwCL(3n)I    ... ..T ... ... ... ... ... ... ... ... ..G ..G ..A ..C ...
PoT I        ... T.T TCC ... ... ..T ... ... ... ..T ... ... ..A ..G ...
Pmi I        ... T.T TCC ... ... ..T ... ... ... ..T ... ... ..A ..G ...
FHEP I       T.. .TG T.T ... ... AAT ... ... ... T.G ... ..G ... ... ..T
PwP(2n)II    ... ATG T.T ... ATA CCC -.. ..G ..A T.G .G. ... GT. ..C ..T
PwCM(2n)III  ... ATG T.T ... ... AAT ... ..G ..A T.G ... ..G CT. ... ..T
Pmi III      ... ATG T.T ... ... AAT ... ..G ..A T.G ... ..G CT. ..A ..T
LCERC1 III   ... ATG T.T ... ... AAT ... ..G ..A T.G ... ..G CT. ..C ..T
PMETA3 III   ..G ATG T.T ... ... AAT ..A .C. ... T.G ..G ..G .T. ..G ...
ESP1 IV      T.. .TT T.T ... ..A AAT ... ... ... T.G ... ..G GTT ... ..T
ETRIV IV     T.. A.T T.C ... ... AAT ... ... ..A T.A ... ..G GTT ..G ..T
```

```
PwP(2n)I     TAC GTC CTG GTG TTT TTG GGT TGT TTT TAT TGT GTT CTC TTT TCT
PwM(2n)I     ... .C. T.. ... ... C.. ... ... ..C ..C ... ..C T.. ... ...
PwCM(2n)I    ... AC. T.. ..A ... ..A ..C ... ... ..C ... ..C T.. ..C ...
PwCL(3n)I    ... AC. T.. ..A ... ..A ..C ..C ... ..C ... ..C T.. ..C ...
PoT I        GGT ... T.. ..T C.. ... ... ... C.. ... ... ..C T.T ..C ...
Pmi I        GGT ... T.. ..T C.. ..A ... ... C.. ... ... ..C T.T ..C ...
FHEP I       ..T T.G T.. ..T ..G ... .C. ... GGC ... ... ..G T.G ... .T.
PwP(2n)II    ..T .CT T.. ..T ..C ..A .C. ..C ... ... ... A.G A.G ... AG.
PwCM(2n)III  ..T T.G T.. ... G.. ... TC. ... GGG ... ... ..G A.T ..A G..
Pmi III      ..T T.G T.. ... G.. ... TC. ... GGG ... ... ..G A.T ..A G..
LCERC1 III   ..T T.G T.. ... G.. ... TC. ... GGG ... ... ..G A.T ..A G..
PMETA3 III   ..T T.G T.. ..T C.. ... TCA ... GG. ... ... ... G.T ..G G.G
ESP1 IV      ..T T.G T.. ..T ..G ..A TC. ... GG. ... ... ... G.T A.. GG.
ETRIV IV     ..T T.G T.. ..T ..G ... TC. ... GC. ... ... ... G.T ..G .G.

PwP(2n)I     CAC AGC TAT TCT GGT ATG AAA GGG GGT GTT GGC ATG CTT TGG TTG
PwM(2n)I     ... ..G ... ... ... ... ... ... ... ... AA. ... ..C ... ...
PwCM(2n)I    ... ... ... ..C ... ... ... ... ..C A.C A.. ... ... ... ...
PwCL(3n)I    ... ... ... ..C ... ... ... ... ..C A.C A.. ... ... ... ...
PoT I        TTT ..T ... ... ... T.. .GG A.A A.A AA. A.A GCT ... ... ...
Pmi I        TTT ..T ... ... ... T.. .GG A.A A.A AA. A.A GCT ... ... ...
FHEP I       TTT ..G .T. GG. ... G.T .GT A.T .T. AAA TTT ... T.G ... ..T
PwP(2n)II    .TT GCT ... AGG ... T.. .GG T.T .T. AA. TAT ... ..C ... ...
PwCM(2n)III  .TG G.T ... AGG ..C ... .GT T.T A.A AAC TTA ... T.A ... ..T
Pmi III      .TG G.T ... AGG ..C ... .GT T.T A.A AAC TTA ... T.A ... ..T
LCERC1 III   .TG G.T ... AGG ..C ... .GT T.T A.A AAC TTA ... T.A ... ..T
PMETA3 III   ..A GC. ... AGC ..G G.T .GT TCT .A. TA. TTG ... T.A ... ..T
ESP1 IV      TTG .TG C.. ATA ... ..T T.T A.. .A. AAA ATT ... T.G ..A ..T
ETRIV IV     TTA .TG C.. AGG ..C G.T TTT A.T .A. AA. ATA ... T.G ..A ..T

PwP(2n)I     TTG GTT GTG ACG AGT GTG ACC GGG TAT AGT CTC TTG AGT GTC GGC
PwM(2n)I     ..A ..G ..A ... ... ... ... ... ... ... ... ... ... ... ...
PwCM(2n)I    ... ..G ... ... ... ... ..G ... ..C ... ... ... ... ..T ...
PwCL(3n)I    ... ..G ... ..C .A. ... ..G ... ..C ... ... ... ... ..T ...
PoT I        ... ..C A.T ..T ... ..T ..A ... ... ... ..T ..A ..G A.A ..T
Pmi I        ... ..C A.T ..T ... ..T ..A ... ... ... ..T ..A ..G A.A ..T
FHEP I       ..A ..G ..T ..T ... A.. ..T ..T ... ..G T.G ..A ... ..T ..T
PwP(2n)II    C.T ... A.T ..A ... AAC .GT ..T ... ... T.T ..A ... A.T ..G
PwCM(2n)III  ... .C. T.A ... ... ..A ..T ..T ... ... T.G ... ..A ..G ..T
Pmi III      ... .C. T.A ... ... ..A ..T ..T ... ... T.G ... ..A ..G ..T
LCERC1 III   ... .C. T.A ... ... ..A ..T ..T ... ... T.G ... ..A ..G ..T
PMETA3 III   ..A ... T.. ..T ... A.A ..A ... ... ... T.G ... ... A.T ..A
ESP1 IV      ... A.. A.T ..T ..G A.T ..A ..C ... ... ..T ..A ..G ..T ..T
ETRIV IV     ..A A.. A.A ... ... A.A ..T ..T ... ... T.A ..A ..G ..T ..T

PwP(2n)I     TGG GGT TCT TAT AAA AAG TAT GCC TTG CTG AGC TGT CTG CGT TCT
PwM(2n)I     ..A ... ..C ... ... ... ... ... ... ... ... ..C T.. ... ...
PwCM(2n)I    ... ..C ..C ..C ... ... ... ..T ... ... ... ... T.. ..C ...
PwCM(3n)I    ... ..C ..C ..C ... ... ... ..T ... ... ... ... T.. ..C ...
PoT I        ... ... ... ... ... ... ... ..T ... T.. ..A ... T.. ..A ...
Pmi I        ... ... ... ... ... ... ... ..T ..A T.. ..A ... T.. ..A ...
FHEP I       ... ... .G. ... ... ... .T. ..T ... G.T ... ... G.T ... ...
PwP(2n)II    ..A ... ... ... ..T ... .T. ..T ..A ..T ... ... T.A ... ...
PwCM(2n)III  ... ..C ... ... ..T ... .T. ..T ... ..A ..G ... G.A ... ..G
Pmi III      ... ... ... ... ..T ... .T. ..T ... A.A ..G ... G.A ... ..G
LCERC1 III   ... ... ... ... ..T ... .T. ..T ... ..A ..A ... G.A ... ..G
PMETA3 III   ... ... ..G ... ..T ... .T. ..T ... T.A ..G ... G.T ... ..G
ESP1 IV      ... ... ... ... ..T ... ... ..T ... ..T ..T ... G.T ... ...
ETRIV IV     ... ... ... ... ..T ... ... ..T ..A ..T ..A ... G.T ... ...
```

```
PwP(2n)I     GCG TTC AGT TCG ATA ACT TTT GAG GCG TGC TTC ATG TGC ATT GTC
PwM(2n)I     ... ..T ... ... ... ... ... ... ..A ... ... ... ... ... ..T
PwCM(2n)I    ... ..T ... ... ... ... ... ... ... ... ... ... ... ..C ...
PwCL(3n)I    ... ..T ... ... ... ... ... ... ... ... ... ... ... ..C ...
PoT I        ... ..T G.. ..T G.G ... ... ..A ..C ..T ..T ... ..T ... ..T
Pmi I        ... ..T G.. ..T G.G ..C ... ..A ..C ..T ..T ... ..T ... ..T
FHEP I       ..T ..T G.G ..T G.T .GG ... ... ..T ..T ..T ... ..T ... ..T
PwP(2n)II    T.A ..T GC. ... G.T .GG ... ... ... ..T ... ... ..T G.G .CA
PwCM(2n)III  ..C ..T G.. ... ..T .GG ... ... ..T ..T ..T ... ..T G.. ..T
Pmi III      ..C ..T G.. ... ..T .GG ... ... ..T ..T ..T ... ..T G.. ..T
LCERC1 III   ..C ..T G.. ... ..T .GG ... ... ..T ..T ..T ... ..T G.. ..T
PMETA3 III   ... ..T G.. ... G.T .G. ... ... ... ..T ..T ... ..T ... ..A
ESP1 IV      ..T ..T G.. ..T G.G .G. ... ..A ..T ..T ..T ... ..T G.. ..T
ETRIV IV     ..C ... G.G ..T G.. .GG ... ..A ..T ..T ..T ... ..T G.. ..T

PwP(2n)I     CTT CTG ATA GCC ATT GTT GGT GGG AGT TAC TCG CTG AGA TAC TAC
PwM(2n)I     ..C T.. .C. ... ... ... ... ... ... ... ... ... ... .TT ...
PwCM(2n)I    ... T.. GC. ... ... ..C ... ... ..C ... ... T.. .A. .TT ...
PwCL(3n)I    ... T.. GC. ... ... ..C ... ... ..C ... ... T.. .A. .TT ...
PoT I        T.A T.. T.G ..T ..A ... A.G ... .T. ..T ..A T.A G.T CCT .GT
Pmi I        T.A T.. T.G ..T ..A ... A.G ... .T. ..T ..A T.A G.T CCT .GT
FHEP I       G.. T.. G.T ..A T.G ... T.G ... ... ..T GGT G.T TCT .GT .TG
PwP(2n)II    G.. A.. G.T ..T T.A A.. .T. ..T ..C ..T GA. A.- .CT CCG CTT
PwCM(2n)III  A.A A.. G.G ..G G.A ..G .T. ..A ..G ... GGT G.A G.T AGT .TG
Pmi III      A.A A.. G.G ..G G.A ..G .T. ..A ..G ... GGT G.A G.T AGT .TG
LCERC1 III   A.A A.. G.G ..G G.A ..G .T. ..A ..G ... GGT G.A G.T AGT .TG
PMETA3 III   A.. A.. G.T ..T G.. ... .TG ..A .C. ..T GGT G.T G.G AG. .TG
ESP1 IV      A.A A.. G.T ..T T.G .C. .TG ..T T.. ..T GGT G.A TTT AGT .TT
ETRIV IV     A.A A.. G.T ..T ..G ... ATG ..T T.. ... GGT G.. .CT GGT ATG

PwP(2n)I     CTG GGA GGG CCT TTT CTC AGA TGT CTG GTC TTT CCT CTT TGT TAC
PwM(2n)I     ... ..G ... T.. ... ... ... ..C T.. ..T ... ..C ..C ..C ...
PwCM(2n)I    ... ... ... ..C ..C ... .A. ... T.. ... ... ... ..C ..C ...
PwCL(3n)I    ... ... A.. ..C ..C ... .A. ... T.. ... ... ... ..C ..C ...
PoT I        T.. .AG C.T T.. ... T.T .TT .C. T.A TCT ..G ..C TG. ... ..T
Pmi I        T.. .AG C.T T.. ... T.T .TT .C. T.A TCT ..G ..C TG. ... ..T
FHEP I       T.T ..T .AA TT. GG. GGT .TG ..A A.. ..T G.. ... G.A GT. ..T
PwP(2n)II    G.T .AG AAT AA. .GG T.T GTT ..G T.A ..T G.. ... .C. GT. ..T
PwCM(2n)III  A.T TC. ..T GTG .GG T.G .TT G.. T.. A.T A.G ... ..C GT. ...
Pmi III      A.T TC. ..T GTG .GG T.G .TT G.. T.. A.T A.G ... ..C GT. ...
LCERC1 III   A.T TC. ..T GTG .GG T.G .TT G.. T.. A.T A.G ... ..C GT. ...
PMETA3 III   G.T A.T A.. G.. .GG T.G .TT AA. T.. A.T ..G ... ..G GT. ..T
ESP1 IV      T.T TCT .A. GTA .GG ..T TTT .CA T.T ..T ..G ... T.G GTG ..T
ETRIV IV     T.T TAT AAT G.. .GG T.T TTT .C. G.T ..T ..A ..A T.G GT. ..T

PwP(2n)I     ATT CTC TGG TTA GTG GGT ATC CTG TGC GAA TGT AAC CGG ACG CCT
PwM(2n)I     ... ... ... ..G ..A ... ... T.. ..T ..G ..C ..T ... ... ...
PwCM(2n)I    ... ... ... ... ... ... ... ... ..T ..G ..C ..T ... ..T ...
PwCL(3n)I    ... ... ... ... ... ... ... ... ..T ..G ..C ..T ... ..T ...
PoT I        ... ..T ..A ..G ... ... .GT T.. ..T ..G ... ..T ... ..T ...
Pmi I        ... ..T ..A ..G ... ..A .GT T.. ..T ..G ... ..T ... .An ...
FHEP I       GG. T.G ... C.G ... ... ..A T.A ..T ... ... ... ..T ..T ..A
PwP(2n)II    GGC A.T ..A C.G ... ... ..T T.. ..T ... ... ... ..T ..T ..C
PwCM(2n)III  GG. T.A ... ... ... ... ..A T.. ... ..G ... ..T ..T .AA ...
Pmi III      GG. T.A ... ... ... ... ..A AC. ... ..G ... ..T ..T ..T ...
LCERC1 III   GG. T.A ... ... ... ... ..A T.. ... ..G ... ..T ..T ..T ...
PMETA3 III   GGG T.G ... ..G ... ... ..T T.A ..T ... ... ..T ..T ..C ..G
ESP1 IV      GG. T.G ..A ..G ..A ..C ..A T.. ..T ..G ... ..T ... ..C ..G
ETRIV IV     GGA T.A ... ..G ... ... ..T T.. ..T ... ... ... ..T ..T ..C
```

```
                                                >JB12>>>
                  JB12par--------------------\
PwP(2n)I      CTG GAT TAT GCT GAG GCT GAA AG
PwM(2n)I      T.. ... ... ... ... ... ... ..
PwCM(2n)I     T.. ... ... ... ... ... ... ..
PwCL(3n)I     T.. ... ... ... ..A ... ... ..
PoT I         T.. ..C ... ... ... ... ... ..
Pmi I         T.. ..C .TA ... ... ... ... ..
FHEP I        T.. ... ... ... ... ... ... ..
PwP(2n)II     A.. ... ... ... ..A ... ... ..
PwCM(2n)III   T.. ... ... ... ... ... ... ..
Pmi III       T.. ... ... ... ... ... ... ..
LCERC1 III    T.. ... ... ... ... ... ... ..
PMETA3 III    T.. ..C ... ..G ... ... ... ..
ESP1 IV       T.. ... ... ..G ... ... ... ..
ETRIV IV      T.A ... ... ... ... ..G ... ..
```

Figure 8.2. Partial ND1 amino acid sequence alignment of representative clones for each ND1 type, as described in Fig. 8.1. A . indicates that there is no amino acid sequence change relative to the amino acids presented on the top line of the set. * indicates the presence of a premature termination codon and - indicates a deletion, in order to maintain an in-frame translation of the nucleotide sequence.

```
PwP(2n)   I   1 IQSFADLLKLVIKFKVALFQGRSWLSWWGVYVLVFLGCFYCVLFSHSYSGMNGGVGMLWL
PwM(2n)   I   2 L.T.............V.............A.........F.............N....
PwCM(2n)  I   3 ..................G......T.........F............IS....
PwCL(3n)  I   4 ..K.................N........T.........F............IS....
     PoT  I   5 ..KL...........SS...........G...L...L...F..F....LSSSNSA...
     Pmi  I   6 .............VSS...........G...L...L...F..F....LSSSNSA...
    FHEP  I   7 L......M.........F..N.........L..L.A.G.....FF.FG.VSSVNF...F
PwP(2n)  II   8 L.........I..Y..MF.I-....C.V...A....A....MM..LA...LSCVNY....
PwCM(2n)III   9 ..........I.....MF..N......L...L..V.S.G...ILALG....SCSNL...F
   Pmi  III  10 ..........I.....MF..N......L...L..V.S.G...ILALG....SCSNL...F
LCERC1  III  11 L.........I.....MF..N......L...L..V.S.G....LALG....SCSNL...F
PMETA3  III  12 L.........I.....MF..N.T....L...L..L.S.G...VLAQA.S.VSSDYL...F
   ESP1  IV  13 L............FVF..N......V...L..L.S.G...VIGLMHI.IYSDNI...F
  ETRIV  IV  14 L.........I....FTF..N......V...L..L.S.A...VLCLMH..VFSDNI...F
               -------10+-------20+-------30+-------40+-------50+-------60+

PwP(2n)   I   1 LVVTSVTGYSLLSVGWGSYNKYALLSCLRSAFSSITFEACFMCIVLLIAIVGGSYSLNYY
PwM(2n)   I   2 ..............................................T..........F.
PwCM(2n)  I   3 ..............................................A..........F.
PwCL(3n)  I   4 ....N.....................................R.......A..........F.
    PoT   I   5 ..I..........I..................G.V.....L......L...S.I...GPC
    Pmi   I   6 ..I..........I..................G.V.............L...S.I...GPC
   FHEP   I   7 .....M...........C...F..V..V....G.VS.........V.V.L.W...GVSCL
PwP(2n)  II   8 ..I..NS...F..I.......F........S.A.VS.......VAVMV.LIV...E-TPL
PwCM(2n)III   9 .AL..................F.....V....G..S.......V.IMVA.VVG..GVGSL
  Pmi   III  10 .AL..................F..I..V....G..S.......V.IMV.V.V...GVGSL
LCERC1  III  11 .AL..................F.....V....G..S.......V.IMV.V.V...GVGSL
PMETA3  III  12 ..L..I.......I.......F.....V....G.VS.........IMV.V.V.TYGVGSL
   ESP1  IV  13 .II..I.....................V....G.VS.......V.IMV.LAV.C.GVFSF
  ETRIV  IV  14 .II..I.....................V....G.VS.......V.IMV.M.M.C.GVTGM
               -------70+-------80+-------90+------100+------110+------120+
```

```
PwP(2n)   I  1 LGGPFLSCLVFPLCYILWLVGILCECNRTPLDYA
PwM(2n)   I  2 ...S..N.....................
PwCM(2n)  I  3 ......N.....................
PwCL(3n)  I  4 ..S...N.....................
    PoT   I  5 .ERS.FIS.SL.C........S.....
    Pmi   I  6 .ERS.FIS.SL.C........S......?...L.
   FHEP   I  7 F.EFGGMWM.V.VV.G...............
PwP(2n)  II  8 VENNWFVW..V.PV.GI.............M...
PwCM(2n)III  9 IS.VW.IG.IM..V.G............N.....
  Pmi   III 10 IS.VW.IG.IM..V.G......T...........
LCERC1  III 11 IS.VW.IG.IM..V.G...............
PMETA3  III 12 VSSAW.IN.IL..V.G...............
   ESP1  IV 13 FSEVW.FSF.L..V.G...............
  ETRIV  IV 14 FYNAWFFSV.L..V.G...............
             ------130+------140+------150+----
```

Table 8.2 Number of differences in amino acids (above the diagonal) and nucleotides (below the diagonal) for different types of ND1 as represented by selected clones, which were also used in Table 8.1 and Fig. 8.1. Clone id. codes are as for Table 8.1 and Fig. 8.1.

Type	Representative clones	Type I 1	2	3	4	5	Type II 6	Type III 7	8	9	10	Type IV 11	12
Type I	1, PwCM(2n)3,4	-	9	8	36	51	57	47	48	49	53	55	55
	2, PwM(2n)	9	-	7	36	49	55	48	47	48	52	53	54
	3, PwP(2n)1,2	8	10	-	36	49	56	46	47	48	52	54	54
	4, PoT	43	42	41	-	53	57	52	53	54	50	55	53
	5, Fhep	90	85	86	97	-	53	44	43	44	41	38	41
Type II	6, PwP(2n)c1,c11	89	84	85	92	81	-	49	48	49	49	54	51
Type III	7, PwCM(2n)1c,4c	91	92	85	92	75	78	-	1	2	21	34	36
	8, Lcerc1	93	90	87	92	73	76	2	-	1	20	33	35
	9, PwP(2n)c8	93	90	87	92	73	76	2	2	-	21	34	36
	10, Pmeta3	91	92	89	98	77	84	54	52	52	-	32	31
Type IV	11, Esp1	97	92	93	88	63	80	68	66	66	88	-	19
	12, Etriv	99	92	93	90	65	84	58	56	56	64	32	-

A number of distinct lineages with high bootstrap support were resolved. One consists entirely of *Paragonimus* clones with strong support (bootstrap confidence limit, BCL = 100), and incorporates ND1 sequences from *P. westermani, P. ohirai* & *P. miyazakii* isolates (Fig. 8.3a & b). This ND1 sequence type could be amplified with both primer pairs, JB11-JB12 and

JB11par-JB12par. This lineage, denoted "type I", maintains phylogenetic structure, as isolates of both *P. ohirai* and *P. westermani* species complexes cluster within their respective species groups. The *P. westermani* species complex (BCL = 100) has a northeast Asian isolates cluster (BCL = 90), which is distinct from the two southeast Asian isolates which are also distinct from each other. However, no strong differentiation between northeast Asian diploid and triploid isolates is achieved. In addition, members of the *P. ohirai* species complex (*P. ohirai* and *P. iloktsuenensis*) are distinguished from each other, (BCL = 76), despite the small difference between them (1 amino acid). This "type I" lineage is considered to represent functional mitochondrial ND1, as the inferred phylogeny of this lineage is largely congruent with the recognised phylogeny for the genus determined by ITS2 and COI sequence data (Blair *et al.*1997a). Furthermore, premature termination codons are absent. *Fasciola hepatica* is also associated with this lineage (BCL = 100).

A second lineage, denoted "type II", contains pseudogenes. These were detected in three individuals of *P. westermani* and one of an Australian echinostome, Pcerc-3 (BCL =100), only when amplified with JB11-JB12 primers. In the sequences from the three *P. westermani* individuals, a single base deletion occurred causing a frame shift. Furthermore, the *Paragonimus* sequences contained premature termination codons. The echinostome isolate (Pcerc3), did not have a deletion, nor did it have premature termination codons, though it clustered with these *Paragonimus* pseudogenes (BCL = 100).

A third lineage of clones, "type III", included representatives of both species complexes of *P. westermani* and *P. ohirai* (as represented by *P. iloktsuenensis*), as well as *P. miyazakii* and an Australian echinostome, LCerc-1 (BCL = 100). Both primer pairs, JB11-JB12 and JB11par-JB12par, amplified these "type III" variants. All members of this lineage were very similar to one another and it may either be another pseudogene lineage or an alternative functional form of ND1 (no insertions or deletions were present, nucleotide sequences had diverged about 30% from the assumed functional "type I" sequence (see Table 8.2) and 20 % of clones in this lineage contained a premature termination codon somewhere in the sequenced fragment). There was no "phylogenetic structure" in this lineage, as variants of clones from all three species represented were intermingled. Also associated with this lineage was Pmeta3 (from Morgan & Blair 1998), another Australian echinostome species (BCL = 99).

The fourth lineage, denoted "type IV", was the *Echinostoma* cluster (*E. sp1, E. paraensi, E. caproni, E. revolutum and E. trivolvis*) with strong bootstrap support (BCL = 99). This also unexpectedly incorporated a pair of clones from the triploid *P. westermani* from Tsushima, Japan (PwJT(3n) clones 4 & 5). It is difficult to explain this anomaly. The *Paragonimus* variant (PwJT(3n) clones 4 and 5) was distinct from *E. paraensi*, suggesting that contamination of samples was not an explanation. It is possible that this is yet another pseudogene cluster, as it was more closely associated with the "type III" cluster than with the "type I" cluster (BCL = 99), however, no premature termination codons occurred in any of these sequences. Further work is required to clarify this matter. ND1 was recently used to identify echinostome isolates from Australia and New Zealand (Morgan & Blair 1998b). The authors concluded from the phylogenetic inferences based on ND1 sequence data that five unidentified echinostome species occur in Australia, based on 9.6 - 30.8% sequence divergence. These findings will be further discussed in the light of data presented here, in section 8.5 below.

Echinostoma hortense, a 28 collar-spined fluke could not be placed in any of the above lineages. Finally, *S. japonicum* behaved like an outgroup (BCL = 100) in the mid-point rooted tree.

8.4.3 Molecular variation summarised

All the *Paragonimus* species investigated here had multiple ND1 variants within individuals. This is not, however, true for another mitochondrially encoded gene, COI (Blair & Agatsuma 1997, Blair *et al.* 1997). There is evidence of an ND1 gene duplication event that predates the speciation event which resulted in *P. miyazakii* and the *P. westermani* and *P. ohirai* species complexes. Indeed, it may even predate the differentiation between the genera *Echinostoma* and *Paragonimus,* which belong to distantly related families.

Three to four ND1 "types" have been identified in this study: I) Apparently functional ND1, showing variation that is consistent with the continuing evolution of the mitochondrial DNA and a phylogeny congruent with that from COI and ITS2 sequence data. II) Deleted clones, assumed to be pseudogenes. III) One tight lineage of clones, 20% of which have premature stop codons when translated. This lineage may be another, independent pseudogene lineage. IV) A lineage of uncertain status, which may encode either functional *Echinostoma* ND1 or another pseudogene. Paths connecting presumed functional ND1 copies (type I),

deleted pseudogenes (type II), presumed pseudogenes (type III), type IV and other trematodes are summarised graphically for convenience (Fig. 8.3b).

Pseudogenes can accumulate mutations equally in all three codon positions, as they are free from the selective constraints which act, particularly on 1st and 2nd codon positions, of functional genes. The high proportion of non-synonymous base changes in "types II and III" relative to "type I" ND1 (shown in Table 8.2 and Fig. 8.2), the presence of premature termination codons in all "type II" clones and some "type III" clones, as well as the lack of phylogenetic structure amongst variants of the "type III" ND1 clade would suggest pseudogenes.

8.4.4. Pseudogenes and nuclear mitochondrial genes (numt) DNA

Pseudogenes are non-functional by-products of the processes of molecular evolution, whereby a functional gene is duplicated. We do not know whether the ND1 pseudogenes in *Paragonimus* and *Echinostoma* are present in the mitochondrial genome, or if they have been integrated into the nuclear genome.

Mitochondrial genes and nuclear homologues vary according to the evolutionary history in the group being studied. For example, in aphids a different pattern of variation of nuclear integrations has been reported (Sunnucks & Hales 1996) to the pattern of variation observed in the ND1 sequences of trematodes reported here. Each CO I & CO II numt pseudogene detected in some species of aphids had a unique sequence. The aphid *numt* variants present in each species were generally most closely related to their mitochondrial counterparts present within the same species. *Paragonimus* and *Echinostoma* species, however, have a different pattern to that seen in aphids, as the different variants fall into distinct, well-supported lineages (Fig. 8.3).

However, the pattern of ND1 evolution reported here closely resembles that of the hominoid D-loop sequences in the nuclear genome (Zischler, Geisert & Castresana 1998), which suggests translocation from the mitochondrial to the nuclear genome, where different evolutionary dynamics exist. In the hominoid lineages the functional mitochondrial sequences show strong phylogenetic structure (c.f. type I in Fig. 8.3), whereas the numt sequences lack such structure (c.f. type III in Fig. 8.3) and are all very similar to each other. In both cases a tight but unresolved cluster of "numt" sequences was formed on the tree (Zischler, Geisert & Castresana 1998). The authors speculated that the hominoid patterns of divergence may be the

result of "a sudden change of the sequence shaping forces ... that leads to a post-transpositional **temporary** increase in sequence change" followed by relative stasis. This phenomenon may be due to the change from mt γ polymerase-based to nuclear α, β, δ and ε polymerase-based replication, repair and recombination of numt sequences (Zischler, Geisert & Castresana 1998).

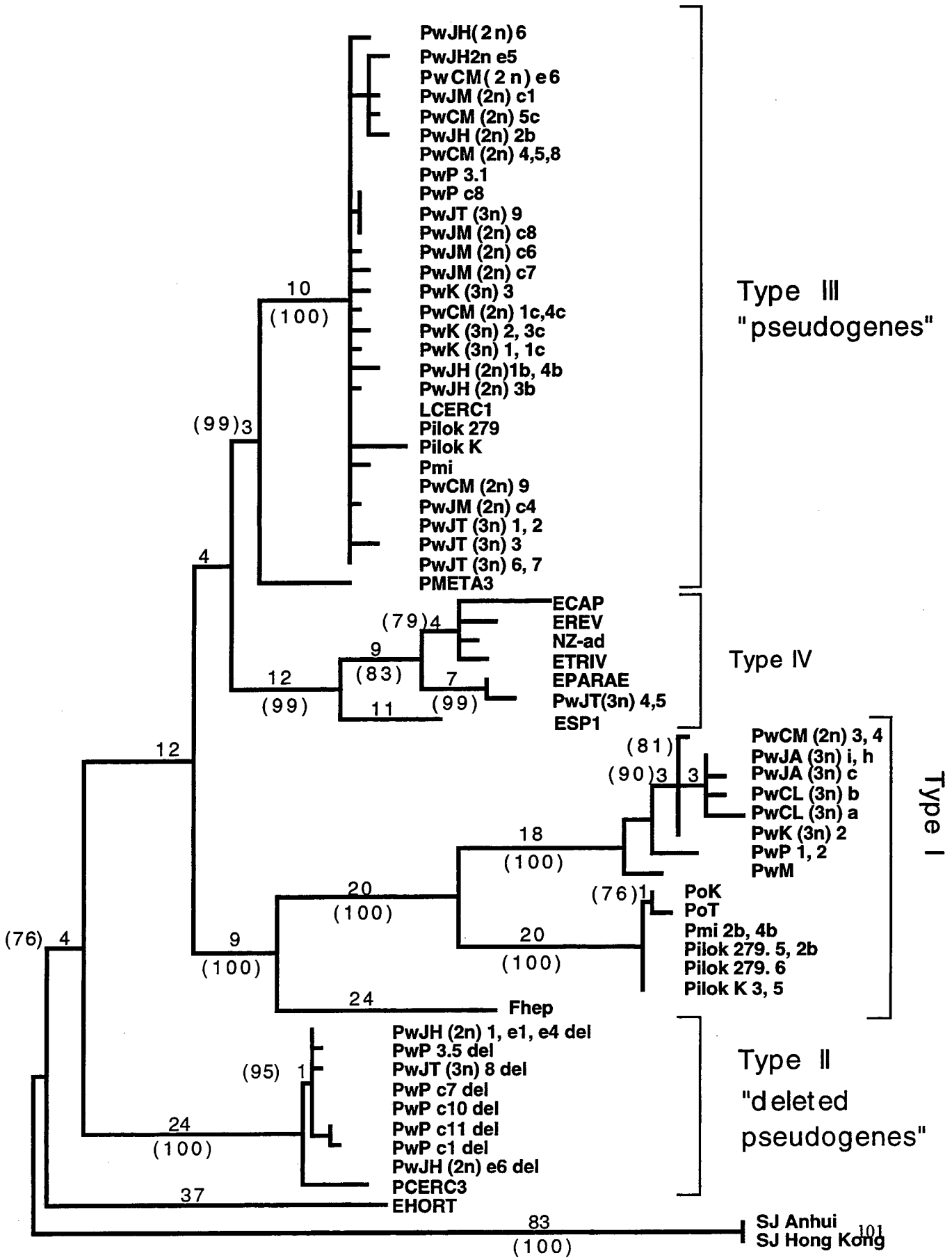
The presence of numts in the genomes of a wide range of taxa (Lopez *et al.* 1994, Zhang & Hewitt 1996, Sorensen & Quinn 1998) and the limited size of mitochondrial genomes suggest that the translocation of mitochondrial genes to the nuclear genome is a likely explanation of the intra-individual ND1 variation reported here. This may be confirmed by amplifying ND1 from purified mt DNA, which should lack numt variants. In addition, ND1 type-specific primers can be designed, which may be used to selectively amplify the desired ND1 type from total genomic DNA.

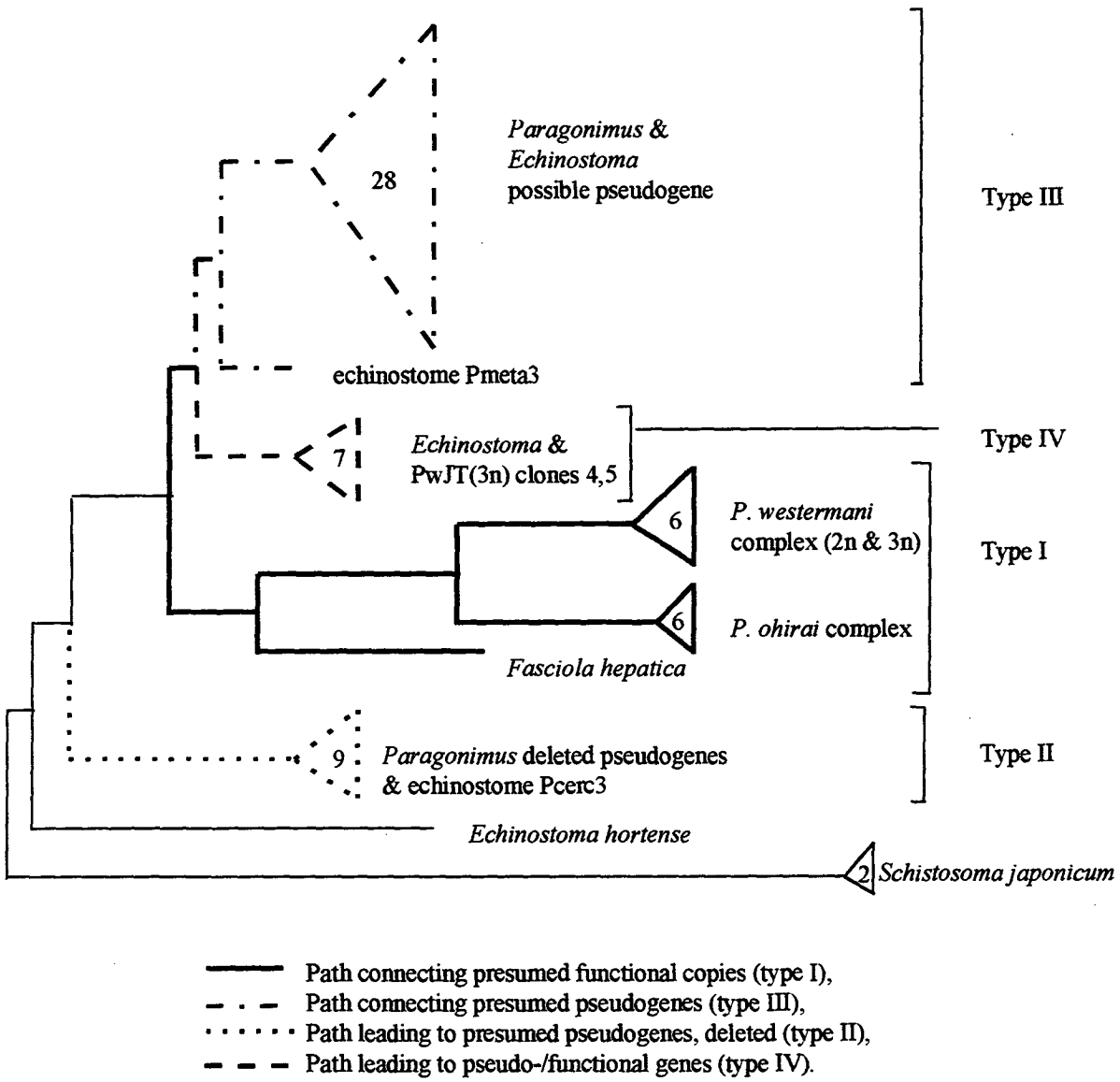
8.4.5 Pseudogenes and heteroplasmy

There is an alternative explanation regarding the location of putative pseudogenes described here. Mixed populations of mitochondria, which might possess different genomes (heteroplasmy), might exist within individuals. Distinct cytochemical differences exist in the mitochondria of the tegumental and parenchymal cells of some of 5 trematode species investigated (*Paragonimus ohirai*, *Echinostoma trivolvis*, *Schistosoma mansoni*, *Fasciola hepatica* and *Zygocotyle lunata*) (Fujino, Fried & Takamiya 1995). *P. ohirai* and *E. trivolvis* each have two well-represented, structurally different forms of mitochondria which were present in different cell types. *S. mansoni* also has two structurally different mitochondrial types, however, one of the forms was in much lower abundance. In contrast, *F. hepatica* and *Z. lunata* had almost exclusively one structural form of mitochondrion.

Figure 8.3a (overleaf)

70% Majority rule consensus tree (midpoint rooting) from 4104 MPT's based on amino acid sequences, from a heuristic search with Tree Bisection Reconnection (TBR) branch-swapping and MULPARS in effect (PAUP 3.11). 1000 bootstrap replicates were performed on the N-J tree, which had identical topology (MEGA). Bootstrap confidence limits (BCL) are shown in parentheses, where BCL > 70. Assigned branch lengths are shown above the lines. The length of the lines connecting nodes is proportional to branch lengths. b) Summary phylogenetic tree of data presented in Fig. 8.3a.


PwJH(2n)6
PwJH2n e5
PwCM(2n)e6
PwJM (2n) c1
PwCM (2n) 5c
PwJH (2n) 2b
PwCM (2n) 4,5,8
PwP 3.1
PwP c8
PwJT (3n) 9
PwJM (2n) c8
PwJM (2n) c6
PwJM (2n) c7
PwK (3n) 3
PwCM (2n) 1c,4c
PwK (3n) 2, 3c
PwK (3n) 1, 1c
PwJH (2n)1b, 4b
PwJH (2n) 3b
LCERC1
Pilok 279
Pilok K
Pmi
PwCM (2n) 9
PwJM (2n) c4
PwJT (3n) 1, 2
PwJT (3n) 3
PwJT (3n) 6, 7
PMETA3
Type III
"pseudogenes"
ECAP
EREV
NZ-ad
ETRIV
EPARAE
PwJT(3n) 4,5
ESP1
Type IV
PwCM (2n) 3, 4
PwJA (3n) i, h
PwJA (3n) c
PwCL (3n) b
PwCL (3n) a
PwK (3n) 2
PwP 1, 2
PwM
PoK
PoT
Pmi 2b, 4b
Pilok 279. 5, 2b
Pilok 279. 6
Pilok K 3, 5
Type I
Fhep
PwJH (2n) 1, e1, e4 del
PwP 3.5 del
PwJT (3n) 8 del
PwP c7 del
PwP c10 del
PwP c11 del
PwP c1 del
PwJH (2n) e6 del
PCERC3
Type II
"deleted
pseudogenes"
EHORT
SJ Anhui
SJ Hong Kong
10
(100)
(99)3
4
(79)4
9
(83)
12
(99)
7
(99)
11
(81)
(90)3
3
18
(100)
20
(100)
(76)1
20
(100)
12
(76) 4
9
(100)
24
(95) 1
24
(100)
37
83
(100)



Based on this information, it is therefore possible that the different types of ND1 sequences found in *Paragonimus* and *Echinostoma* occur in their different types of mitochondria. Both these genera are facultatively anaerobic, which is probably an adaptation to the variable availability of oxygen in the host tissues, at the site of establishment (a cyst within the lung and the lumen of the intestine, respectively) (Takamiya *et al.* 1994). It appears that *S. japonicum* and *F. hepatica* only have one form of ND1 (all schistosomes inhabit the aerobic vascular system and *F. hepatica* inhabits the anaerobic bile duct of their respective hosts). The predominance of a single mitochondrial type in schistosomes is further supported

by a population study of ND1 variants in 155 individuals, representing seven Chinese strains of *S. japonicum* (Bøgh *et al.* 1998). This showed that ND1 variants differed by less than 1.5% within and between populations, which is in marked contrast to the ND1 sequence variation detected within individuals of *Paragonimus* species, which have at least two well-represented mitochondrial types.

In the bivalve *Mytilus edulis*, in which mitochondria are inherited from both parents (Zouros *et al.* 1992), two mitochondrial genes, COIII and 16S rRNA, from the heteroplasmic mitochondria of members of the *Mytilus edulis* species complex, have been characterised by sequencing and phylogenetic analyses (Stewart *et al.* 1995, Rawson & Hilbish 1995). Both of these genes are represented by two distinct lineages in the three species studied to date, analogous to the distinct ND1 lineages identified here. Furthermore, the lineages of both mitochondrial genes have diverged prior to the divergence of the three species (*M. edulis*, *M. galloprovincialis* and *M. trossulus*) as is the case in ND1 lineages of *Echinostoma* and *Paragonimus* species. In addition, the paternal (P) mitochondrial lineage of both genes appears to be under less functional constraint than the same genes of the maternal (M) mitochondrial lineage, as the rate of mutational change appears to be greater in the P lineage though they do not contain frameshift mutations or premature termination codons.

8.5 Conclusions

These findings are of importance, because they shed light on the complex processes of evolution at the molecular level, which may impact on our interpretation of speciation events and enable us to discriminate between phylogenetically useful genes and pseudogenes that confound phylogenetic analyses due to paralogy. These findings therefore have implications for phylogenetic studies of other organisms, particularly trematodes, where mitochondrial sequences are used to differentiate between species and strains. Indeed, on the basis of these findings it seems that further studies of *Echinostoma* are also warranted and that phylogenetic inferences based on ND1 data (Morgan & Blair 1998a, 1998b) may need to be re-assessed in this group. Finally, the type II ND1 "pseudogene" / alternatively functional gene warrants further investigation to determine whether it is nuclear or mitochondrial. If it is mitochondrial we need to determine if the different mitochondrial types contain different ND1 variants.

CHAPTER 9

Clonal diversity in parthenogenetic triploid *Paragonimus westermani* (Trematoda: Digenea)

9.1 Abstract

Phylogenetic studies have not been able to differentiate between diploid and triploid *Paragonimus westermani* isolates from Northeast China (Chapters 5 and 8). I have used molecular techniques to test the hypothesis that the parthenogenetic triploid form of *P. westermani* has arisen only once. Sources of data for comparison were firstly, partial mitochondrial NADH Dehydrogenase subunit 1 (ND1) sequences, secondly, restriction fragment length polymorphisms (RFLPs) of ribosomal DNA (rDNA) digests flanking the internal transcribed spacers (ITS) and thirdly, "fingerprint" patterns observed when genomic digests were probed with $(ATT)_{10}$ and $(ATGT)_7$. In all cases there were distinct differences between triploid isolates from southwest Japan, northeast China and Korea. These findings are considered in the context of previous cytogenetic, allozyme, mitochondrial-RFLP and partial cytochrome c oxidase subunit I (COI) sequence studies and indicate that triploid lineages may have arisen independently on more than one occasion. The alternative explanation is that the triploids did have a single origin, but that different clonal lineages have undergone subsequent mutations. In view of recent studies on tetraploid *P. westermani*, the independent multiple origins of triploids is favoured.

9.2 Introduction

Paragonimus westermani is a typical hermaphroditic trematode (Digenea), which displays a three-host life-cycle. Sexually reproducing diploids as well as tetraploids and parthenogenetic triploid forms of the species have been described (Terasaki *et al.* 1995, 1996). Diploid strains are distributed throughout east and southeast Asia, whilst triploids are restricted to northeast and east China, Japan, Korea and Taiwan (reviewed in Blair, Xu & Agatsuma 1998). The tetraploid form has been found in Liaoning, China, in sympatry with the diploid and triploid forms and has been shown to contain alleles that are present in both diploid and triploid *P. westermani*

(Blair, Xu & Agatsuma 1998 and references therein). Not only do all three ploidy forms occur in sympatry, but they can co-occur in the same intermediate crustacean host genus in NE China and Korea.

Generally polyploidy results when closely related sympatric species hybridise. On the basis of allozyme and mitochondrial RFLP studies it has been suggested that the triploid form is clonal due to a single hybridisation event between *P. westermani* and some other strain or species of *Paragonimus* (Hirai *et al.* 1985, Hirai & Agatsuma 1991, Agatsuma *et al.* 1994). Cytogenetic studies have indicated that triploids have two of three chromosome homologues resembling diploid *P. westermani* chromosomes (by C-banding), while the third set of chromosomes are unique (Hirai *et al.* 1985), indicating the likely hybrid origin of triploids. However, the differences noted among homologous chromosomes may simply be due to a chromosomal polymorphism (Tan & Li 1990, Hirai & Agatsuma 1991, Terasaki *et al.* 1996). Furthermore, recent work by Blair *et al.* (1997) indicates the possible independent origins of triploid populations from varied ancestors, on the basis of partial mitochondrial gene (COI) sequences.

It is important to note that polymorphism in asexual species cannot be directly compared to that in sexual species as it may either be inflated by accumulated mutations since the loss of sexual reproduction or it may be reduced by periodic selective events (Birky 1996). However, tandemly repeated genes, such as the ribosomal RNA genes, appear to undergo sequence homogenisation at high frequency during both meiosis and mitosis, so that these sequences are expected to behave in the same way in both sexual and asexual species. Likewise, organelles such as the mitochondria, which are present in multiple copies in all cells of an organism, undergo random replication and random segregation which should ensure that identical sequences are maintained within an individual, so that the evolution of mitochondrial genes should not differ whether they are in sexual or asexual species (Birky 1996).

In this study additional molecular techniques have therefore been used to address the question of the single origin of triploid *P. westermani,* particularly RFLPs of the rRNA genes, partial sequences of a mitochondrial gene (ND1) and simple sequence repeat (SSR) fingerprints were compared between isolates.

9.3 Materials & Methods

9.3.1 DNA extraction

DNA extraction, PCR, cloning and sequencing procedures have been described in Materials and Methods (Chapter 3). Table 9.1 indicates ploidies and locations of worms used in this study.

Table 9.1 *Paragonimus westermani* isolates investigated in this study, indicating whether ITS1 rDNA RFLP, ND1 sequence and microsatellite fingerprint data were obtained for particular isolates. Number of individuals for which data were obtained is indicated. Generally ND1 and ITS1 data were obtained from the same individual (Table 3.1), but fingerprinting was done on an additional series of individuals.

	Number of Individuals		
Isolate	ITS1 rDNA RFLPs	ND1 sequences	Fingerprints
P. westermani (3n)			
Amakusa, Japan	1	1	3
Liaoning, NE China	1	1	5
Bogil Island, Korea	nd	1	4
P. westermani (2n)			
Mie, Japan	1	nd	nd
Minchin, E China	nd	1	nd
Malaysia	2	1	nd
Philippines	1	1	nd
P. iloktsuenensis, Japan	1	2	nd
P. ohirai, Japan	2	2	nd

Key: nd = no data obtained

9.3.2 Genomic digests and Southern blots

9.3.2a rDNA variants

Genomic DNA from diploid and triploid individuals of *Paragonimus westermani,* from *P. ohirai* and *P. iloktsuenensis* (*P. ohirai* species complex) and *P. miyazakii* was double digested with *Hind* III and *Xba* I to cut on either side of the rDNA coding regions flanking the ITS. *Hind* III should cut near the 3' end of the 18S rDNA (based on published trematode sequences), whilst *Xba* I cuts downstream of the ITS2. Individuals have multiple copies of ITS1 which differ in length, due to the presence of variable numbers of a 120bp repeat (van Herwerden, Blair & Agatsuma 1998b). Digestion should therefore excise a fragment consisting of ITS and flanking rDNA fragments, which is expected to vary in length within individuals, depending on the number of internal repeats embedded within ITS1 variants. All genomic digests were separated on 0.8% agarose gels, DNA was transferred to Hybond N membranes (Amersham), as per Southern (1975) and fixed by baking at 80 °C for 2 hrs. The probe used was an α-P^{32} dATP random-primed labelled PCR product amplified from the ITS1 of *P. westermani* using primers parfor and parrev (Table 3.2), which amplify the "post-repeat" fragment of ITS1 3' to the region of repeats (van Herwerden, Blair & Agatsuma 1998). Hybridisation was done at 50°C and the membranes were washed at 50°C to 0.5 X SSC, 0.1% SDS, exposed to a Phosphor Imager screen (Molecular Dynamics, California) overnight and scanned.

9.3.2b Fingerprinting of simple sequence repeats (SSRs)

Genomic DNA from 3-5 individuals from each of three isolates of triploid *P. westermani* (see Table 9.1) was digested with *Alu* I and *Hinf* I (as described by Turner *et al.* 1990). Southern blots were prepared as described above and probed with γ P^{32} end-labelled oligonucleotide probes $(ATGT)_7$ or $(ATT)_{10}$ (Ramachandran *et al.* 1997). Hybridisation and washes were performed at 48 °C and 40 °C respectively. Filters were washed in 0.5 X SSC, 0.1% SDS and 2 X SSC, 0.1% SDS respectively, before autoradiography (at -70°C, in the presence of 2 intensifying screens, overnight or for 5 days).

9.3.3 PCR, cloning and sequencing of ND1

ND1 sequences were obtained as described in Chapter 3. Multiple ND1 lineages were characterised, as described in Chapter 8, however, only functional "type I" ND1 sequences were used for this part of the study. Genbank numbers for sequences used here are AF051840 - AF051843, AF052242 - AF052243, AF063786, AF063787, AF063793 and AF063796.

9.4 Results

Table 9.1 summarises which molecular data were obtained for particular isolates in this study.

9.4.1 ITS - rDNA

RFLPs for the Japanese triploid from Amakusa and the Chinese triploid from Liaoning differ in size and number of bands from each other and from all diploid isolates of *P. westermani,* each of which was also unique (Fig. 9.1, lanes 1-6). *P. miyazakii* and members of the *P. ohirai* complex each exhibited single bands, which differed between but not within the species (Fig. 9.1 lanes 7-10).

9.4.2 ND1 sequences

Not all ND1 sequences obtained were used in this study, as multiple ND1 lineages existed within individuals (data presented in Chapter 8). The sequences compared here were all regarded as orthologous. Diploid isolates from China (Minchin), the Philippines (Sorsogon) & Malaysia (Ulu Langat) were also characterised for comparison (Tables 9.1 & 9.2). As shown in Table 9.2, no two isolates shared the same ND1 sequence. The Korean triploid is most similar in sequence to the Chinese diploid with only 1 base difference. Clearly differences between isolates from NE and SE Asia are considerable. ND1 sequences (both nucleotide & amino acid) of the SE Asian isolates from the Philippines and Malaysia also differ substantially from each other, as would be expected on the basis of COI and ITS2 data (Blair *et al.* 1997).

9.4.3 Fingerprinting data

Microsatellite fingerprint data is compared between 12 individuals of three triploid isolates, using two simple sequence repeats (SSRs) as probes of genomic material that has been digested, separately, with two frequent cutting restriction enzymes, *Alu* I & *Hinf* I (Table 9.3, Figure 9.2). Southern blots of genomic digests were probed with the SSRs $(ATGT)_7$ and $(ATT)_{10}$. There is no variation detected with either of the tri- or tetra nucleotides, amongst Chinese triploids from Liaoning. However, some Korean and all Japanese triploids did differ from the Chinese triploids, when *Alu* I digests were probed. When *Hinf* I digests were probed, all Korean and Chinese isolates appeared identical, with the Japanese triploids being distinctly different. Furthermore, it is important to note that there is variation in patterns of Korean and Japanese individuals within triploid populations, when *Alu* I digests were investigated.

Table 9.2 Pairwise differences (total length 473 bp) among ND1 sequences above the diagonal and number of amino acid changes below the diagonal for diploid and triploid isolates of *P. westermani* from East Asia.

Isolate	1	2	3	4	5	6
1. E China (2n), Minchin	-	4	4	1	56	60
2. NE China (3n), Liaoning	3	-	4	4	54	57
3. Japan (3n), Amakusa	3	0	-	5	54	58
4. Korea (3n), Bogil Island	0	3	2	-	56	60
5. Philippine(2n), Sorsogon	8	9	7	7	-	29
6. Malaysian (2n), Ulu Langat	9	8	8	7	9	-

Figure 9.1 (overleaf)
Autoradiograph of Southern blot of restriction digested (*Xba* I & *Hind* III) individuals of the *P. westermani* species complex, the *P. ohirai* species complex (*P. ohirai* & *P. iloktsuenensis*) & *P. miyazakii*, probed with a P^{32} labelled ITS1 fragment. Sizes indicated are for the sizes of ITS1 plus flanking rDNA fragments excised by digestion. The specimens shown are (1 - 6): *P. westermani* (1) Malaysia, Kuala Lumpur (2n) ; (2) Sorsogon, Philippines (2n); (3) Malaysia, Ulu Langat (2n); (4) Japan, Mie (2n); (5) Japan, Amakusa (3n); (6) NE China, Liaoning (3n); (7,8) *P. ohirai*, Japan; (9) *P. iloktsuenensis*, Japan & (10) *P. miyazakii*, Japan.

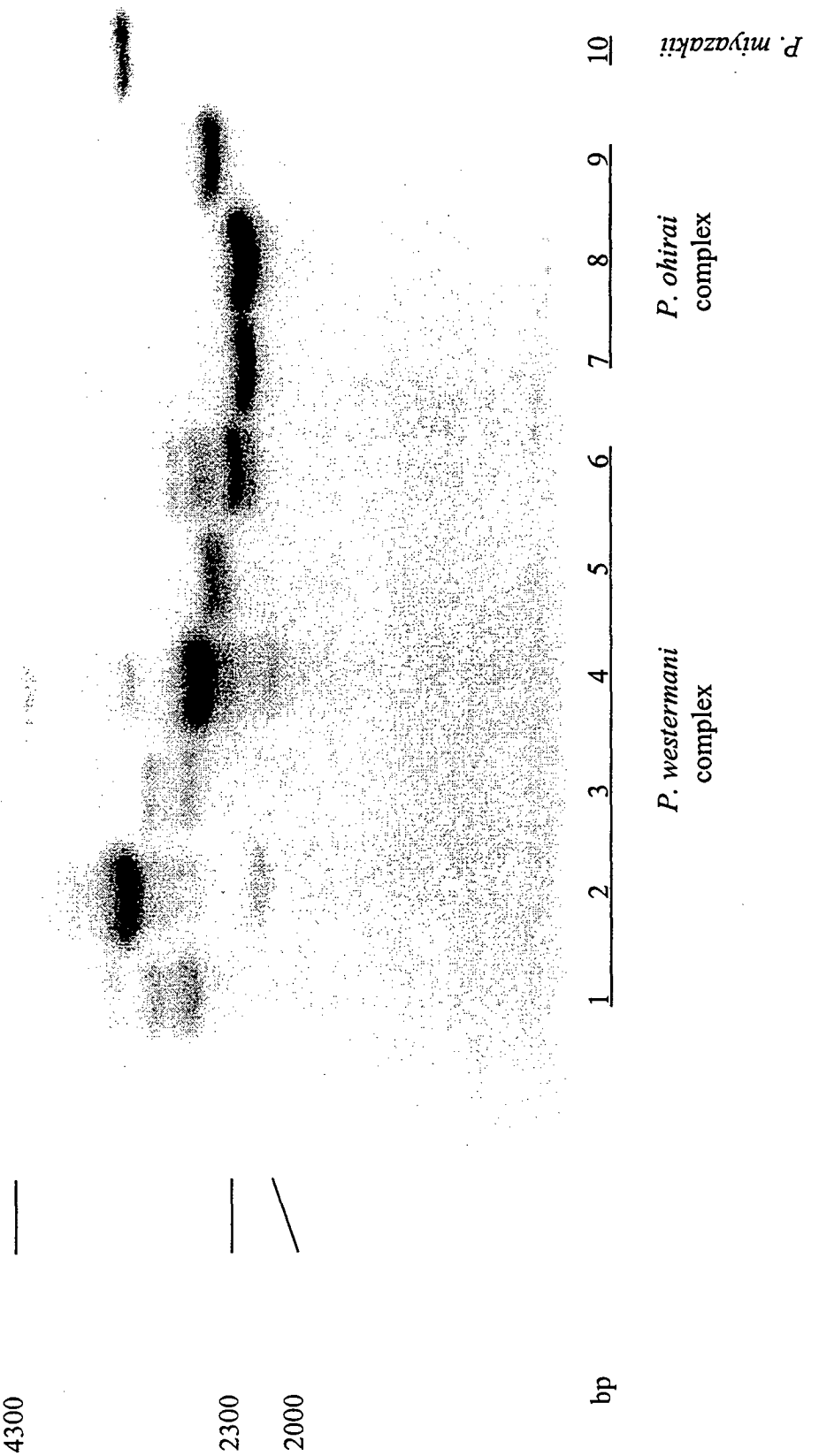

4300
2300
2000
bp
1
2
3
4
5
6
7
8
9
10
P. westermani complex
P. ohirai complex
P. miyazakii

Table 9.3 Comparison of three triploid *P. westermani* isolates, using two microsatellite probes to determine differences between fingerprints of individuals. Comigrating bands for each treatment are denoted by the same letter of the alphabet.

Isolate	Number of individuals	$(ATGT)_7$ *Alu* I	$(ATGT)_7$ *Hinf* I	$(ATT)_{10}$ *Alu* I	$(ATT)_{10}$ *Hinf* I
NE China (Liaoning)	5	abcdefgh	klmno	pq	st
Korea (Bogil Island)	1	abcdefgh	klmno	pqr	st
	3	abc-ef--ij	klmno	pqr	st
Japan (Amakusa)	2	abcdefgh	lm	pq	tu
	1	ef	lm	p	tu

9.5 Discussion

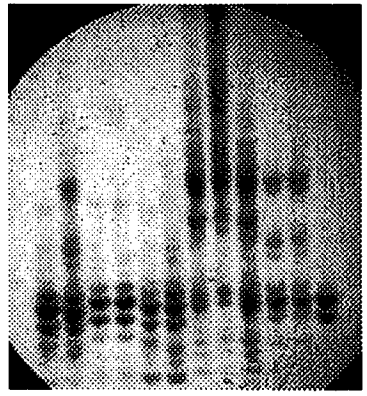
Processes of molecular evolution of both tandemly repeated rDNA and mitochondrial genes are expected to produce effectively "homogenised" sequences within individuals, irrespective of whether reproduction is sexual (diploids) or parthenogenetic (triploids) (Birky 1996). In this study I have shown, using three different molecular techniques, that triploid *P. westermani* are not genetically identical.

9.5.1 rDNA RFLPs

Polymorphism detected in the nuclear rDNA gene fragments show that all diploid and triploid *P. westermani* individuals investigated displayed multiple bands that differ in both number and size among isolates, regardless of ploidy. In contrast, *P. ohirai* isolates from 2 Japanese locations were identical, but have a different RFLP fragment to *P. iloktsuenensis*. This contrasts with previous findings (Blair, Agatsuma & Watanobe 1997), in which *P. iloktsuenensis* and *P. ohirai* were indistinguishable by ITS2 and CO I sequences.

Figure 9.2 (overleaf)
Fingerprints of *P. westermani* triploids from NE China (Liaoning), n=5, Korea (Bogil Island), n=4 and Japan (Amakusa), n=3 based on a) $(ATGT)_7$ probed *Alu* I fingerprints; b) $(ATGT)_7$ probed *Hinf* I fingerprints; c) $(ATT)_{10}$ probed *Alu* I fingerprints; d) $(ATT)_{10}$ probed *Hinf* I fingerprints

a



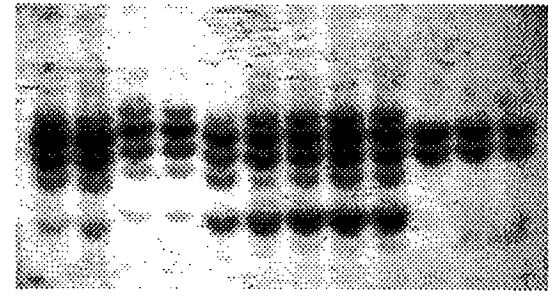
1 2 3 4 5 1 2 3 4 1 2 3

China Korea Japan

***Alu* I**

$(ATGT)_7$

b

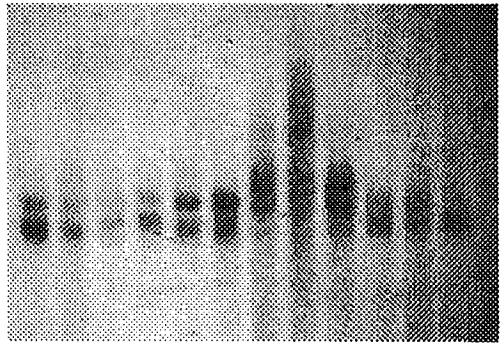


1 2 3 4 5 1 2 3 4 1 2 3

China Korea Japan

***Hinf* I**

c

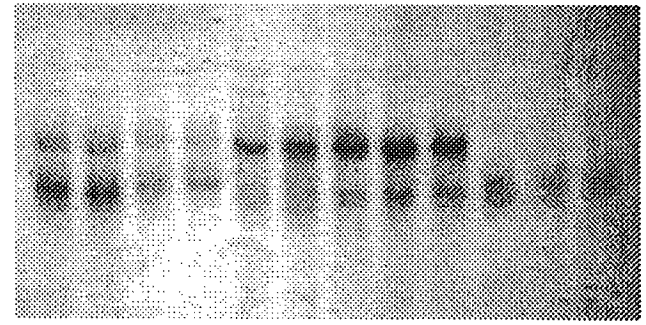


1 2 3 4 5 1 2 3 4 1 2 3

China Korea Japan

$(ATT)_{10}$

d



1 2 3 4 5 1 2 3 4 1 2 3

China Korea Japan

9.5.2 ND1 sequences

Each diploid and triploid isolate from NE China, Korea and Japan had a unique ND1 sequence. However, differences among these were small relative to differences between E Asian and SE Asian isolates.

9.5.3 Fingerprinting data

Triploids from NE China, Korea and Japan differed in one or more of the SSR markers used in this study. In addition, variation among Japanese and among Korean individuals was noted.

9.6 Conclusions

It appears (on the basis of rDNA RFLPs and ND1 sequences) that a recent single clonal origin of the parthenogenetic triploid *P. westermani* is unlikely, given the genetic variation observed. This variation is no less among triploid isolates than it is among diploid isolates. Furthermore, SSR fingerprints showed that there is genetic variation among triploid individuals within and between populations.

A possible mechanism has been proposed for the origin of triploids. Triploid *P. westermani* cannot produce normal sperm (Fujino & Ishii 1982) but can receive sperm from diploid individuals when sharing a cyst (Miyazaki *et al.* 1981, Terasaki *et al.* 1996). Although the triploid eggs do not undergo reduction division (Sakaguchi & Tada 1980), fertilization by a haploid sperm may theoretically produce tetraploid offspring, which may in turn contribute diploid gametes for fertilization by haploid gametes, producing parthenogenetic triploids. The recent discovery of tetraploid individuals in the wild (Terasaki *et al.*1989, 1995) further supports the possibility that triploids may be derived from relatively rare crosses between tetraploid and diploid individuals. Furthermore, the tetraploids are not identical to one another, but have independent origins (Agatsuma *et al.* 1992).

CHAPTER 10

General Discussion

10.1 Overview

10.1.1 Aims

The aims of this study have been to evaluate the use of molecular techniques to distinguish among and infer phylogenies of the Asian trematode genera *Schistosoma* and *Paragonimus.* These Asian trematodes were specifically targeted in this study, because the species level taxonomy, transmission and distribution of East Asian species were already known. Secondly, robust phylogenies of these genera can be evaluated in the context of an existing biogeographic hypothesis of coevolution between transmitting snail hosts and their parasites. Thirdly, I wanted to evaluate the hypothesis that there was a single clonal origin of parthenogenetic triploid *P. westermani,* using several different molecular techniques. The fourth aim of the study arose from the observations of unexpected intra-individual variation of both ITS1 and ND1 sequences. Specifically, I wished to understand why both regions sequenced displayed intra-individual variation within the *Paragonimus* species groups, despite the fact that both are expected to be homogeneous within individuals and species.

10.1.2 Phylogenetic hypotheses

ITS1 and ND1 sequences were selected for the phylogenetic inference, as both had been used successfully in the past for the inference of phylogenetic relationships at and below the species level in trematode genera (Morgan & Blair 1998a, 1998b, Bowles & McManus 1993). Phylogenetic inferences made from ITS1 sequence data of members of the *S. japonicum, P. westermani* and *P. ohirai* species complexes not only supported findings based on previous molecular phylogenetic studies (ITS2 and COI sequences), but in addition differentiated between species within the groups. For example, *P. ohirai* and *P. iloktsuenensis* had not previously been differentiated by phylogenetic studies. This was achieved, despite the presence of repeat sequences embedded within ITS1, by using "post-repeat" ITS1 sequences only for phylogenetic tree construction in both *Schistosoma* and *Paragonimus.* ITS1 "post-repeat" sequence data from the *P. ohirai* species complex lacked the intra-individual variation that was a

complicating factor for phylogenetic inference in the *P. westermani* group, despite excluding the repeat sequences from phylogenetic analysis.

Similarly, ND1 sequence variants in *Paragonimus* species also had the potential to confound phylogenetic inferences by producing a gene tree of the variant ND1 lineages, rather than a species tree. However, the presumed functional ND1 variant (Type I) could be identified and did resolve the lineages within both the *P. westermani* and *P. ohirai* species complexes. Diploid and triploid *P. westermani* lineages from northeast Asia could however, not be distinguished from each other by phylogenetic inference with either COI and ITS2 in previous studies or ITS1 and ND1 sequence data in this study, though they are not identical. This indicates that there is no genetic substructure among the Northeast Asian *P. westermani* populations and importantly, that the triploids cannot be phylogenetically distinguished from the diploids. Interestingly, it was also possible to differentiate between those members of the *P. ohirai* species complex investigated with ND1 (BCL of 76), supporting the ITS1 sequence and ITS-rDNA RFLP data, thereby improving on the phylogenetic resolution of ITS2 and COI sequence data. Therefore, three distinct molecular techniques have collectively differentiated members of the *P. ohirai* species complex investigated here, indicating that they are not likely to be synonymous, as was previously thought. Having obtained phylogenetic inferences for the *S. japonicum* and *P. ohirai* species complexes, I could then consider the phylogenetic data in the context of Davis' biogeographic hypothesis.

10.1.3 Biogeographic hypothesis, coevolution and host switching

Davis (1980, 1992) proposed that both *Schistosoma* and *Paragonimus* species that are transmitted by the pomatiopsid snails, share similar patterns of biogeography and dispersal throughout east and southeast Asia, and that both genera may therefore have coevolved with their rapidly evolving snail hosts. The snails evolved down the Southeast Asian rivers, as they flowed southwards and eastwards, subsequent to the Himalayan orogeny.

Phylogenetic inferences for members of the species groups investigated here indicate that coevolution of the Asian *Schistosoma japonicum* and *P. westermani* species complexes with their respective transmitting snail hosts has probably occurred,

as proposed by Davis. All molecular data to date, including data presented in this study, indicate that members of the species complexes which are geographically distinct (NE vs SE Asia), and are transmitted by different snail hosts, are phylogenetically differentiated. The presence of *S. japonicum* in the Philippines, where it is transmitted by *Oncomelania hupensis*, as are all *S. japonicum* strains, is however anomalous. We may expect that the Philippine *S. japonicum* strain and its snail host may have evolved into distinct species from the NE Asian strains according to the biogeographic hypothesis. This anomaly may indicate that the parasite was recently introduced to the Philippines with its transmitting snail host, by human activities.

Members of the *P. westermani* species complex also appear to have co-evolved with their thiarid or pleurocerid snail hosts, as predicted by Davis, with NE and SE Asian strains being phylogenetically quite distinct. The ceritheacean snails transmitting *P. westermani* are however not of the same superfamily as the rissoacean snails transmitting the Asian *Schistosoma* and members of the *P. ohirai* species complex. Furthermore, there is no evidence indicating whether these two snail superfamilies have a shared biogeography, as the ceritheacean snails may have been present in Asia prior to the arrival of the schistosome transmitting snails which were introduced to Asia via the Indian plate. I suggest that these ceritheacean snails may have been present in Asia prior to the arrival of the rissoacean snails, which had rafted across on the Indian plate (Davis 1980), because the ceritheacean snails do not transmit any schistosomes whereas the rissoacean snails transmit both schistosomes and members of the *P. ohirai* species complex. Should this be true, I further suggest that there may have been a pre-existing lineage of *Paragonimus* in Asia (precursors of the *P. westermani* complex), which was subsequently joined by another distinct lineage from India (precursors of the *P. ohirai* complex). Clearly, the biogeography of the pre-existing group of *Paragonimus westermani* would be distinct from the rissoacean transmitted *Paragonimus ohirai* species if this is the case. Members of the *P. ohirai* species complex, may on the other hand have the same biogeographic history as the Asian schistosomes, as they are also transmitted by rissoacean snails. However, members of this complex appear to have switched hosts rather than to have coevolved with their snail hosts, because they are transmitted naturally and / or experimentally by snails from very different rissoacean families (Assimineidae and Pomatiopsidae) (see Chapter

2 section 2.2). Although members of the *P. ohirai* species complex were indistinguishable by COI and ITS2 sequences, ITS1 and ND1 sequences did differentiate between the two species, though they were phylogenetically very close.

Following Davis' logic, the *P. ohirai* species complex, which includes three morphologically very similar species may be expected to consist of at least two species, on the basis of the snail hosts utilised: one including the Assimineidae-transmitted *P. ohirai* and *P. iloktsuenensis* and another for *P. sadoensis*, which is transmitted by the Pomatiopsidae (*Oncomelania minima*). This is however not the case, as discussed above and suggests that the Davis Hypothesis of coevolution does not adequately explain the observed transmission of members of the *Paragonimus ohirai* complex. Indeed, host-switching may better explain the broad host-specificity of species such as *Paragonimus ohirai.*

Host-switching may be either prevented or facilitated by the presence of other trematodes inhabiting the same snail species, for example, *S. japonicum* is disadvantaged by the presence of *P. ohirai* in *Oncomelania nosophora* (Hata *et al.* 1988). Conversely, *P. westermani* co-occurs with other trematodes in the *Semisulcospira* host more often than predicted by chance alone (Tomimura *et al.* 1989 in Blair, Xu & Agatsuma 1998). The application of modern systematics to the *Paragonimus* transmitting snails and their parasites should enable us to determine whether the host-parasite specificity is indeed so broad. If it is, this would indicate that host-switching has masked coevolution, which might explain the broad host specificity of some *Paragonimus* species. If this has happened frequently in nature, any pattern of coevolution such as that suggested by Davis, will be obscured.

10.1.4 The proposed single clonal origin of triploid *P. westermani*

The theory of the single clonal origin of parthenogenetic triploid *P. westermani* is shown to be unlikely, as the triploid isolates from different geographic locations are not genetically identical when ITS-rDNA RFLPs, ND1 sequences and SSR fingerprints are investigated. Additionally, isolates from the same geographic locations are not always genetically identical (particularly note Korean and Japanese isolates), which suggests that the triploid lineages may have had multiple independent origins. Alternatively, these parthenogens may have arisen by a single event, but have since

diverged due to a high mutation rate, a possibility which cannot be excluded. However, the absence of any phylogenetic differentiation between diploid and triploid lineages with all molecular data used to date (ITS1 and 2, COI and ND1) suggests that there is the potential for occasional gene flow between them, via diploid / tetraploid intermediaries, as described in Chapter 9.

In addition to the main objectives of this study, all of which have been addressed above, some interesting issues of broader application have been raised by this work, due to the extensive amount of intra-individual variation observed among both ND1 and ITS1 sequences. These issues of molecular evolution are discussed below (section 10.2), because I believe that further research is warranted to investigate these aspects.

10.2 Molecular evolution

Two broad future research areas that have been raised by this study are 1) ITS1 structure and function, particularly relating to processes of concerted evolution in some trematodes; 2) Mitochondrial pseudogenes and their possible presence in either or both the nuclear and the alternative mitochondrial genomes, in some trematodes.

10.2.1a ITS1 variation in trematodes and implications for processes of concerted evolution

The tandemly repeated rDNA genes have formed the basis of much of the work that has been done on processes of concerted evolution to date (Dover 1982, 1986, Dover & Tautz 1986). However, the evolutionary patterns of concerted evolution have not been studied much in natural or experimental populations (Elder & Turner 1995). The trematodes would form a particularly interesting system for modelling processes of concerted evolution in natural or experimental populations, as sister species in several genera (*Schistosoma*, *Paragonimus* and *Dolichosaccus*) appear to experience different rates of ITS1 sequence homogenisation. Some species or species groups display intra-individual variation (*S. japonicum*, *P. westermani* and *D. helocirrus*), whereas sister species or species groups in each of these genera either lack or display reduced intra-individual variation in the ITS1 (*S. mansoni*, *P. ohirai* and *D. symmetrus*). It would be interesting to investigate some possible underlying causes of these differences in intra-

individual variation. There are two issues that can be investigated in this context. Firstly, to determine the rate of recombination as indicated by the frequency of chiasma formation (FX), in species of the above-mentioned species pairs for which FX has not yet been determined. Secondly, the sister species of each of the three pairs could be investigated for the presence of multiple Nucleolar Organiser Regions (NORs), as the only species investigated to date, *Paragonimus ohirai* and *Schistosoma mansoni* each has a single NOR. If present in sister species, multiple NORs may be another mechanism whereby sequence homogenisation between clusters of ribosomal genes is reduced.

10.2.1b Putative regulation of rDNA transcription by enhancer motifs present in the ITS1 of some trematodes

In vivo transcription and quantitation of the ITS1 transcripts from rRNA variants of individual trematodes with and without multiple repeats embedded in the ITS1, should indicate if transcription is indeed enhanced by the presence of additional repeats. This can be done for the three species of the *S. japonicum* species complex, as well as for *Paragonimus westermani* and *P. ohirai* , as outlined in Chapter 6.

10.2.2a ND1 pseudogenes and presumed numt DNA in trematodes

The detection of multiple lineages of ND1 in Asian *Paragonimus* and some Australian *Echinostoma* species has prompted the need for further studies in order to determine if ND1 variants have been integrated into the nuclear genome, perhaps at different times, as has been determined for mitochondrial genes in many taxa. This can be done by obtaining fresh samples of parasites, so that mitochondrial DNA can be purified. Further investigations can then be pursued as outlined in Chapter 8.

10.2.2b ND1 pseudogenes and heteroplasmy in trematodes

Different types of mitochondria have been distinguished in tissues of trematode species by stains that react with cytochrome c oxidase (aerobic metabolism) (Takamiya *et al.* 1994). It is possible that the different mitochondria have different forms of ND1, a gene which codes for a component of the electron transport chain in the mitochondrial membrane. As trematodes are facultatively anaerobic, different

mitochondrial genes may be active in the aerobic and anaerobic mitochondrial types. ND1 pseudogenes or an alternative functional ND1 may therefore be present in these distinct mitochondrial types. This may be investigated, as outlined in Chapter 8.

10.3 Concluding comments

In this study I have demonstrated that ITS1 and ND1 should be used with caution for the inference of highly resolved phylogenetic structure in some trematodes, contrary to the clear and congruent phylogenetic signals obtained from COI and ITS2 sequence data in previous studies. Unexpectedly and for different reasons both ND1 and ITS1 produced potentially confounding data for phylogenetic inference, particularly in the *P. westermani* species complex. The variation detected in both ND1 and ITS1 raised some important and fascinating questions about these genes in trematodes. These markers, together with SSR fingerprints, were particularly useful to demonstrate genetic differences between triploid isolates previously assumed to be of single origin. I have also shown in this study that much further work is required to frame detailed and particular questions about the coevolution of these trematodes and their snail hosts.

Literature Cited

Adams, C.G. (1981) An outline of tertiary palaeogeography pp. 221-235. In *The evolving earth - chance, change and challenge*. British Museum (Natural History) P.H. Greenwood & L.R.M Cocks (eds.) Cambridge University Press, Cambridge.

Adlard, R. D., Barker, S. C., Blair, D. & Cribb, T. H. (1993) Comparison of the second internal transcribed spacer (ribosomal DNA) from populations and species of Fasciolidae (Digenea). *International Journal for Parasitology* **23**: 423-425.

Agatsuma, T. & Habe, S. (1985) Electrophoretic studies on enzymes of diploid and triploid *Paragonimus westermani*. *Parasitology* **91**: 489-497.

Agatsuma, T., Ho, L., Jian, H., Habe, S., Terasaki, K., Shibahara, T., Sugiyama, H., Kim, D. & Kawashima, K. (1992) Electrophoretic evidence of a hybrid origin for tetraploid *Paragonimus westermani* discovered in Northeast China. *Parasitology Research* **78**: 537-538.

Agatsuma, T., Yang, L., Kim, D. & Yonekawa, H. (1994) Mitochondrial DNA differentiation of Japanese diploid and triploid *Paragonimus westermani*. *Journal of Helminthology* **68**: 7-11.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1989) *Molecular biology of the cell*. 2nd ed. Garland Publishing, New York. 1219 pages.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403 - 410.

Arnheim, N. (1983) Concerted evolution of multigene families. In: *Evolution of genes and proteins*. (Eds: Nei, M. and Koehn, R. K.) Sinauer Associates, Massachusetts, 38-61.

Arnold, M.L. (1997) *Natural hybridization and evolution.* Pp 155 – 181. Oxford series in Ecology and Evolution, Oxford University Press, Oxford.

Avise, J.C. Mitochondrial DNA and the evolutionary genetics of higher animals. (1986) *Philosophical Transactions Royal Society London series B* **312**: 325 - 342.

Avise, J. C., Arnold, J., Ball, M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. and Saunders, N. C. (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**: 489-522.

Avise, J.C. (1994) *Molecular markers, natural history and evolution.* Chapman & Hall, New York. 511 pages.

Barker, S.C. & Blair, D. (1996) Molecular phylogeny of *Schistosoma* species supports traditional groupings within the genus. *Journal of Parasitology* **82**: 292 - 298.

Barton, D. (1994) Three species of the genus *Dolichosaccus* Johnston, 1912 (Digenea:Telorchiidae) from the introduced toad *Bufo marinus* (Amphibia: Bufonidae) in Australia, with the erection of Meditypus n. subg. *Systematic Parasitology* **29**: 121-131.

Behm, C.A. (1991) Fumarate reductase and the evolution of electron transport systems, pp 88-108 In *Metazoan Life without oxygen*, C. Bryant ed. Chapman & Hall, London.

Birky, C.W. Jr. (1996) Heterozygosity, heteromorphy and phylogenetic trees in asexual eukaryotes. *Genetics* **144**: 427-437.

Blair, D., Campos, A., Cummings, M. P. and Laclette, J. P. (1996) Evolutionary biology of parasitic platyhelminths: the role of molecular phylogenetics. *Parasitology Today.* **12**: 66-71.

Blair, D., Agatsuma, T. & Watanobe, T. (1997) Molecular evidence for synonymy of three species of *Paragonimus*, *P. ohirai* Miyazaki, 1939, *P. iloktsuenensis* Chen, 1940 and *P. sadoensis* Miyazaki *et al.*, 1968. *Journal of Helminthology* **71**: 305-310.

Blair, D., Agatsuma, T., Watanobe, T., Okamoto, M. & Ito, A. (1997a) Geographical genetic structure within the human lung fluke, *Paragonimus westermani,* detected from DNA sequences. *Parasitology* **115**: 411-417.

Blair, D., van Herwerden, L., Hirai, H., Taguchi, T., Habe, S., Hirata, M., Lai, K., Upatham, S. & Agatsuma, T. (1997b) Relationships between *Schistosoma malayensis* and other Asian schistosomes deduced from DNA sequences. *Molecular and Biochemical Parasitology* **85**: 259 - 263.

Blair, D., Xu, Z-B. & Agatsuma, T. (1998) Paragonimiasis and the genus *Paragonimus*. (in press) *Advances in Parasitology*.

Blair, D., Bray, R.A. & Barker, S.C. (1998) Molecules and morphology in phylogenetic studies of the hemiuroidea (Digenea: Trematoda: Platyhelminthes) *Molecular Phylogenetics & Evolution* **9**: 15-25.

Bϕgh, H.O., Zhu, X.Q., Qian, B-Z. & Gasser, R.B. (in press) Scanning for nucleotide variations in mitochondrial DNA fragments of *Schistosoma japonicum* by single strand conformation polymorphism. *Parasitology*.

Bowles, J., Blair, D. and McManus, D. P. (1992) Genetic variants within the genus *Echinococcus* identified by mitochondrial DNA sequencing. *Molecular and Biochemical Parasitology* **54**: 165-174.

Bowles, J. & McManus, D.P. (1993). Rapid discrimination of *Echinococcus* species and strains using a PCR-based method. *Molecular and Biochemical Parasitology* **57**: 231 - 239.

Bowles, J., Blair, D. and McManus, D. P. (1994) Molecular genetic characterisation of the cervid strain ("northern form") of *Echinococcus granulosus*. *Parasitology* **109**: 215-221.

Bowles, J., Blair, D. and McManus, D. P. (1995a) A molecular phylogeny of the human schistosomes. *Molecular Phylogenetics & Evolution* **4**: 103 - 109.

Bowles, J., Blair, D. and McManus, D. P. (1995b) A molecular phylogeny of the genus *Echinococcus*. *Parasitology* **110**: 317 - 328.

Bowles, J. & McManus D.P. (1993) NADH Dehydrogenase 1 gene sequences compared for species and strains of the genus *Echinococcus*. *International Journal for Parasitology* **23**: 969-972.

Brower, A.V.Z. & De Salle, R. (1994) Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. *Annals of the Entomological Society of America* **87**: 702-716.

Brown, J.R., Daar, I.O., Krug, J.R. & Maquat, L.E. (1985) Characterisation of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. *Molecular & Cellular Biology* **5**: 1694-1706.

Brown, W.M. (1985) The mitochondrial genome of animals. pp. 95-130. In *Molecular evolutionary genetics*. R.J. MacIntyre, ed. Plenum Press, New York.

Busby S.J. & Reeder R.H. (1993) Spacer sequences regulate transcription of ribosomal gene plasmids injected into *Xenopus* embryos. *Cell* **34**: 989 - 96.

Cabot E.L. & Beckenbach A.T. (1989) Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. *Computer Applications in the Biosciences* **5**: 233 - 4.

Cabot, E.L. (1997) Eyeball Sequence Editor v 3.1s. University of Rochester, New York.

Clarke, A.G. (1997) Neutral behaviour of shared polymorphism. *Proceedings of the National Academy of Science USA* **94**: 7730 - 7734.

Clarke, A.G. & Kao, T-h. (1991) Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proceedings of the National Academy of Science USA* **88**: 9823 - 9827.

Coleman, A. W. and Mai, J. C. (1997) Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting relatedness. *Journal of Molecular Evolution.* **45:** 168-177.

Davis, G.M. (1979) The origin and evolution of the gastropod family Pomatiopsidae, with emphasis on the Mekong River Triculinae. *Monograph of the Academy of Natural Sciences of Philadelphia* **20**: 1-120.

Davis, G.M. (1980) Snail hosts of Asian *Schistosoma* infecting man: evolution and coevolution. *Malacological Review* **suppl. 2**: 195-238.

Davis, G.M. (1992) Evolution of prosobranch snails transmitting Asian *Schistosoma*: co-evolution with *Schistosoma*: a review. *Progress in Clinical Parasitology* **3**: 145-204.

Davis, G.M. (1994) Molecular genetics and taxonomic discrimination. *The Nautilus* **supplement 2**: 3-23.

Davis, G.M., Chen, C-E., Yu S.H. & Li, L. (1994a) Molecular genetic and anatomical relationships among Pomatiopsid (Gastropoda: Prosobranchia) genera from southern China. *Proceedings of the Academy of natural sciences of Philadelphia* **145**: 191-207.

Davis, G.M., Chen, C-E., Kang, Z-B. & Liu, Y-Y. (1994b) Snail hosts of *Paragonimus* in Asia and the Americas. *Biomedical and Environmental Sciences* 7: 369-382.

Davis, G.M., Spolsky, C. & Zhang, Y. (1995) Malacology, biotechnology and snail-borne diseases. *Acta Medica Philippina* **31**: 45-60.

Davis, G.M., Zhang, Y, Hua, G.Y. & Spolsky, C. (1995) Population genetics and systematic status of *Oncomelania hupensis* (Gastropoda: Pomatiopsidae) throughout China. *Malacologia* **37**: 133-156.

Després, L., Imbert-Establet, D., Combes, C. & Bonhomme, F. (1992) Molecular evidence linking hominid evolution to recent radiation of schistosomes (Platyhelminthes: Trematoda). *Molecular Phylogenetic Evolution* **1**: 295 - 304.

Devereux, J., Haeberli, P. & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**: 387-395.

Dietrich P., Soares M.B., Affonso M.H.T. & Floeter-Winter L.M. (1993) The *Trypanosoma cruzi* ribosomal RNA-encoding gene: analysis of promoter and upstream intergenic spacer. *Gene* **125**: 103 - 107.

dos Reis M.G., Davis R.E., Singh H., Skelly P.J. & Shoemaker C.B. (1993) Characterisation of the *Schistosoma mansoni* gene encoding the glycolytic enzyme, triosephosphate isomerase. *Molecular & Biochemical Parasitology* **59**: 235-242.

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

Dover, G.A. (1986) The spread and success of non-Darwinian novelties. In *Evolutionary processes and theory* (eds. S. Karlin & E. Nevo) Academic Press, New York & London.

Dover, G.A. & Tautz, D. (1986) Conservation and divergence in multigene families: alternatives to selection and drift. *Philosophical Transactions of the Royal Society London series B* **312**: 275-289.

Drew, A.C., Brindley, P.J., Lewis F.A., Liang Y. & Minchella, D.J. (1998) Tandemly repeated genomic sequence demonstrates inter- and intra-strain genetic variation in *Schistosoma japonicum*. *Tropical Medicine and International Health* **3**: 373-380.

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

Felsenstein, J. (1983) Parsimony in systematics: biological and statistical issues. *Annual Review Ecological Systematics* **14**: 313-333.

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

Felsenstein, J. (1989) PHYLIP - Phylogeny inference package (Version 3.2). *Cladistics* **5**: 164-166.

Fritz, G.N., Conn, J., Cockburn, A. & Seawright, J. (1994) Sequence analysis of the ribosomal DNA Internal Transcribed Spacer 2 from populations of *Anopheles nuneztovari* (Diptera: culicidae). *Molecular Biology & Evolution* **11**: 406 - 416.

Fujino, T. & Ishi, Y. (1982) Ultrastructural studies on spermatogenesis in a parthenogenetic type of *Paragonimus westermani* (Kerbert 1878) proposed as *P. pulmonalis* (Baelz 1880). *Journal of Parasitology* **68**: 433-441.

Fujino, T., Fried, B. & Takamiya, S. (1995) Cytochemical localisation of cytochrome c oxidase activity in mitochondria in the tegument and tegumental and parenchymal cells of the trematodes *Echinostoma trivolvis, Zygocotyle lunata, Schistosoma mansoni, Fasciola gigantica and Paragonimus ohirai*. *Journal of Helminthology* **69**: 195 - 201.

Galtier, N. & Gouy, M. (1995) Inferring phylogenies from DNA sequences of unequal base compositions. *Proceedings of the National Academy of Science USA* **92**: 11317 - 11321.

Garey, J.R. & Wolstenholme, D.R. (1989) Platyhelminth mitochondrial DNA: evidence for early evolution origin of a tRNA ^{ser} AGN that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. *Journal of Molecular Evolution* **28**: 374 - 387.

Gasser, R., Nansen, P. & Guldberg, P. (1996) Fingerprinting sequence variation in ribosomal DNA of parasites by DGGE. *Molecular and Cellular Probes* **10**: 99-105.

Gasser, R.B., Chilton, N.B., Hoste, H. & Beveridge, I. (1993) Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. *Nucleic Acids Research* **21:** 2525-2526.

Gentry, A.W. & Sutcliffe, A.J. (1981) Pleistocene geography and mammal faunas pp. 237-251 in *The evolving earth - chance, change and challenge*. British Museum (Natural History) P.H. Greenwood & L.R.M Cocks (eds.) Cambridge University Press, Cambridge.

Gonzalez, I.L., Chambers, C., Gorski, J.L., Stambolian, D., Schmickel, R.D. & Sylvester, J.E. (1990) Sequence and structure correlation of human ribosomal transcribed spacers. *Journal Molecular Biology* **212**: 27 - 35.

Hedges, S.B. & Maxson, L.R. (1997) Complementary uses of molecules and morphology: a reply to Lee. *Molecular Phylogenetic Evolution* **8**: 445.

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

Hillis, D. M. and Dixon, M. T. (1991) Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quarterly Review of Biology.* **66**, 411-446.

Hillis, D.M., Moritz, C., Porter, C.A. & Baker, R.J. (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**: 308 - 310.

Hillis, D.M., Bull, J.J., White, M.E., Badgett, M.R. & Molineux, I.J. (1992) Experimental phylogenetics: generation of a known phylogeny. *Science* **255**: 589-592.

Hillis, D.M. & Heulsenbeck (1992) Signal, noise and reliability in molecular evolutionary analyses. *Journal of Heredity* **83**: 189-195.

Hillis, D.M. & Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**: 182 - 192.

HillisD.M., Huelsenbeck, J.P. & Swofford D.L. (1994) Hobgoblin of phylogenetics? *Nature* **369**:363-4.

Hirai, H. & Agatsuma, T. (1991) Triploidy in *Paragonimus westermani*. *Parasitology Today* **7**: 19-21.

Hirai, H., Sakaguchi, Y., Habe, S. & Imai, H.T. (1985) C-banding analysis of six species of lung flukes, *Paragonimus* spp. (Trematoda: Platyhelminthes), from Japan and Korea. *Zeitschrift Parasitenkunde* **71**: 617-629.

Hirai, H. (1988) *Paragonimus ohirai*: identification of Nucleolar Organizer Regions (NORs) and silver nitrate staining pattern in spermatogenesis. *Experimental Parasitology* **67**: 281-286.

Hirai, H., Spotila, L.D. & LoVerde, P.T. (1989) *Schistosoma mansoni*: chromosomal localisation of DNA repeat elements by in situ hybridisation using biotinylated DNA probes. *Experimental Parasitology* **69**: 175-188.

Hirai, H., Hirata, M., Aoki, Y., Tanaka, M. & Imai, H.T. (1996) Chiasma analyses of the parasite flukes, *Schistosoma* and *Paragonimus* (Trematoda), by using the chiasma distribution graph. *Genes and Genetic Systems*. **71**: 181 - 188.

Hooker, C. & Brindley, P.J. (1996) Cloning and characterisation of strain-specific transcripts encoding triosephosphate isomerase, a candidate vaccine antigen from *Schistosoma japonicum*. *Molecular and Biochemical Parasitology* **82**: 265 - 269.

Hooker, C., Yang, W., Becker, M.M., Waine, G.J., Kalinna, B.H., Nimmo, K.A., McManus, D.P. & Brindley, P.J. (1995) *Schistosoma japonicum* - heterogeneity in paramyosin genes. *Acta Tropica* **59**: 131 - 141.

Hope, M., Duke, M. & McManus, D.P. (1996) A biological and immunological comparison of Chinese and Philippine *Schistosoma japonicum*. *International Journal for Parasitology* **26**: 325 - 332.

Hotez, P.J, Zheng, F., Long-qi, X., Ming-gang, C., Shu-hua, X., Shu-xian, L., Blair, D., McManus, D.P. & Davis, G.M. (1997) Emerging and reemerging helminthiases and the public health of China. *Emerging Infectious Diseases* **3**: 303-310.

Huelsenbeck, J.P., Bull, J.J. & Cunningham, C.W. (1996) Combining data in phylogenetic analysis. *Trends in Ecology & Evolution* **11**: 152 – 158.

Janzen, D.H. (1980) When is it coevolution? *Evolution* **34**: 611-612.

Kane, R.A. & Rollinson, D.A. (1994) Repetitive sequences in the rDNA internal transcribed spacer of *Schistosoma haematobium, Schistosoma intercalatum* and *Schistosoma mattheei. Molecular and Biochemical Parasitology* **63**: 153 - 6.

Kane, R.A., Ridgers, I.L., Johnston, D.A. & Rollinson, D. (1996) Repetitive sequences within the first Internal Transcribed Spacer of ribosomal DNA in schistosomes contain a Chi-like site. *Molecular and Biochemical Parasitology* **75**: 265 - 269.

Kumar, S., K. Tamura, and M. Nei. (1993) *MEGA: Molecular evolutionary genetics analysis.* Version 1.02. The Pennsylvanian State University, University Park, PA.

Kuperus, W.R. & Chapco, W. (1994) Usefulness of Internal Transcribed Spacer regions of ribosomal DNA in Melanopline (Orthoptera: Acrididae) systematics. *Annals Entomology Society America* **87**: 751 - 754.

Latchman D.S. 1995 *Eukaryotic transcription factors* (2nd edition) Academic Press, Hartcourt Brace & Co. Publishers, London.

Lewin B. (1990) Chapters 32 - 33 in: *Genes IV* , Oxford University Press, Oxford.

Li, W. H. (1997) *Molecular evolution.* Sinauer Associates, Sunderland Massachusetts, USA. 487 pages.

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

Littlewood, D.T. & Johnston, D.A. (1995) Molecular phylogenetics of the four *Schistosoma* species groups determined with partial 28S rRNA gene sequences. *Parasitology* **111**: 167 - 75.

Long, E.O. & Dawid, B. (1980) Repeated genes in eukaryotes. *Annual Review Biochemistry* **49**: 727 - 764.

Lopez, J.V., Yuhki, N., Masuda, R. , Modi, W.S. & O'Brien, S.J. (1994) Numt: a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal Molecular Evolution* **39**: 174-190.

Luton, K., Walker, D. & Blair, D. (1992) Comparison of ribosomal internal transcribed spacers from two congeneric species of flukes (Platyhelminths: Trematoda: Digenea). *Molecular & Biochemical Parasitology* **56**: 323 - 328.

Lyons L, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ. 1996. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics* **15**: 47-56.

Maynard Smith, J. (1978) *The evolution of sex*. Cambridge University Press, Cambridge, 222 pages.

McManus, D.P. & Hope, M. (1993) Molecular variation in the human schistosomes. *Acta Tropica* **53**: 255 - 76.

Merelender A.M., Woodruff, D.S., Upatham, E.S., Viyanant V. & Yuan H.C. (1987) Large genetic distance between Chinese and Philippine *Schistosoma japonicum*. *Journal of Parasitology* **73**: 861-863.

Michot B., Bachellerie J-P. & Raynal F. (1983) Structure of mouse rRNA precursors. Complete sequence and potential folding of the spacer regions between 18S and 28s RNA. *Nucleic Acids Research* **11**: 3375 - 91.

Michot B., Després L., Bonhomme F. & Bachellerie J-P. (1993) Conserved secondary structures in the ITS2 of trematode pre-rRNA. *FEBS* **316**: 247 - 52.

Miyazaki, I., Habe, S. & Terasaki, K. (1981) On the pairing of adult lung flukes. *Medical Bulletin of the Fukuoka University.* **8**: 159-167.

Miyazaki, I. (1991) *Helminthic zoonoses.* International Medical Foundation of Japan, Tokyo.

Morgan, J. A. T. and Blair, D. (1995) Nuclear rDNA ITS sequence variation in the Trematode genus *Echinostoma*: an aid to establishing relationships within the 37-collar-spine group. *Parasitology* **111**: 609-615.

Morgan, J.A.T. & Blair, D, (1998a) Relative merits of nuclear ribosomal internal transcribed spacers and mitochondrial COI and ND1 genes for distinguishing among *Echinostoma* species (Trematoda). *Parasitology* **116** (in press).

Morgan, J.A.T. & Blair, D, (1998b) Mitochondrial ND1 gene sequences used to identify echinostome isolates from Australia and New Zealand. *International Journal for Parasitology* **28**: 493-502.

Morgan, J. A. T. and Blair, D. (1998). Trematode and monogenean rRNA ITS2 secondary structures support a four domain model. *Journal of Molecular Evolution.* (in press)

Morrison, D. A. and Ellis, J. T. (1997) Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. *Molecular Biology and Evolution.* **14:** 428-441.

Moss T. (1983) A transcriptional function for the repetitive ribosomal spacer in *Xenopus laevis*. *Nature* **302**: 223 -28.

Musters, W., Boon, K., van der Sande, C.A.F.M., van Heerikhuizen, H. & Planta, R.J. (1990) Functional analysis of transcribed spacers of yeast ribosomal DNA. *EMBO J.* **9**: 3989 - 3996

Nadler, S. A. (1990) Molecular approaches to studying helminth population genetics and phylogeny. *International Journal for Parasitology* **20:** 11-29.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Nei, M., Gu, X. & Sitnikova, T. (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Science, USA* **94**: 7799 - 7806.

Ohama, T., Owasa, S., Watanabe, K. & Jukes, T.H. (1990) Evolution of the mitochondrial genetic code IV. AAA as an Asparagine codon in some animal mitochondria. *Journal of Molecular Evolution* **30**: 329-332.

Patterson C. (1988) Homology in classical and molecular biology. *Molecular Biology & Evolution* **5**: 603-625.

Picon M., R.R. Gutell, R. Ehrlich & R. Zaha. 1996 Characterisation of a flatworm ribosomal RNA-encoding gene: promoter sequence and small subunit rRNA secondary structure. *Gene* **171**: 215 - 220.

Ramachandran, S., Beukeboom, L.W., Gerace, L., Pavlovic, N, Carranza, S. & Michiels, N.K. (1997) Isolation and characterisation of microsatellite loci from the planarian *Dugesia polychroa* (Schmidt)(Platyhelminthes: Tricladia). *Molecular Ecology* **6**: 389-391.

Rawson, P.D. & Hilbish, T.J. (1995) Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. *Molecular Biology and Evolution* **12**: 893-901.

Rhodes D. & A. Klug. (1993) Zinc finger structure. *Scientific American* **268**: 37 - 39.

Rinder, H., Rausch, R.L., Takahashi, K., Kopp, H., Tomschke, A. & Löscher, T. (1997) Limited range of genetic variation in *Echinococcus multilocularis*. *Journal of Parasitology* **83**: 1045-1050.

Saiki R.K., Gelfand D.H., Stoffel S., Scharf S., Higuchi R., Horn G.T., Mullis K.B. & Ehrlich H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA Polymerase. *Science* **239**: 487-491.

Saitou, N. & Nei , M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.

Sakaguchi, Y. & Tada, I. (1980) Karyotypic studies of lung flukes, *Paragonimus iloktsuenensis*, *P. sadoensis* & *P. westermani*, with special reference to gametogenesis in *P. westermani*. *Japanese Journal of Parasitology* **29**: 251-256.

Sambrook J., Fritsch EF. & Maniatis T. (1989) *Molecular cloning: A laboratory manual* (2nd Edition) Cold Spring Harbour: Cold Spring Harbour Laboratory Press, USA.

Schlötterer C., Hauser M-T., von Haeseler A. & Tautz D. (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Molecular Biology & Evolution* **11**: 513 - 22.

Simpson, A.J.G., Dame, J.B., Lewis, F.A & McCutchan, T.F. (1984) The arrangement of ribosomal RNA genes in *Schistosoma mansoni* identification of polymorphic structural variants. *European Journal Biochemistry* **139**: 41 - 5.

Sneath, P.H.A. & Sokal, R.R. (1973) *Numerical Taxonomy*. W.H. Freeman & Co. , San Francisco.

Sollner-Webb B. & J. Tower. (1986) Transcription of cloned eukaryotic ribosomal RNA genes. *Annual Reviews Biochemistry* **55**: 801 - 830.

Sorensen, M.D. & Quinn, T.W. (1998) Numts: a challenge for avian systematics and population biology. *The Auk* **115**: 214-221.

Southern E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**: 503.

Spolsky, C.M., Davis, G.M. & Zhang, Y. (1996) Sequencing methodology and phylogenetic analysis: cytochrome b gene sequence reveals significant diversity in Chinese populations of *Oncomelania* (Gastropoda: Pomatiopsidae). *Malacologia* **38**: 213-221.

Stewart, D.T., Saavedra, C., Stanwood, R.R., Ball, A.O. & Zouros, E. (1995) Male and female mitochondrial DNA lineages in the blue mussel *Mytilus edulis* species group. *Molecular Biology and Evolution.* **12**: 735-747.

Streelman, J.T., Karl, S.A. (1997) Reconstructing labroid evolution with single-copy nuclear DNA. *Proceedings of the Royal Society London series B.* **264**: 1011-1020.

Sunnucks, P. & Hales, D.F. (1996) Numerous transposed sequences of mitochondrial Cytochrome Oxidase I-II in Aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology & Evolution* **13**: 510-524.

Swofford, D. L. (1993) *PAUP: Phylogenetic analysis using parsimony version 3.1* Computer program distributed by the Illinois Natural History Survey, Campaign, Illinois.

Swofford, D. L., Olsen, G. J., Waddell, P. J. and Hillis, D. M. (1996) Phylogenetic inference. In *Molecular systematics* 2nd edition. edited by Hillis, D. M., Moritz, C. and Mable, B. K.. Sinauer Associates Inc., Sunderland, Massachusetts U.S.A., 407-514.

Takamiya, S., Wang, H., Hiraishi, A., Yu, Y., Hamajima, F. & Aoki, T. (1994) Respiratory chain of the lung fluke *Paragonimus westermani*: Facultative anaerobic mitochondria. *Archives of Biochemistry and Biophysics* **312**: 142 - 150.

Tan, Q. & Li, D.Y. (1990) Study on chromosome C-banding in *Paragonimus westermani*: comparison of the diploid and triploid types. *Chinese Journal of Parasitology and Parasitic Diseases* **8**: 35-37.

Tang, J., Laurent, T., Back, C. & Unnasch, T.R. (1996) Intraspecific heterogeneity of the ribosomal DNA Internal Transcribed Spacer in the *Simulium damnosum* (Diptera: Simuliidae) complex. *Molecular Biology and Evolution* **13**: 244 - 252.

Tautz D., Tautz C., Webb D. & Dover G.A. (1987) Evolutionary divergence of promotors and spacers in the rDNA family of four *Drosophila* species. Implications for molecular coevolution in multigene families. *Molecular Biology* **195**: 525-542.

Terasaki, K., Habe, S. & Ho, L.Y. (1989) Ploidy of chromosomes in the lung fluke, from China - confirmation of tetraploids. In: *Paragonimus in Asia: Biology, genetic variation and speciation* (Paragonimus research report, Number 2) (K. Kawashima ed.) pp. 26-33. Fukuoka, Japan: Kyushu University of Health Sciences.

Terasaki, K., Habe, S., Ho, L., Jian, H., Agatsuma, T. Shibahara, T. Sugiyama, H. & Kawashima, K. (1995) Tetraploids of the lung fluke *Paragonimus westermani* found in China. *Parasitology Research* **81**: 627-630.

Terasaki, K., Shibahara, T., Nada, Y. & Kayano, H. (1996) The oocyte of triploid fluke receiving intrusion of sperm from a diploid fluke - evidence for the origin of tetraploids in *Paragonimus westermani*. *Journal of Parasitology* **82**: 947-950.

Thomaz, D., Guiller, A. & Clarke, B. (1996) Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proceedings of the Royal Society London series B.* **263**: 363-368.

Thweatt, R. & Lee, J.C. (1990) Yeast precursor ribosomal RNA. Molecular cloning and probing the higher order structure of the internal transcribed spacer I by kethoxal and dimethylsulfate modification. *Journal of Molecular Biology* **211**: 305 - 320.

Trueman, J. (1997) *Cladistic* methods for phylogenetic analysis.* A workshop in the series Computational tools for biologists. Bioinformatics group, Research School of Biological Sciences, Australian National University.

Turner, B.J., Elder, J.F., Laughlin, T.F. & Davis, W.P. (1990) Genetic variation in clonal vertebrates detected by simple sequence fingerprinting. *Proceedings of the National Academy of Science, USA.* **87**: 5653-5657.

van Herwerden L., Blair D. & Agatsuma T. (1998a) Intra- and inter-specific variation in nuclear ribosomal internal transcribed spacer 1 of the *S. japonicum* species complex. *Parasitology* **116**: 311-317.

van Herwerden, L., Blair, D. & Agatsuma, T. (1998b) Intra- and inter- individual variation in ITS1 of *Paragonimus westermani* (Trematoda: Digenea) and related species: implications for phylogenetic studies. *Molecular Phylogenetics & Evolution* (in press).

van Nues, R. W., Rientjes, J. M. J., Morré, S. A., Mollee, E., Planta, R. J., Venema, J. and Raué, H. A. (1995) Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. *Journal of Molecular Biology.* **250**: 24-36.

Veldman, G. M., Klootwijk, J., van Heerikhuizen, H. and Planta, R. J. (1981) The nucleotide sequence of the intergenic region between 5.8s and 26s rRNA genes of the yeast ribosomal RNA operon. Possible implications for the interaction between 5.8s and 26s rRNA and the processing of the primary transcript. *Nucleic Acids Research.* **9**: 4847-4862.

Vogler, AP. & De Salle, R. (1994) Evolution and phylogenetic information content of the ITS-1 region in the Tiger Beetle (Cindela dorsalis). *Molecular Biology and Evolution* **11** : 393 - 405.

Wakeley, J. (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology & Evolution* **11:** 158-163.

Walker, T.K, Rollinson, D. & Simpson, A.J.G. (1986) Differentiation of *Schistosoma haematobium* from related species using cloned ribosomal RNA gene probes. *Molecular and Biochemical Parasitology* **20**: 123 - 131.

Wesson, D.M., Porter, C.H. & Collins, F.H. (1992) Sequence and secondary structure comparisons of Internal Transcribed Spacer rDNA in mosquitoes (Diptera: culicidae). *Molecular Phylogenetics & Evolution* **1**: 253 - 269.

Wesson, D. M., McLain, D. K., Oliver, J. H., Piesman, J. and Collins, F. H. (1993) Investigation of the validity of species status of *Ixodes dammini* (Acari: Ixodidae) using rDNA. *Proceedings of the National Acadamy of Sciences of the USA.* **90**: 10221-10225.

World Health Organisation (1997) *Foodborne trematode infections control.* WWW site http://www.who.ch/programmes/ctd/act/tremact.htm

Woodruff, D.S., Merenlender, A.M., Upatham, E.S. & Viyanant, V. (1987) Genetic variation and differentiation of three *Schistosoma* species from the Philippines, Laos and Peninsular Malaysia. *American Journal Tropical Medicine Hygiene* **36**: 345 - 54.

Yong, H.S., Greer, G.J. & Ow, Y.C.K. (1985) Genetic diversity and differentiation of four taxa of Asiatic blood flukes (Trematoda: Schistosomatidae). *Tropical Biomedicine* **2**: 17-23.

Zhang, D-X. & Hewitt, G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution* **11**: 247 - 251.

Zhang X.H. & Chinnappa C.C. (1994) Triosephosphate isomerase of *Stellaria longipes* (Caryophyllaceae). *Genome* **37**: 148-156.

Zischler, Geisert & Castresana (1998) A hominoid specific nuclear insertion of the mtiochondrial D-loop: implications for reconstructing ancestral mitochondrial sequences. *Molecular Biology and Evolution* **15**: 463-469.

Zouros, E., Freeman, K.R., Ball, A.O. & Pogson, G.H. (1992) Direct evidence for extensive paternal mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Nature* **359**: 412-414.

Zouros, E., Ball, A.O., Saavedra, C. & Freeman, K.R. (1994) An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proceedings of the National Academy of Science, USA.* **91**: 7463-7467.