Population Genetics of Death Valley Pupfishes (Cyprinodontidae:*Cyprinodon* Spp.) and the Identification of a New Retrotransposable Element Family

By

David D. Duvernell

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

Biology

APPROVED:

_____

Bruce J. Turner, chairman

_____ _____

Eric M. Hallerman **Muriel Lederman**

_____ _____

Bruce Wallace **David A. West**

March, 1998

Blacksburg, Virginia

Population Genetics of Death Valley Pupfishes (Cyprinodontidae:*Cyprinodon* Spp.) and the Identification of a New Retrotransposable Element Family

By

David D. Duvernell

Abstract

Study of the genetic relationships and evolutionary histories of pupfish populations (Cyprinodontidae: *Cyprinodon* spp.) from the remnant aquatic habitats of Death Valley was approached by exploring the genetic structure and divergence within and among populations using mitochondrial and nuclear DNA markers. The findings of these studies illustrate the influences of population size and isolation time in the divergence of small, fragmented populations largely via genetic drift. The information revealed in this study has implications for assessing priorities in the conservation of the unique evolutionary heritage among populations of the Death Valley pupfishes.

A new retrotransposable element family was identified and characterized. This family of genetic elements was uncovered during a search of the pupfish genome for transposable elements to be used as molecular markers for population analyses. The description of this element family, named "Swimmer 1" (SW1), provides new insights into the evolution of long interspersed nuclear elements (LINEs) in vertebrates. Therefore, a full characterization of the SW1 element family was undertaken in the Japanese medaka (*Oryzias latipes*) as well as in the pupfish genome. The Japanese medaka is a model organism widely used for genetic and developmental biology studies.

Key Words: effective population size, evolutionarily significant units, transposable elements.

Granting Agencies and Collecting Permits

This work was supported in part by grants from the Endangered Species Program of the California Department of Fish and Game, the Jeffress Memorial Trust, Sigma Xi, and the Theodore Roosevelt Memorial Fund. Collections of pupfishes in the Death Valley region were under the auspcies of the following permits or special licenses issued by units of the U.S. Dept. of the Interior or by state agencies: Endangered Species Permit PRT-769851 and Special Use Permit 56034, U.S. Fish & Wildlife Service; Collecting Permit A9103, Death Valley National Monument, National Park Service; 6840, California Desert District Office, Bureau of Land Management; S9494, Nevada Dept. of Wildlife, and a Memorandum of Understanding with the Calif. Dept. of Fish and Game.

Acknowledgments

I would like to thank my advisor, Bruce J. Turner, for his contributions and influence during all aspects of this project. You were always available when I had questions or needed help. I also want to thank the other members of my advisory committee, Drs. Eric M. Hallerman, Muriel Lederman, Bruce Wallace and David A. West, for their many helpful suggestions and comments.

I owe a huge debt of gratitude to all of the people involved in the collection of the samples from Death Valley and elsewhere. I only wish I could have helped with this aspect of the project. There are several other people to whom I owe my appreciation for their contributions to either the conceptual development of the study or the laboratory methodology. Many thanks to Jason Bond, John Elder, Anthony Furano, Hiroshi Hori, Beth McCoy, Matt Pelletier, Brenda Shirley and Holly Wichman.

Table of Contents

List of Figures

List of Tables

Background and Introduction

Population Genetic Inferences and Molecular Markers

Molecular data exhibit many attributes which make them useful for studying population genetics (reviewed by Avise, 1994). One of the most important of these features is that molecular techniques allow access to genetic variation directly at the DNA level. Consequently, expression of molecular characters is entirely genetically based. Hence, screenings of molecular markers avoid many of the problems inherent in phenotypic variation such as low heritability and phenotypic plasticity. Another important attribute of molecular data is that most variants appear to conform to the expectations of neutral theory (Kimura, 1968a, 1968b; King and Jukes, 1969). Since most molecular genetic variants appear to confer no fitness differences to the organisms in which they are found, this variation apparently is not affected by external selective pressures and environmental influences as may be true of phenotypic variance.

As a consequence of their predominantly neutral pattern of inheritance, molecular markers are generally less prone to convergence and parallel evolution. Consequently, molecular data have proven valuable in uncovering patterns of evolutionary convergence in morphological characters. For example, molecular analyses of mtDNA variation have revealed that benthic and limnetic forms of threespine sticklebacks (*Gasterosteus aculeatus*) have evolved multiple times in freshwater environments on islands off the coast of British Columbia (Thompson *et al.*, 1997). In this case, had morphological characters alone been used to infer evolutionary relationships, then the

independent origins of these convergent ecophenotypes would not have been recognized. In a similar study, dwarf and normal morphotypes of whitefish (*Coregonus cupeaformis*) were revealed to have evolved multiple times in the St. John River drainage of northern Quebec (Pigeon *et al.*, 1997). Studies like these reveal the dangers of using phenotypic characters alone when attempting to infer historical patterns of evolution and species relatedness.

The phylogeographic approach to population genetics combines the use of intraspecific phylogenetic techniques with a biogeographic interpretation of molecular patterns for studying population divergence (Avise *et al.*, 1987; Avise 1994). This approach has proven influential in revealing the dominant role of historical events in shaping genetic composition and structures of modern populations (Avise *et al.*, 1984; Bermingham & Avise, 1986; Avise *et al.*, 1987). However, the phylogeographic approach for studying population relationships and their genetic distinctiveness is limited when recently isolated populations have apparently undergone rapid divergence. The amount of time necessary for two populations to establish reciprocal monophyly at neutral loci is estimated as approximately 4Ne generations (Avise *et al.*, 1984). However, selection may drive rapid adaptive changes at some loci over time scales much shorter than could occur via mutation and drift alone (Bernatchez, 1995), and it is these undetected, phenotypically consequential differences among populations which may be most important for assessing population distinctiveness. Consequently extreme caution must be exercised when using molecular data to assess the genetic distinctiveness of populations which may have undergone recent divergence.

Another important application of intraspecific phylogenies and molecular genetic diversity has been in the assessment of current and historical effective population sizes. Assuming constant mutation rates, the level of genetic diversity present at neutrally evolving markers provides an indication of the rate of fixation of genetic variants as a consequence of genetic drift. Therefore, estimates of diversity can provide inferences about historical population sizes and may reveal evidence of founder and bottleneck events in population histories, as was shown among some members of the *mbuna* species flock of Lake Malawi (Moran and Kornfield, 1995). The sample distributions of pairwise differences in DNA sequences also can be used to detect size expansions or contractions in population histories (Slatkin & Hudson, 1991).

Most molecular population genetic studies of animals have relied on the analysis of mitochondrial DNA sequence variation. Mitochondrial DNA molecules possess a number of unique attributes that make it particularly well suited to these kinds of analyses, including a uniparental, nonrecombining mode of inheritance and a relatively high mutation rate (Avise, 1986; Moritz *et al.*, 1987). The mtDNA D loop or control region often provides the highest level of variation for population level studies, especially when PCR-based techniques are employed. This non-coding region serves as the origin of replication for the mtDNA genome and consists of a central conserved region flanked by highly variable regions (Lee *et al.*, 1995). The variable region from the tRNA gene to the central conserved region exhibits sufficient variation for many population analyses and has found wide usage in

population and incipient speciation studies (eg. Meyer *et al.*, 1990; Toline & Baker, 1995; Strecker *et al.*, 1996).

Retrotransposable Elements and Evolutionary Processes

The nuclear DNA marker developed for population analysis of Death Valley pupfishes was the SW1 retrotransposable element family, a teleost LINE family identified and characterized for the first time in this study. Retrotransposable element families are of interest for studying molecular population genetics because of their unique mutation generating properties (Kidwell & Lisch, 1997). When element sequences become transpositionally active, they generate apomorphic insertional mutations in the genomic sites where they become integrated. The transposition rates of element families can vary extensively and may, under certain circumstances, be increased substantially. There is evidence that for some element families, transpositional activity may be enhanced in populations subject to environmentally-induced stresses or severe inbreeding (Fontdevila, 1993). Such conditions often are encountered in small, isolated populations, and it has been suggested that this mutational process may even be a critical component of allopatric speciation (Fontdevila, 1993).

Insertional events of short interspersed nuclear element (SINE) families (Ohshima *et al.*, 1996; Okada & Hamada, 1997; Okada *et al.*, in press) have been employed with great success as phylogenetic markers for the determination of previously intractable relationships among some species of salmonids (Murata *et al.*, 1993; Murata *et al.*, 1996; Takasaki *et al.*, 1997) and ungulates (Shimumura *et al.*, 1997). In a similar application, insertional events of Tc1, another DNA

transposable element family, were used as markers to reconstruct interstrain evolutionary histories in *Caenorhabditis elegans* (Egilmez *et al.*, 1995). This study extends the use of transposable element insertion polymorphisms, possibly for the first time, to the study of issues in intraspecific population genetics.

Review of Pupfishes

The cyprinodontid genus *Cyprinodon* Lacépède comprises about 30 pupfishes distributed in arid regions of northern Mexico and the southwestern United States, and in Atlantic coastal habitats of North and Central America (Turner and Liu, 1977). This genus includes numerous narrowly endemic, isolated and morphologically distinctive forms throughout this range (Wildekamp, 1995). The most extensively studied members of the genus have been the desert pupfishes of the great basin region of western North America in general, and the pupfishes of Death Valley in particular. This group of pupfishes is interesting because of the high degree of local endemism and morphological divergence of many of its members, the diversity of habitats they occupy, and the threatened or endangered status of many of the species. The most comprehensive review of these pupfishes was in Miller's (1948) monograph of the cyprinodont fishes of the Death Valley system. Numerous studies also have detailed their courtship behaviors (Liu, 1969), protein electrophoretic variation (Turner, 1974; Echelle and Echelle, 1993), physiological adaptations (reviewed by Feldmeth, 1981, and Soltz and Hirshfield, 1981), and interspecific reproductive compatabilities (Liu, 1969; Turner and Liu, 1977).

The genetic relationships among the Death Valley populations and their evolutionary histories primarily have been inferred from morphological similarities of the fishes and the hydrological relationships of their habitats (Miller, 1981). However, the genetic basis of these morphological features is unknown. Further, there is evidence that the desert habitats in which they reside have changed extensively through time, so that current environmental conditions and hydrological relationships could be misleading for inferring evolutionary processes and constraints (Smith, 1981). In this study, molecular variation has been surveyed in order to resolve the genetic relationships and histories of Death Valley pupfish populations, and to provide an independent historical framework from which phenotypic data and population/environment interactions might be interpreted.

Chapter 1: Mitochondrial DNA Sequence Variation and Population Structure

Introduction

The "Death Valley pupfishes" (genus *Cyprinodon*) are a group of relict populations in the remnant aquatic habitats of the Death Valley drainage system of southern California and Nevada (USA). These populations represent a monophyletic assemblage (Echelle & Dowling, 1992) comprising three described species, *C. nevadensis*, *C. diabolis* and *C. salinus*, and nine subspecies, seven of which are extant .

During the Pleistocene, the Death Valley region was decidedly less arid, and habitats were probably at least intermittently connected by an extensive lacustrine and riverine system. Isolation of contemporary populations has occurred recently as habitats became fragmented, beginning about 10-20,000 years ago (Soltz & Naiman, 1978; Miller, 1981). Several of the now-isolated spring habitats may have become isolated as recently as the past few hundred years (Soltz & Naiman, 1978).

Continued interest in the Death Valley pupfishes as a model for rapid allopatric divergence and/or speciation stems from studies which have uncovered marked differences among at least some of the populations in morphological (Miller, 1948; LaBounty & Deacon, 1972), physiological (reviewed by Feldmeth, 1981, and Soltz & Hirshfield, 1981) and/or behavioral features (Liu, 1969). It has been hypothesized that rapid evolution of phenotypic changes in this system has been facilitated by variable environmental conditions in the isolated

habitats (selective pressures) as well as small effective population sizes (stochastic drift).

Despite the obvious organismal differences, it has proven quite difficult to document either genetic variation within, or divergence among the relict pupfish populations. Extensive allozyme studies revealed only limited variation and diagnostic differences were virtually absent (Turner, 1974, 1983; Echelle & Echelle, 1993). A phylogenetic analysis of mtDNA variation by Echelle and Dowling (1992) uncovered only limited restriction endonuclease site variation within Death Valley, making population comparisons difficult.

Within the mtDNA molecule, the D-loop or control region is often the most variable region (Brown, 1985), and has successfully been used to discern historical events such as genetic bottlenecks as well as contemporary population structure in a number of studies (eg. Rosel & Block, 1996; Toline & Baker, 1995; Fajen & Breden, 1992). We have analyzed mtDNA D-loop variation in population samples from sixteen localities in the Death Valley region. Our work focused on three primary goals:

i. to assess contemporary as well as historical relative effective population sizes based on the patterns of haplotype diversity within and among populations.

ii. to evaluate the phylogeographic relationships of extant mtDNA haplotypes in order to examine the roles of lineage sorting versus *in situ* divergence in population structure with regard to the mitochondrial genome.

iii. to assess the level of contemporary population structure and divergence relevant to defining management units (MUs) and

evolutionarily significant units (ESUs) (*sensu* Moritz, 1994a,b, 1995) for conservation purposes.

Materials and Methods

Death Valley pupfish samples were obtained from 16 populations distributed among three geographical regions (Salt Creek drainage, Amargosa River drainage and Ash Meadows, Fig. 1.1). Collection locations and taxonomy of samples are provided in Figure 1.1, and sample sizes are indicated in Table 1.1. Collections were made in April, 1994 under permits listed in the acknowledgment section. All specimens were pithed, degutted and fixed in absolute methanol in the field. Fixative was changed at least twice in the 36-48 hrs following initial fixation.

Additional specimens included: *C. macularius* **from the Cochella Valley Preserve (Riverside Co., California),** *C. radiosus* **from White Mountain Research Station (Mono Co., California),** *C. fontinalis* **from Ojo de Carbonera (Chihuahua, Mexico), and C. variegatus from Big Sabine Point (Escambia Co., Florida). Data for** *C. artifrons* **are from Strecker** *et al*. **(1996).**

mtDNA surveys

The mtDNA D-loop region was PCR amplified using primers L15926 (Kocher *et al*.**, 1989) and H16498 (Meyer** *et al*.**, 1990). These primers amplify a segment of the variable region of the mtDNA D-loop, spanning from the threonine tRNA gene to the central conserved region. Unique mtDNA haplotypes were identified and screened in all populations via direct sequencing of PCR products and/or single-strand conformation polymorphism (SSCP) analysis (Hongyo** *et al*.**, 1993, Fan** *et al*.**, 1993). The ND2 gene was PCR amplified and sequenced using primers ND2B-L and**

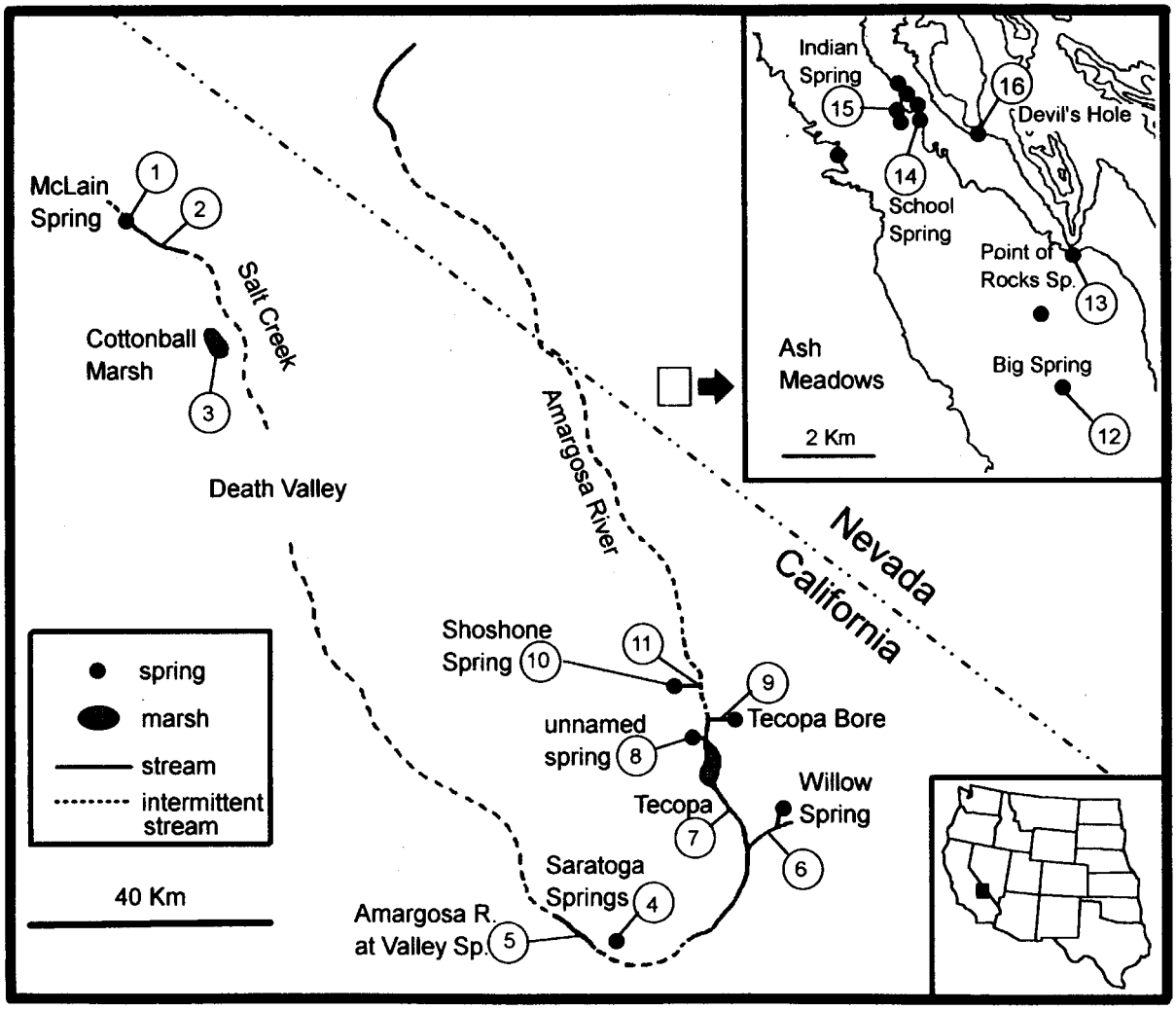

Fig. 1.1 Geographic distribution of pupfish population samples collected from the Death Valley region. Sample numbers refer to collection sites identified as follows: *C. s. salinus* **- (1) McLean Spring, (2) Salt Creek;** *C. s. milleri* **- (3) Cottonball Marsh;** *C. n. nevadensis* **- (4) Saratoga Springs;** *C. n. amargosae* **- (5) Amargosa River at Valley Spring, (6) Willow Spring Creek at China Ranch, (7) Amargosa River at Tecopa, (8) unnamed spring near Tecopa Spring Road, (9) Tecopa "Bore";** *C. n. shoshone* **(?) - (10) Shoshone Spring, (11) Amargosa River at Shoshone;** *C. n. mionectes* **- (12) Big Spring, (13) Point of Rocks Spring;** *C. n. pectoralis* **- (14) School Spring, (15) Indian Spring;** *C. diabolis* **- (16) Devil's Hole.**

Table 1.1 Geographic distribution of mtDNA haplotypes as well as haplotypic ($\hat{h}$) and nucleotide (π) diversity indices for each population sample. Death Valley populations are as designated in Figure 1.1. Data for *C. artifrons* from Strecker *et al.* (1996).

	mtDNA Haplotypes														
Population	A	B	C	D	E	F	G	H	I	J	K	Total	$\hat{h}$	(±SE)	π	(±SE)
1	13	-	-	-	-	-	-	-	-	-	-	13	0		0	
2	15	-	-	-	-	-	-	-	-	-	-	15	0		0	
3	18	-	-	-	-	-	-	-	-	-	-	18	0		0	
4	-	12	6	-	-	-	-	-	-	-	-	18	0.46	(0.055)	0.0056	(0.0010)
5	-	-	-	23	8	-	-	-	-	-	-	31	0.39	(0.054)	0.0024	(0.0004)
6	-	-	4	25	2	-	-	-	-	-	-	31	0.33	(0.070)	0.0013	(0.0004)
7	-	-	11	16	3	-	-	-	-	-	-	30	0.58	(0.036)	0.0021	(0.0003)
8	-	-	-	18	-	-	-	-	-	-	-	18	0		0	
9	-	-	8	13	-	-	-	-	-	-	-	21	0.48	(0.039)	0.0015	(0.0000)
10	-	-	17	-	-	-	-	-	-	-	-	17	0		0	
11	-	-	18	-	-	-	-	-	-	-	-	18	0		0	
12	-	-	-	-	-	1	8	-	-	-	-	9	0.21	(0.116)	0.0040	(0.0030)
13	-	-	-	-	-	-	-	8	3	-	-	11	0.42	(0.090)	0.0078	(0.0024)
14	-	-	-	-	-	-	-	-	-	16	-	16	0		0	
15	-	-	-	-	-	-	-	-	-	7	-	7	0		0	
16	-	-	-	-	-	-	-	-	-	-	10	10	0		0	
								C. variegatus:				14	0.79	(0.033)	0.0105	(0.0017)
								C. artifrons:				16	0.83	(0.046)	0.0067	(0.0009)

ND2E-H (T. Dowling, Pers. Comm.). This region was sequenced in select individuals that exhibited each of the haplotypes identified by D-loop sequences and provided additional informative characters for phylogenetic analysis.

For SSCP, 5 μl of mtDNA D-loop PCR product (approximately 100-500 ng) were mixed with 0.5 μl of 1 M methylmercury hydroxide, 4.0 μl of 15% w/v Ficoll loading buffer, and 10.5 μl H_2O. Electrophoresis was conducted on a 20% polyacrylamide gel (39:1 acrylamide:bis acrylamide) at 300 volts for 14 hours with the gel temperature held constant at 14°C. Gels were stained with Sybr Green II (Molecular Probes, Inc., Eugene, OR) and the DNA was visualized by transillumination at 302nm. The relationships between D-loop DNA sequences and the corresponding SSCP electrophoretic band mobilities were established by sequencing one or more specimens exhibiting each unique SSCP band in every population where the bands were detected.

Data analysis

All estimates of population genetic variation and population divergence were based on variation detected in the D-loop region of the mtDNA molecule. Population structure was assessed using statistics based on both haplotype and nucleotide differences. These two methods potentially have different powers for detecting population divergence, depending on sample sizes and mutation rates (Hudson *et al.*, 1992). Mitochondrial haplotype frequency differences among populations were examined using a heterogeneity chi-square test with Monte Carlo simulations (1000 randomizations, Roff & Bentzen, 1989). Chi-square

tests as well as estimates of haplotype diversity within populations were calculated using REAP (ver. 4.0, McElroy *et al.*, 1992). Estimates of nucleotide diversity within populations were calculated using DnaSP (ver.2.5, Rozas & Rozas, 1997). Sequence-based statistics for detecting population structure were calculated using an analysis of molecular variance format (AMOVA, ver. 1.55) described by Excoffier *et al.* (1992). Nucleotide diversities and Φ-statistics were estimated using a Tajima-Nei distance matrix (Tajima & Nei, 1984) constructed with MEGA (ver. 1.02, Kumar *et al.*, 1993).

A phylogenetic analysis of all eleven mtDNA haplotypes was conducted using the combined sequences from both the D-loop and the ND2 genes. It was necessary to combine the data because the number of informative sites in either data set alone were insufficient to reliably resolve the relationships of haplotypes found within Death Valley. Additionally, D-loop and ND2 sequences of *C. fontinalis*, *C. radiosus*, *C. macularius*, and *C. variegatus* were included in the analysis in order to determine if the populations targeted in this study formed a monophyletic group. Haplotype sequences were aligned with Clustal V (Higgins *et al.*, 1991). The combined sequences were then analyzed by means of maximum parsimony (MP) using PAUP (ver. 3.1.1, Swofford, 1993) with *C. variegatus* designated as the outgroup (Echelle & Dowling, 1992). For the MP bootstrap analysis, heuristic searches were performed using the "random addition of taxa" option with 500 replications. Sequences were also analyzed by the neighbor-joining (NJ) method (Saitou & Nei, 1987) using a Tajima-Nei distance matrix (Tajima & Nei, 1984). Statistical significance of the NJ branch topology was

tested using the bootstrap algorithm (500 replicates) available in MEGA (ver. 1.02, Kumar *et al.*, **1993).**

Results

The analysis of mtDNA variation in 278 Death Valley pupfish individuals is based on 335 bp of the control region sequence; 79 individuals were assayed by direct sequencing of PCR products and an additional 199 individuals were assayed by SSCP. In total, eleven unique mtDNA sequences were identified. The D-loop sequences varied at 18 nucleotide-positions including 14 transitions, 3 transversions, and a single insertion/deletion. Pairwise divergence among haplotypes ranged from 0.003 (±0.003) to 0.025 (±0.009) corresponding to 1 to 9 mutational steps. Additionally, single representatives of the eleven haplotypes identified by unique D-loop sequences differed across their ND2 genes (1047bp) at 41 base positions including 35 transitions and 6 transversions.

Additionally, D-loop and ND2 sequences were determined for single individuals of *C. macularius*, *C. radiosus*, *C. fontinalis*, **and** *C. variegatus*. **A population analysis of D-loop sequence variation in fourteen individuals from the** *C. variegatus* **population uncovered five haplotypes. All sequences have been deposited in GenBank under accession numbers AF028281-AF028313.**

Population genetic diversity

The frequencies of haplotypes in sampled populations along with estimates of haplotype diversity and nucleotide diversity are shown in Table 1.1. Haplotype diversities varied among populations; the highest levels of variation occurred in both the Amargosa River populations and the large spring populations at Saratoga Springs and Ash Meadows (Table 1.1). In contrast, estimates of nucleotide diversity were markedly

higher for the large spring populations than they were in the river populations (Table 1.1). Haplotype and nucleotide diversity estimates for *C. variegatus* **and** *C. artifrons* **are also reported in Table 1.1.**

Distribution of haplotypes and population variation

All mtDNA haplotypes were localized in one of the three disjunct geographic regions, with no indication of haplotype sharing among the Salt Creek drainage, Amargosa River drainage or Ash Meadows populations. A nested analysis of molecular variation (AMOVA) indicated that 62.52% of the variation was distributed among the three geographic regions and 25.71% of the variation occurred among populations within regions. Only 11.77% of the observed variation was distributed within populations. Haplotype distributions and consequent genetic structuring and divergence among populations within each of these three regions are summarized below (numbers in parentheses refer to population samples indicated in Fig. 1.1):

i. Salt Creek Drainage: **Populations of** *C. s. salinus* **at McLean Spring (1) and Salt Creek (2) and** *C. s. milleri* **at Cottonball Marsh (3) were all indistinguishably monomorphic for haplotype A.**

ii. Amargosa River Drainage: **Amargosa River drainage populations, comprising two or three described subspecies,** *C. n. amargosae*, *C. n. nevadensis*, **and nominal** *C. n. shoshone* **(see Taylor** *et al.*, **1988; Taylor & Pedretti, 1994 for status of** *C. n. shoshone*), **exhibited four haplotypes (B-E). Haplotype C was the most widespread, occurring in six of the eight population samples and in all three subspecies (Table 1.1). Haplotypes D and E also were common among several populations of** *C. n. amargosae*, **with the highest level of diversity occurring in the**

Tecopa (7) and Willow Spring (6) populations, where haplotypes C, D and E were all present (Table 1.1). The Saratoga Springs (4) population (*C. n. nevadensis)* was polymorphic for haplotype C and a fourth unique haplotype (B) that was not found in any of the other populations (Table 1.1). Populations of the nominal *C. n. shoshone* from Shoshone Spring (10) and its outflow at the Amargosa River (11) were both monomorphic for haplotype C (Table 1.1).

Chi-square heterogeneity tests as well as estimates of Φ_{st} (Table 1.2) revealed significant differences in all intersubspecific population comparisons between Saratoga Springs, Shoshone Springs and all river populations nominally listed as *C. n. amargosae*. However, pairwise comparisons of heterogeneity among river populations of *C. n. amargosae* were predominately not significant (Table 1.2).

iii. ***Ash Meadows*****: Five isolated populations were sampled at Ash Meadows, including all three taxa in that region (*C. n. mionectes*, *C. n. pectoralis* and *C. diabolis*). Five of the six haplotypes identified at Ash Meadows were unique to single populations. The *C. n. mionectes* populations at Big Spring (12) and Point of Rocks Spring (13) were polymorphic for two unique mtDNA haplotypes each (F & G, H & I, respectively) and *C. diabolis* (16) was monomorphic for haplotype K. Haplotype J was shared between the two *C. n. pectoralis* populations at School (14) and Indian (15) Springs. Despite the close proximity of populations in this region, the pairwise estimates of Φ_{st} at Ash Meadows (Table 1.2) were among the highest in Death Valley.**

Table 1.2 Pairwise tests for homogeneity of mtDNA haplotype frequencies among Death Valley Pupfish populations. Estimates of Φ_{st} appear above the diagonal; NS refers to estimates not significantly different from zero in randomization tests*. Probability values for χ^2 tests conducted for select pair-wise comparisons appear below the diagonal; ns refers to χ^2 values that were not significant*. Populations are as designated in Figure 1.1.

	Salt	Amargosa River						Ash Meadows			
Pops:	1,2,3	4	5	6	7	8	9	10,11	12	13	14,15	16
1,2,3	-	0.744	0.892	0.935	0.906	1.000	0.940	1.000	0.899	0.793	1.000	1.000
4		-	0.536	0.578	0.537	0.597	0.531	0.647	0.507	0.370	0.757	0.767
5		0.000	-	NS	NS	NS	NS	0.541	0.673	0.703	0.875	0.872
6		0.000	ns	-	NS	NS	NS	0.692	0.739	0.761	0.926	0.922
7		0.000	0.001	ns	-	0.320	NS	0.383	0.696	0.734	0.881	0.888
8		0.000	ns	ns	0.002	-	0.329	1.000	0.807	0.766	1.000	1.000
9		0.000	0.000	ns	ns	0.006	-	0.580	0.716	0.735	0.920	0.924
10,11		0.000	0.000	0.000	0.000	0.000	0.000	-	0.858	0.812	1.000	1.000
12									-	NS	0.849	0.849
13									0.000	-	0.789	0.789
14,15									0.000	0.000	-	1.000
16									0.000	0.000	0.000	-

*** For each statistical test, critical values were adjusted (overall α=0.05) using the sequential Bonferroni correction (Rice, 1989) for multiple tests.**

Phylogenetic analysis

Only ten parsimony informative sites were identified among the D-loop sequences of the eleven Death Valley haplotypes. An additional thirteen informative sites were obtained by combining the D-loop and ND2 sequences for phylogenetic analysis. Several topological features are revealed by the MP and NJ analyses (Fig. 1.2). First, all eleven haplotypes from Death Valley formed a monophyletic group with the next closest relative being *C. fontinalis*. **Second,** *C. salinus* **and** *C. nevadensis/C. diabolis* **were reciprocally monophyletic, but none of the taxonomically defined groups within the** *C. nevadensis/C. diabolis* **clade were reciprocally monophyletic. Finally, the two** *C. n. mionectes* **populations and the** *C. n. nevadensis* **population exhibited polymorphisms for unique pairs of divergent haplotypes. Haplotypes from the two** *C. n. mionectes* **populations (F, G, H and I) were placed into three distinct clades and the two haplotypes from Saratoga Springs (B and C) were isolated in separate clades.**

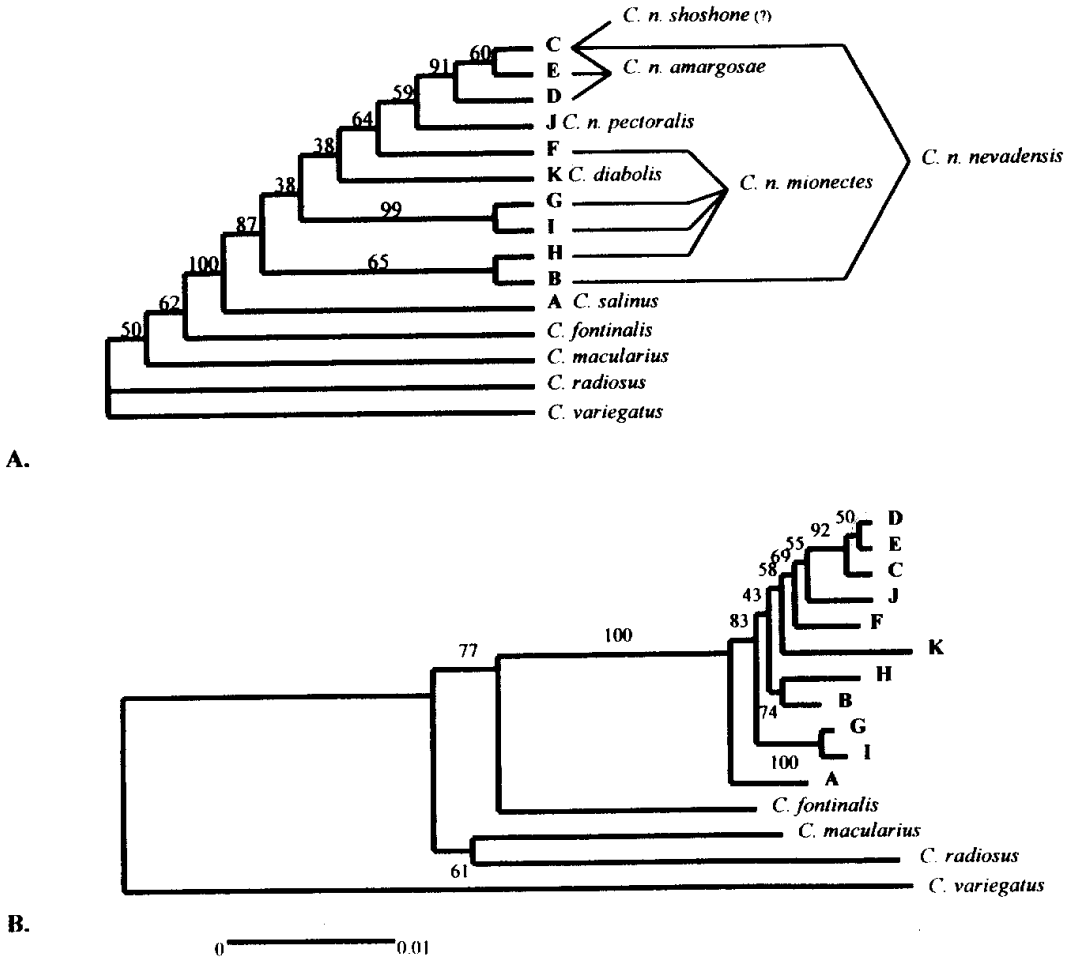

Fig. 1.2 Results of MP (A) and NJ (B) analyses of combined D-loop and ND2 sequence data. Numbers indicate the percentage of bootstrap replicates out of 500 that support each branch node. Nominal taxa and haplotype designations are indicated at the tips of the trees.

Discussion

Analysis of mtDNA control region sequences uncovered substantial genetic variation among the pupfish populations of Death Valley. Genetic diversity within populations was generally low, with most variation distributed among populations, resulting in genetic structure and divergence. Accordingly, the majority of populations, including those associated with the Amargosa River, were monomorphic or polymorphic for only weakly divergent haplotypes (max. pairwise divergence within populations = 0.006 ± 0.004). However, it is noteworthy that some of the large spring populations, including those at Saratoga Springs, Big Spring and Point of Rocks Spring, were polymorphic for mtDNA haplotypes which were clearly more divergent (divergence = 0.012 ± 0.006 to 0.018 ± 0.008). Additionally, none of the classically defined taxanomic units within the *C. nevadensis/C. diabolis* clade were reciprocally monophyletic, and a general lack of congruence was noted between gene phylogenies and the corresponding presumptive relationships of populations based on regional hydrology.

Isolation of the populations which comprise this group has undoubtedly occurred quite recently. There is little doubt that many of the extant habitats were larger and probably at least intermittently connected as recently as the late Pleistocene, with a number of the habitats becoming isolated in only the past few hundred to several thousand years (Soltz & Naiman, 1978). Therefore, it is perhaps not surprising that insufficient time has elapsed for populations to have established mtDNA reciprocal monophyly, and much of the pattern of divergence is most easily explained by stochastic, and in some cases,

incomplete lineage sorting of ancestral variation common to all members of the *C. nevadensis/C. diabolis* clade as first suggested by Echelle & Dowling (1992). This distribution of mtDNA diversity is reminiscent of the patterns of incomplete mtDNA lineage sorting which have been revealed in sympatric species flocks of similar age such as the *mbuna* of Lake Malawi (Moran & Kornfield, 1993; Parker & Kornfield, 1997) and the pupfishes of Laguna Chichancanab (Strecker *et al.*, 1996).

Current versus historical population sizes

Haplotype diversity was generally low in all contemporary Death Valley populations. However, the high level of sequence divergence among haplotypes found across isolated populations, and particularly within as well as among populations found at Ash Meadows and Saratoga Springs, supports the notion that ancestral pupfish populations in Death Valley may have been quite large. A comparison of these populations to populations of coastal pupfish relatives illustrates this point. Estimates of haplotype diversity are markedly higher in coastal populations of *C. variegatus* from Big Sabine Point, Florida (0.79±0.033; Table 1.1) and *C. artifrons* from the Yucatan Peninsula (0.83±0.046; Table 1.1) than in any of the contemporary pupfish populations (max.=0.58±0.036; Table 1.1). However, estimates of nucleotide diversity in some of the Death Valley populations (max.= 0.0078±0.0024; Table 1.1) rival those seen in the same coastal populations; although nucleotide diversity was higher in *C. variegatus* (0.0105±0.0017; Table 1.1), it was not in *C. artifrons* (0.0067±0.0009; Table 1.1).

These data indicate that, although contemporary population sizes may be quite low, effective population sizes in the ancestral Death Valley pupfishes may have rivaled contemporary coastal pupfish populations. It is perhaps not unlikely that ancestral Death Valley pupfishes, in their lacustrine environments, may have exhibited population dynamics and effective population sizes similar to their coastal relatives.

Nucleotide diversity also varied extensively within Death Valley. The differing patterns of nucleotide diversity between the polymorphic spring populations at Saratoga Springs and Ash Meadows versus those populations more closely associated with the Amargosa River may be indicative of differences in the rate or time period in which lineage sorting has occurred in those populations. Computer models have illustrated the effects of population demographics on the rate of lineage sorting (Avise *et al.*, 1984). Lineage extinction rates are strongly influenced by population size fluctuations and are substantially increased when populations are subject to size contractions (Avise *et al.*, 1984).

The Amargosa River is subject to extensive seasonal fluctuations with flash flood conditions often occurring in the wet season and extreme desiccation prevailing during the dry season (Soltz & Naiman, 1978). Consequently, populations are known to fluctuate in some habitats from many millions of individuals at the onset of the dry season to only a few thousand to tens of thousands of individuals on average by the beginning of the next breeding season (Soltz & Naiman, 1978). It is possible that under these conditions, any ancestral polymorphisms that were shared among spring/marsh and riverine habitats

could have been lost in the latter through rapid lineage extinction once gene flow ceased between those populations.

In contrast, the Ash Meadows spring flows are seasonally much more stable (Soltz & Naiman, 1978). Only recently, when spring habitats became isolated as a result of decreased water discharge, would effective population sizes have been reduced. This process would have occurred independently in each spring and could have resulted in a random distribution of divergent haplotypes such as that revealed by the phylogenetic analysis. Similarly, the two divergent haplotypes found at Saratoga Springs may indicate that this habitat, which presently consists of a small number of interconnected marshy pools, was also larger and more lacustrine in the recent past.

Historic bottleneck event in Salt Creek ancestor

In contrast to the Amargosa River and Ash Meadows populations, the Salt Creek drainage populations of *C. salinus* **were most striking in their complete lack of mtDNA variation. Such a lack of variation may be indicative of one or more historical bottleneck events. Although Salt Creek is subject to extensive seasonal fluctuations equal to or greater than those observed along the Amargosa River, adjacent McLean Spring supports one of the most abundant and stable pupfish populations in Death Valley. Likewise, nearby Cottonball Marsh appears to support a reasonably large population. A similar lack of D-loop sequence variation was also described in the northern red belly dace (Phoxinus eos; Toline & Baker, 1995), which was subject to severe bottleneck or founder events resulting from displacement, in it's northern range, associated with Pleistocene glaciation. Since both the McLean**

Spring/Salt Creek and Cottonball Marsh populations exhibited identical haplotypes it is likely that they were monomorphic prior to their separation about 2000 years ago (LaBounty & Deacon, 1972).

Higher level phylogenetic relationships

Our phylogenetic analysis with D-loop and ND2 sequences confirms the relationships among the Death Valley pupfishes and *C. macularius, C. radiosus* **and** *C. fontinalis* **suggested by mtDNA RFLP data (Echelle & Dowling, 1992). The closest relative to the Death Valley pupfishes in both studies is** *C. fontinalis*, **which is found much further south in the Ojo de Carbonera drainage of western Mexico. On the other hand,** *C. radiosus*, **which occurs in the adjacent Owens River valley and is sometimes considered a component of the Death Valley group, appears to be more closely related to** *C. macularius*. **Confirmation of those relationships provides important support for the double invasion hypothesis proposed by Echelle and Dowling (1992) with respect to the colonization of the Death Valley region by pupfishes.**

Contemporary population structure

Knowledge of the genetic structure of Death Valley populations has largely been based on morphological characters (Miller, 1943, 1948, 1950, 1981; LaBounty and Deacon, 1972; Soltz & Hirshfield, 1981; Taylor *et al.*, **1988; Taylor & Pedretti, 1994). Additionally, some populations exhibit unique, heritable, physiological, behavioral, and life history features which tend to reinforce the impression of significant biological divergence. However, the underlying genetic and historical basis for these phenotypic traits has remained largely unknown.**

Moreover, determination of the extent of genetic structure among isolated/semi-isolated populations that are morphologically similar has been difficult, particularly in regions where occasional flooding events may provide dispersal corridors among otherwise isolated populations. This problem is particularly relevant among several populations of *C. n. amargosae* and proximate subspecies occupying the Amargosa River drainage.

Analyses of mtDNA haplotype distributions and frequencies within the Amargosa River drainage indicated significant genetic structure among only a few of the geographically isolated *C. n. amargosae* populations and most estimates of Φ_{st} among those populations were not significantly different from zero. Additionally, those Φ_{st} estimates that were significant were the lowest detected throughout Death Valley. Therefore, genetic structuring among populations within the Amargosa River and its tributaries is rather limited despite the presence of potential barriers to gene flow in the contemporary hydrography of the drainage. However, isolated spring habitats along this drainage, including two populations which have been recognized as morphologically distinct subspecies, did exhibit significant structuring with respect to the river populations despite the absence of any fixed mtDNA haplotype differences.

Conservation issues

Analyses of mtDNA haplotype distributions and frequencies confirm that, in fact, the pupfishes of Death Valley are genetically structured in a manner consonant with the hydrography of the region, and that

almost all of the diversity is distributed among geographic regions and/or among populations within regions. These data indicate that throughout most of Death Valley, many of the isolated populations are demographically independent and consequently should be designated as separate management units (MUs; Moritz, 1995) for conservation. However, the data also serve to illustrate that the designation of independent management status based solely on mtDNA markers can be misleading. For example, demographic independence was not revealed among populations in McLean Spring/Salt Creek and Cottonball Marsh where mtDNA variation was absent. Yet these populations are morphologically distinctive (nominally differentiated at the subspecific level), and as previously noted, have been completely isolated for some time (LaBounty & Deacon, 1972).

Our mtDNA sequence data reveal that the largest portion of the genetic diversity in Death Valley is distributed among the various populations. Therefore we feel this data supports the broad consensus among conservation biologists that many individual pupfish populations must be preserved if the total evolutionary potential of this complex is to be maintained (Minckley & Deacon, 1991). Such long term conservation goals are increasingly based on the recognition of "evolutionarily significant units" (ESUs, Waples, 1991). Moritz has proposed (Moritz, 1994b, 1995; Moritz *et al.*, 1995) that such units be based on reciprocal monophyly of mtDNA haplotypes in the populations concerned. However, as Moritz himself points out (see "caveats and comments", Moritz *et al.*, 1995), a lack of reciprocal monophyly among most units within Death Valley precludes their designation as separate ESUs. Moreover, this definition does not take into consideration the

potential contribution of stochastic lineage sorting in the initial differentiation of isolated populations. Although the relationship between population divergence and molecular divergence is soundly based in theory (Moritz, 1994b), connections between molecular divergence and adaptively significant changes remain much less clear (Hard, 1995). Consequently, we prefer a more integrated approach to the use of molecular and phenotypic data such as that suggested by Bernatchez (1995). Under such a broader definition, the heritable, potentially adaptively significant, phenotypic distinctiveness of Death Valley pupfish populations will continue to be critical in the discrimination of ESUs in this system.

Chapter 2: *Swimmer 1*, a new low copy number LINE family in teleost genomes with sequence similarity to mammalian L1

Introduction

The genomes of virtually all organisms contain an array of retrotransposons that are capable of autonomous transposition via RNA intermediates (Finnegan, 1992; McDonald, 1993). One such class of mobile genetic elements is the "non-LTR" retrotransposons, a diverse group of retro-elements that contain polypeptide-encoding open reading frames (ORFs), but lack long terminal repeats (LTRs) and have an adenine-rich tail at the 3' end of the coding strand. We report here the salient features of "Swimmer 1" (SW1), a new family of non-LTR containing retrotransposons from the genomes of two teleost fishes, the Japanese medaka (*Oryzias latipes*) and the Desert pupfish (*Cyprinodon macularius*).

The non-LTR retrotransposons are considered to be among the oldest groups of retro-elements (Eickbush, 1994) and are known throughout the eukaryotes. They form a monophyletic group among the retro-elements, based on phylogenetic analyses of their reverse transcriptase genes (Xiong & Eickbush, 1990; Eickbush, 1994). However, despite their apparent common origin, these elements vary both structurally, for example in the number and kinds of ORFs they contain (Eickbush, 1994), and functionally, in their insertion site specificity (or lack thereof, Eickbush, 1994) and overall genomic copy number (ranging from just a few tens to many hundreds of thousands per genome; Hutchison *et al.*, 1989). These differences may reflect variation among non-LTR retrotransposon families in their evolutionary dynamics.

One of the most extensively studied groups of non-LTR elements is L1 (LINE-1), a prominent family which occurs throughout mammals (Hutchison *et al.*, 1989). The structure and evolutionary dynamics of this family have been studied in some detail in a wide array of mammalian species. The SW1 family is phylogenetically the closest known relative of the mammalian L1 family. However, several distinctive features of this new family suggest that SW1 has evolved with different dynamics than those inferred for L1, or has been subject to different constraints.

Materials and Methods

Isolation of Reverse Transcriptase Sequences

A portion of a reverse transcriptase (RT) gene was amplified from an Amargosa pupfish (*Cyprinodon nevadensis*) genome using PCR with degenerate primers (Wichman & Van Den Bussche, 1992; Wichman, pers. comm.) designed to anneal to amino acid motifs "PVKKP" and "YVDD" in conserved RT domains 1 and 5, respectively (Xiong & Eickbush, 1988, 1990). PCR conditions followed those described by Wichman and Van Den Bussche (1992). Amplicons ranging from 300 to 450 bp were excised from a 1% agarose gel, purified (Wizard Columns, Promega, Madison, WI), cloned (TA cloning kit, Invitrogen, Carlsbad, CA), and sequenced. One cloned amplicon (309 bp) out of eight was identified by a GenBank BLAST search as having a deduced amino acid sequence similar to those of other RT sequences.

Homologous sequences were subsequently isolated from the medaka (*Oryzias latipes)* genome by PCR with non-degenerate primers (SW1*Cm*-3400 & SW1*Cm*-3681, appendix 1) designed from the Amargosa pupfish sequence. These amplicons were cloned and sequenced as well, and two out of ten clones had deduced amino acid sequences similar to those of the Amargosa pupfish.

Screening of Genomic Libraries

A lambda phage library of the medaka genome (Lambda Dash II vector, Stratagene, La Jolla, CA) was acquired from the laboratory of Hiroshi Hori (Dept. Biol., Nagoya Univ., Nagoya, Japan). A second genomic library of Desert pupfish (*C. macularius*) DNA was constructed

in the Lambda GEM-11 vector (Promega, Madison, WI) following manufacturer's protocols. Genomic libraries were transferred to nylon membranes and hybridized to radiolabeled probes (Random Primers DNA Labeling System, Gibco BRL, Gaithersburg, MD) derived from the plasmid clones described above. Nylon membranes were hybridized overnight at 65°C following the procedures of Church and Gilbert (1984), and then washed three times with 2X SSC, 0.1% SDS at 65°C. Hybridized membranes were exposed to Kodak BioMax films (Rochester, NY) at -80°C for 12 hours with intensifying screens.

Southern Hybridizations

Medaka and pupfish genomic DNAs (10ug each) were digested with the appropriate restriction enzyme (*Bam*H I or *Pvu* II) according to supplier specifications. Digested DNAs were size fractionated on 1% agarose gels (1X TAE), alkaline transferred *in vacuo* to nylon membrane (Hybond-N, Amersham, Arlington Heights, IL) according to standard protocols (Ausubel *et al*., 1987), and probed according to procedures followed for the library screening. Probes were prepared by PCR amplification of various regions of the SW1 lambda phage clones. The primer pairs used to amplify these clones are indicated in Figure 2.1c and Table 2.1.

Nucleotide Sequencing

The nucleotide sequences of cloned amplicons and genomic library isolates were determined by the dideoxy method using Thermo Sequenase (Amersham, Arlington Heights, IL). Primers were designed using Primer Select (DNASTAR); primer sequences are provided in Table 2.1 and primer

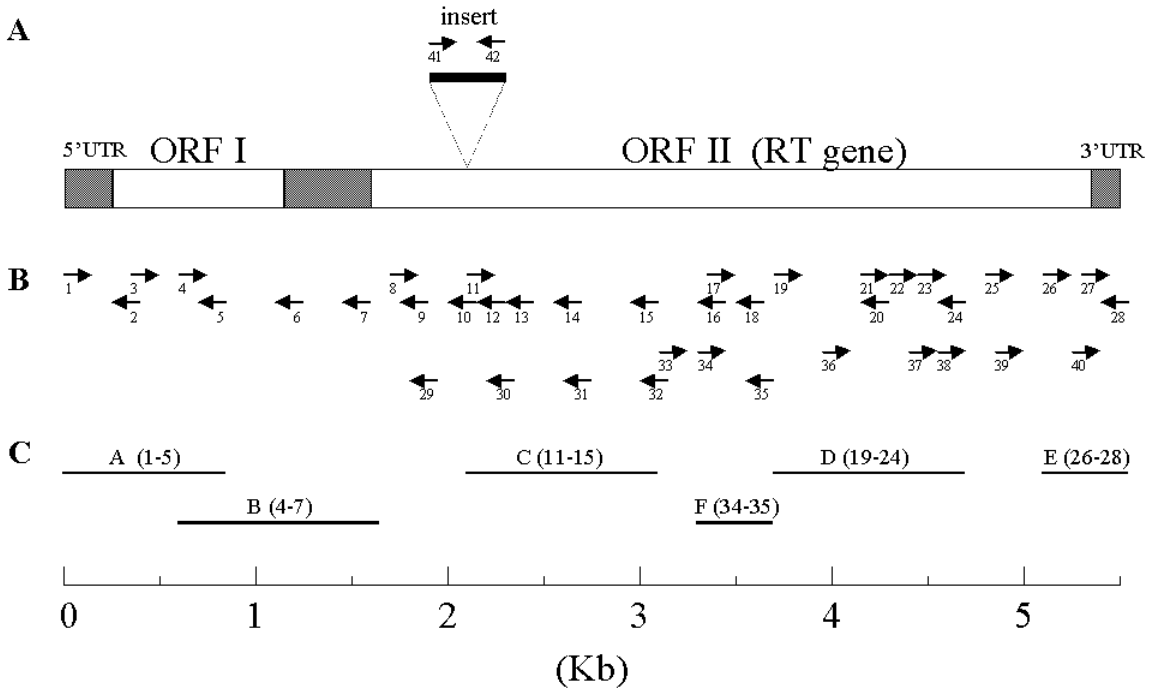

Fig. 2.1 Diagram of the composite SW1 element in medaka. *A*, **structural features which are typical of non-LTR elements are indicated, including: i) 5'UTR, presumably where the internal promoter region is located, ii) ORFI which encodes an RNA binding protein in L1, iii) ORFII which encodes the RT gene, iv) 3'UTR which ends in a poly-A region. Also indicated is the location of a putative transposable element like insert identified in clone λSW1*Ol*13.** *B*, **arrows indicate the location and orientation of primers used for sequencing the medaka (1-28), and pupfish (29-40) SW1 elements as well as the putative insert (41, 42) in clone λSW1*Ol*13. Numbers correspond to primer numbers indicated in Table 2.1.** *C*, **lines indicate the location of probes A-F used in the Southern blot experiments; numbers in parentheses indicate the primers used to construct each probe.**

Table 2.1 Primer Names and Sequences

	Name		Sequence
1.	SW1*Ol*-0020	+	ATCGGGATGGCTGTGTAGTG
2.	SW1*Ol*-0375	-	TCTTCCATTCGGTTATTCACTTTA
3.	SW1*Ol*-0482	+	CTGGAAAGCCGCGAAAGGAGA
4.	SW1*Ol*-0787	+	GACTACCCTCCACTCATTCTC
5.	SW1*Ol*-0810	-	CTTGAGAATGAGTGGAGGGTAGTC
6.	SW1*Ol*-1202	-	TTATCAGAAACGGAAACCA
7.	SW1*Ol*-1538	-	CATTTTGAACACTGAACTCTGAA
8.	SW1*Ol*-1814	+	GCTGGGGCAGTGAATTATCAACAC
9.	SW1*Ol*-1879	-	ATATACCTTGCCTTCTTTG
10.	SW1*Ol*-2012	-	TCCGCCGCATATCAAGGTTCC
11.	SW1*Ol*-2101	+	GGAGGAAATGGGCATAGTTGAT
12.	SW1*Ol*-2177	-	GCATTATGCGGATGTGAGTAGTGA
13.	SW1*Ol*-2315	-	TTTTCTGTCCAGCCAAGTAG
14.	SW1*Ol*-2644	-	ACCGCCAGTGTATTGATGTCG
15.	SW1*Ol*-3041	-	TTTCCATCGTTTTATACCATTCTG
16.	SW1*Ol*-3407	-	TCACCGAGTCAAATGCCTTCT
17.	SW1*Ol*-3487	+	CAAACCAACGGCCAGGATTA
18.	SW1*Ol*-3621	-	CTGATCCACTGCCCAAGAGG
19.	SW1*Ol*-3854	+	GAATGGCAGGCAGACTCAAT
20.	SW1*Ol*-4227	-	TGGGCAGCACAGAAATAGTC
21.	SW1*Ol*-4227	+	AGCTGAGGCCTCTTATCTG
22.	SW1*Ol*-4265	+	ACTGCAGGGTGGAAAGAT
23.	SW1*Ol*-4515	+	CCAAAGGGCTAACTGATT
24.	SW1*Ol*-4639	-	ATAGTGCCGGACCTGTAGA
25.	SW1*Ol*-4917	+	TTTTTACTACACCCTCTCAA
26.	SW1*Ol*-5164	+	AGCCAGTAAGAAAGCAATAACCAG
27.	SW1*Ol*-5323	+	AGAATATGTTAAGCCTGTA
28.	SW1*Ol*-5522	-	ATTGAAAAAGCACAGGATAAA
29.	SW1*Cm*-1937	-	TGCCTGGAGGTACGTAAACATTGA
30.	SW1*Cm*-2283	-	ACGGGATTGTGGTCTGAG
31.	SW1*Cm*-2724	-	CTTAGCCTGTAGGAGAGTAGC
32.	SW1*Cm*-3081	-	ACCCAGTTAAATGCTCTCAGTA
33.	SW1*Cm*-3172	+	TTATCGGCCAATCAGTGTTTTA
34.	SW1*Cm*-3400	+	CTCGGTACGGTGGGCCTTTTTAT
35.	SW1*Cm*-3681	-	AATAGGGCCAGTTTTTGT
36.	SW1*Cm*-4062	+	TACCAGTTAAAATACCTCAGAAAT
37.	SW1*Cm*-4448	+	AGGTGGTGCGCTTACGACTCTG
38.	SW1*Cm*-4614	+	ACAGGTATTTGCAGGTTAGACATT
39.	SW1*Cm*-4953	+	GTGGGGACGCATGTTGGAGA
40.	SW1*Cm*-5303	+	AAAATCTGGACTAAATGGACTGAG
41.	SW1*Ol*7a	+	TTTCTGTTTCCTGTTTTATTTTGA
42.	SW1*Ol*-b	-	ATGGTGCAGGGTGGCTTGAG

Note: Numbers to left of primer names correspond to numbers indicated in Figure 2.1. Primer names include nucleotide position of the 5' end of primer sequences relative to clone λSW1*Ol*7. Symbols following primer names indicate the primer orientation.

positions are indicated in Figure 2.1b. Sequencing templates were prepared by PCR amplification of the cloned inserts using primers that were complementary to the T3 and T7 RNA promoter sites in the multiple cloning cassettes of both the plasmid and the lambda vectors. Templates for PCR consisted of boiled bacterial colonies or minipreps of lambda phage DNA (Grossberger, 1987). The Expand Long Template PCR System (Boehringer Mannheim, Indianapolis, IN) was used to amplify phage DNA inserts that ranged from 12-20 kb. Amplification products were purified for sequencing using GENECLEAN (BIO 101, Inc., Vista, CA). Nucleotide sequences were identified by BLAST search (National Center for Biotechnology Information, Bethesda, MA) to DNA and protein databases in GenBank using the BLASTN and BLASTX algorithms.

Sequence Analysis

Nucleotide and inferred amino acid sequence alignments were performed using CLUSTALW (Higgins *et al.*, 1991). Phylogenetic analyses were performed on the complete RT amino acid sequences by means of maximum parsimony (MP) using PAUP (ver. 3.1.1, Swofford, 1993). The DRE element (Marschalek *et al.*, 1992) was designated as the outgroup based on global phylogenetic analyses of retro-elements (Xiong & Eickbush, 1990; Eickbush, 1994). For the MP bootstrap analysis, heuristic searches were performed using the "random addition of taxa" option with 500 replications.

The inferred RT amino acid sequences were also analyzed by the neighbor-joining (NJ) method (Saitou & Nei, 1987) using a Poisson-correction distance matrix. Statistical significance of the NJ branch topology was tested using the bootstrap algorithm (500 replicates)

available in MEGA (ver. 1.02, Kumar *et al.*, 1993). Estimates of the relative frequency of synonymous (K_s) and nonsynonymous (K_a) substitutions (Nei & Gojobori, 1986) were calculated using DnaSP (ver. 2.5, Rozas & Rozas, 1997).

Results

Identification of Structural Features in the SW1 Family

Characterization of the structural features of the SW1 element family is based on the complete sequences of four inserts obtained from a medaka genomic library (λSW1*Ol*6, λSW1*Ol*7, λSW1*Ol*8, λSW1*Ol*13). The complete reverse transcriptase and 3' untranslated region (UTR) sequences of a library-derived pupfish insert (λSW1*Cm*1) were also determined.

The 5' and 3' ends of the SW1 elements were identified by the presence of 12-15 bp short direct repeats (SDRs) of the presumed host DNA sequence at the insertion sites. A poly-A tail, ranging from 9 to 13 bp in length, was present immediately upstream of the 3' SDRs. These two features are typical of non-LTR elements (Hutchison *et al.*, 1989; Eickbush, 1992). There was no sequence similarity between the flanking DNA sequences of cloned SW1 elements, and a BLAST search, using 50-100 base pairs of flanking sequence, did not uncover any matches in GenBank.

The full length SW1 sequences were approximately 5.5 kb in length and each contained the four regions typical of most non-LTR retrotransposons: a 5'UTR, ORFI, ORFII, and a 3'UTR (Fig. 2.1a). Initial BLAST searches of inferred amino acid sequences for ORFs I and II indicated that these sequences were most similar to respective regions of the mammalian L1 family. ORF II encodes the reverse transcriptase (RT) enzyme characteristic of all retro-elements (Eickbush, 1994). The reading frames of the RT sequences were

uninterrupted and of identical length in two of the four cloned inserts from the medaka library (λSW1*Ol*7, λSW1*Ol*8). These sequences code for a protein 1259 amino acids in length. RT genes in the other two medaka clones (λSW1*Ol*6, λSW1*Ol*13) as well as the pupfish clone were interrupted by insertions and deletions, some of which resulted in frameshift mutations.

A phylogenetic analysis of complete RT amino acid sequences (Fig. 2.2 a,b) of the medaka (λSW1*Ol*7) and pupfish (λSW1*Cm*1) SW1 elements as well as several mammalian L1 representatives (including human, rat and mouse) and other phylogenetically related (Xiong & Eickbush, 1990; Eickbush, 1994) non-LTR retrotransposons, including Tx1, Cin4 and DRE, revealed that medaka and pupfish SW1 elements are monophyletic and are more closely related to the L1 element family than to any other known retro-elements. A partial alignment of RT amino acid residues is shown in figure 2.3a. This alignment includes the seven conserved domains within the RT gene that have been identified in various groups of retro-elements (Xiong & Eickbush, 1988,1990; Eickbush, 1992). A comparison of the percent identity among RT amino acid sequences derived from SW1 (medaka and pupfish) and L1 (human and rat) elements confirmed that the highest sequence identity occurs in the middle third of the gene where the conserved RT domains are located (Fig. 2.4).

The ORF I sequence of L1 encodes a p40 protein with RNA binding function (Hojoh and Singer, 1996). An alignment of the inferred amino acid sequences of ORF I from SW1*Ol*, L1*Rn*, and L1*Hs* (Fig. 2.3b) indicated that the C-terminal two-thirds of ORF I is more highly conserved than the N-terminal region, where several gaps were

introduced into the alignment. The C-terminal region of ORF I is more highly conserved among L1 members as well (Demers, Matunis & Hardison, 1989).

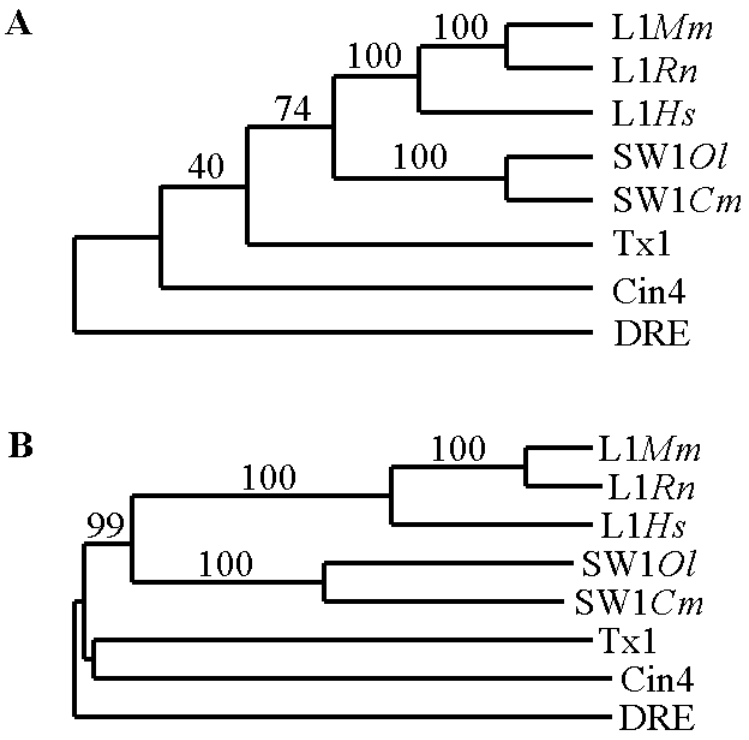

Fig. 2.2 Results of MP (*A*) and NJ (*B*) analyses of complete inferred amino acid sequences from the reverse transcriptase genes of select non-LTR containing elements. Numbers indicate the percentage of bootstrap replicates out of 500 that support each branch node. Sequences used: SW1*Ol*, λSW1*Ol*7; SW1*Cm*, λSW1*Cm*1; L1*Mm*, *Mus musculus* (M29324; Shehee *et al.*, 1987); L1*Rn*, *Rattus norvegicus* (U83119; Ilves *et al.*, 1992); L1*Hs*, *Homo sapiens* (U93569; Sassaman *et al.*, 1997); Tx1, *Xenopus laevis* (M26915; Garrett *et al.*, 1989); Cin4, *Zea mays* (Y00086; Schwarz-Sommer *et al.*, 1987); DRE, *Dictyostelium discoideum* (102250; Marschalek *et al.*, 1992).

A

```
SW1Ol  WVMEKKMTPLSWNEAIISIIPKEGKDRLDCANYRPVSVLNIDYKLFTSIISRRLETILPM
SW1Cm  WVMEKKVIPSSWSEAIISIIPKENRDKLECSNYRPISVLNIDYKLFTSILSKRLEIILPD
L1Mm   KIEVEGTLPNSFYEATITLIPKPQKDPTKIENFRPISLMNIDAKILNKILANRIQEHIKA
L1Rn   KIETDGALPNSFYEATITLIPKPHKDTTKKENFRPISLMNINAKILNKILANRIQEHIKT
L1Hs   SIEKEGILPNSFYEASIILIPKPGRDTTKKENFRPISLMNIDAKILNKILANRIQQHIKK
Tx1    EAFKKGELPLSCRRAVLSLLPKKGDL-RLIKNWRPVSLLSTDYKIVAKAISLRLKSVLAE
Cin4   FYHHKCKSLHLVNEANIVLLPKRENP-DRIDLFRPISLINSCMKIITKIMATRLAPRMNE
DRE    NHFWNTTIPKDFKQGILITIYKNKGDPNNLDNYRPITLLNVDYKIYSKIINNRILKLLNK
                Domain 1                          Domain 2

SW1Ol  LMHKDQTGFIKQRQTQDSIRKVLNIIHQVVQQ-----KQETLVISLDAEKAFDSVRWTFL
SW1Cm  LINKDQTGFIRQRQTQDNIRKTLHIMKYVSQH-----KLETLILSLDAEKAFDSVRWAFL
L1Mm   IIHPDQVGFIPGMQGWFNIRKSINVIHYINKLK----DKNHMIISLDAEKAFDKIQHPFM
L1Rn   IIHHDQVGFIPGMQGWFNIRKTINVIHYINKLK----EQNHMIISLDAEKAFDKIQHPFM
L1Hs   LIHHDQVGFIPGMQGWFNIRKSINVIQHINRAK----DKNHVIISIDAEKAFDKIQQPFM
Tx1    VIHPDQSYTVPGRTIFDNVFLIRDLLHFARRT-----GLSLAFLSLDQEKAFDRVDHQYL
Cin4   IVSTTQNAFIQKRSIHDNFLYVQKVIKKLHKS-----KQAALFVKLDISKAFDSLNWAYL
DRE    IISPFQTGFVPRRLLHDNIITLNSTIEIIKREINTKEDMEPIITFYDFEKAFDSISHNAI
                                                        Domain 3

SW1Ol  YKVLGKFGFCKSIIETISGLYNKPTARIKINGDLTETITLERGTRQGCNMSALLFALYIE
SW1Cm  YKVLSKFGFHPNIVDTFAALYSKPTAKIKVNGDLTNSFTLQRGSRQGCGASPLLFALFLE
L1Mm   IKVLERSGIQGPYLNMIKAIYSKPVANIKVNGEKLEAIPLKSGTRQGCPLSPYLFNIVLE
L1Rn   IKVLERIGIQGPYLNIVKAIYSKPVANIKLNGEKLEAIPLKSGTRQGCPLSPYLFNIVLE
L1Hs   LKTLNKLGIDGMYLKIIRAIYDKPTANIILNGQKLEAFPLKTGTRQGCPLSPLLFNIVLE
Tx1    IGTLQAYSFGPQFVGYLKTMYASAECLVKINWSLTAPLAFGRGVRQGCPLSGQLYSLAIE
Cin4   LDVLKALGFTQKWRDWIATILGSSSSKIIINGQQTKEIKHMRGVRQGDPLSPFLFILAMD
DRE    LRTLAHLKLPLKMVLTIMNLLNESETSVYINNSLSKSFTSKRGTKQGDPISPTIFALVVE
                                                        Domain 4

SW1Ol  PLGQWIRQRA---DIKGVKVSGKEQKLSLFADDLLLTISQPTKTLPIIMDSLKDFGTLSG
SW1Cm  PLDQWIRQRS---DISGVTMTAGEQKLALFADDVLIFLTQPNQTLPRLMTVLEEYGSLSG
L1Mm   VLARAIRQQK---EIKGIQIGKEEVKISLFADDMIVYISDPKNSTRELINLINSFGEVAG
L1Rn   VLARPIRKQK---EIKGIQIGKEEVKISLFADDMIVYLSDPKSSTREQLKLINNFSKVAG
L1Hs   VLARAIRQEK---EIKGIQLGKEEVKLSLFADDMIVYLENPIVSAQNLLKLISNFSKVSG
Tx1    PFLCLLRKRL---TGLVLKEPDMRVVLSAYADDVILVAQDLVD-LERAQECQEVYAAASS
Cin4   PLQRMIERAAHEGLLGQVLPNGAKFRCSLYADDAGVFVRADKLDLKVLKRILEAFEWCSG
DRE    CMATTIINDR-----CINGVTKETIKILQFADDTATIAYNFMD-HFLMNEWIKKFCQATS
                                      Domain 5             Domain 6

SW1Ol  YKINVNKIQVLTLNYS--PPQNIKDEYKWEWQADSIKYLGIAL
SW1Cm  YKINVNKTQILRLNFN--PSTRIKNMYKWIWDSEHIKYLGVVL
L1Mm   YKINSNKSMAFLYTKNKQAEKEIRETTPFSIVTNNIKYLGVTL
L1Rn   YKINSNKSVAFLYTKEKQAEKEIRETTPFIIDPNNIKYLGVTL
L1Hs   YKINVQKSQAFLYNNNRQTESQIMGELPFTIASKRIKYLGIQL
Tx1    ARINWSKS-SGLLEGSLKVDFLPPAFRDISWESKIIKYLGVYL
Cin4   LKINFEKTEIFPIRYPESLWSNLMEVFPGKYSNFPGKYLGLPL
DRE    AKINQTKCSCITFKWNTRTLYTVI--------KSNERYLGFDF
                                     Domain 7
```

B

```
SW1Ol  ----------------------------------------------------MAEYN
L1Rn   MVRGKRRNPSNRNQDYMASSEPNSPTKTNMEYPNTPEKQDLVSKSYLIMMLEDFKKDVKN
L1Hs   M--GKKQNRKTGNSKTQSTS-P--PPKERSSSPAT--EQSWMENDFDELREEGFRRSNYS

SW1Ol  ILLQELRAFRQE-----------NNEKLESIKEDIAKVN--------------------
L1Rn   SLREQVEAYREESQKSLKEFQENTIKQLKELKMEIEAIKKEHMETTLDIENQRKRQGAVD
L1Hs   ELREDIQTKGKE--------VENFEKNLEECITRISSTEKCLKELMELKTKARELR-EEC

SW1Ol  ----NRMEEAEGRIEKAEERIQTMEDVMVELMQVHVKLTDKLTDLESRERRENIRIYGVP
L1Rn   TSITNRIQEMEERISRAEDSIEIIDSTVKDNVKRKKLLVQNIQEIQDSMRRSNLRIIGIE
L1Hs   RSLRSRCDQLEERVSAMEDEMNEMKREGKFREKRIKRNEQSLQEIWDYVKRPNLHLIGVP

SW1Ol  ETS-ERDSPSMSAFVETLLREGLKLEGAENINIERAPSLTRAAPPNGASPRSILVKFLSF
L1Rn   ESEDSQLKGPVNVFNKIIEENFPNLKKEIPIGIQEAYRTPNRLDQKRNTSRHIIVKTPNA
L1Hs   ESDVENGTKLENTLQDIIQ-NFPNLARQANVQIQEIQRMPQRYSSRRATPRHIIVRFTKV

SW1Ol  KTKEQILRKAWQQKGFTWKGKQISLDNDYPPLILKKRREYAAIRRILKDKQIQFQTLFPA
L1Rn   QNKERILKAVREKGQVTYKGRPIRITPDFSPETMKARRSWTDVIQTLREHKCQPRLLYPA
L1Hs   EMKEKMLRAAREKGCVTLKGKPIRLTADLSAETLQARREWGPIFNILKEKNFQPRISYPA

SW1Ol  RLKVKYADGVKIYNTATEASEDMSERGFPVEVIKPPESVLERYKQLNTWNRVTRGTGRTA
L1Rn   KLSINIDGETKIFHDKTKFTQYLSTNPALQRIINGKAQHKEASYTLEEARN---------
L1Hs   KLSFISEGEIKYFIDKQILRDFVTTRPALKELLK-EALNMERNNRYQPLQNHAKM-----

SW1Ol  PGPPGPSYKEKLRAFRRTGADPAVE
L1Rn   -------------------------
L1Hs   -------------------------
```

Fig. 2.3 Partial alignment of deduced amino acid sequences of (*A*) a portion of ORF II encompassing the seven conserved domains (Xiong and Eickbush, 1988, 1990), and (*B*) ORF I. Amino acid residues which are shared between either SW1*Ol* or SW1*Cm* and other non-LTR retrotransposons are indicated in bold. Sequences included in the alignments are the same as those indicated in Figure 2.2.

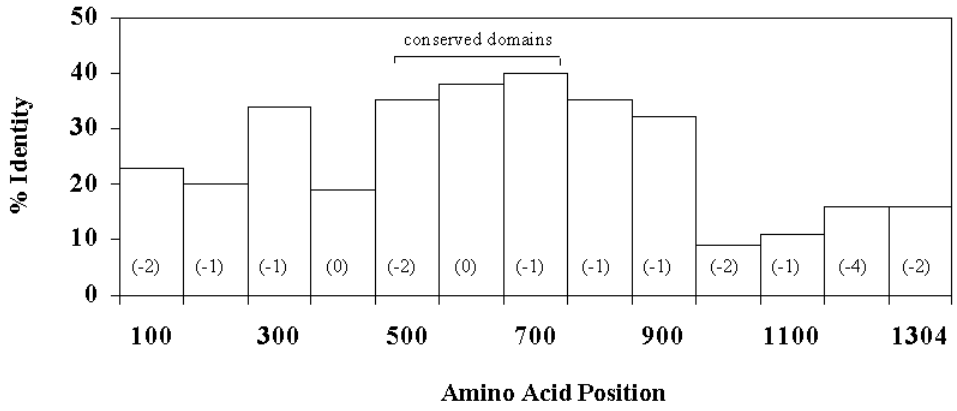

Fig. 2.4 A comparison of the percent identity among ORF II inferred amino acid sequences derived from SW1 (medaka and pupfish) and L1 (human and rat) elements (sequences and alignments are the same as indicated in Fig. 2.2). The average percent identity among the four pairwise comparisons is shown in one-hundred amino acid segments along the entire ORF II sequence. Numbers indicated in parentheses represent gap penalties that were subtracted from the average percent identity for each segment. The regions which contain the conserved domains defined by Xiong and Eickbush (1988, 1990) are noted with a bracket.

The 5'UTRs of SW1*Ol* clones were approximately 265 bp in length. This region serves as the transcriptional promoter site in L1 elements and appears to have been acquired repeatedly from non-L1 sources (Adey *et al.*, 1994). A BLAST search of GenBank using the 5'UTR sequences from SW1*Ol* sequences did not uncover any matches. Mammalian L1 promoters exhibit a high G + C content (>50%) and qualify as CpG islands (Furano *et al.*, 1988); these features appear to be important for the suppression of L1 transcription through hypermethylation (Woodcock *et al.*, 1997). Although the G + C content of the SW1*Ol* 5'UTR was higher (47%) than that of the remainder of the element sequence (38%), the incidence of CpG and GpC did not appear higher in this region, which therefore does not qualify as a CpG island.

It has been suggested that certain secondary structural features of the 3'UTRs of non-LTR elements may be essential for transcript recognition during reverse transcription (Luan & Eickbush, 1995). Among the L1 family members, a highly conserved G-rich polypurine tract (PPT), which forms stable tetraplexes, has been implicated for this function (Howell & Usdin, 1997). The 3'UTRs of SW1*Ol* and SW1*Cm* displayed no sequence similarity to each other, and neither of these elements exhibited the conserved G-rich PPTs characteristic of L1 members.

Analysis of Genomic Members of the Medaka SW1 Family

Pairwise nucleotide sequence divergence among the four SW1*Ol* clones was low (range = 0.007±0.001 to 0.015±0.002). In addition, although some of the ORF II sequences contained insertions and/or deletions, both ORF I and ORF II of SW1*Ol*7 and SW1*Ol*8 were intact and

thus potentially competent for replication. Synonymous substitutions (K_s) exceeded nonsynonymous substitutions (K_a) in all pairwise comparisons of both ORF I and ORF II sequences (Table 2.2). Taken together, these data indicate that the sequences have been under selective pressure and probably were generated quite recently.

We used Southern blot analyses to determine the composition of the SW1 family in medaka and pupfish genomes. Restriction enzymes were selected that did not cleave the SW1 sequences. When genomic DNAs from several medaka laboratory strains (an outbred orange variant mutant line, and inbred lines HO5 and HNI; compliments of Hiroshi Hori, Dept. Biol., Nagoya Univ., Nagoya, Japan) were digested with *Bam*H I and challenged with a portion of the RT gene (probe C, Fig. 2.1c), approximately six to eight bands were visualized per individual (Fig. 2.5a). The variable banding patterns of the orange variant mutant individuals contrasted sharply with the invariant patterns of the two inbred strains (HO5 and HNI). DNAs from three individuals from the orange variant mutant strain subsequently were challenged with five consecutive probes which collectively spanned the entire SW1 element (probes A-E, Fig. 2.1c). The banding patterns revealed by each of the probes were identical for each individual. When *Pvu* II digests of sixteen *C. macularius* individuals were probed with a single *C. macularius* derived probe (probe F, Fig. 2.1c), a similar number of bands, with a level of polymorphisms similar to that seen in the orange medaka strain, was detected (Fig. 2.5b). The small number of bands detected in the medaka and pupfish genomes suggested a low copy number for the SW1 family, perhaps not exceeding 20 copies per genome. Additionally, the band profiles generated in the medaka, with each of

Table 2.2 Ratio of synonymous (K_s) to nonsynonymous (K_a) substitution rates in ORFs I and II among SW1*Ol* genomic isolates.

	λSW1*Ol*6	λSW1*Ol*7	λSW1*Ol*8	λSW1*Ol*13
λSW1*Ol*6	----	2.25	2.69	1.11
λSW1*Ol*7	1.48	----	3.30	1.63
λSW1*Ol*8	2.56	3.02	----	2.53
λSW1*Ol*13	3.08	3.68	3.27	----

Above the diagonal, estimates of K_s/K_a for ORF I. Below the diagonal, estimates of K_s/K_a for ORF II. Both K_s and K_a values were calculated using DnaSP (Rozas & Rozas, 1997) by the method of Nei and Gojobori (1986).

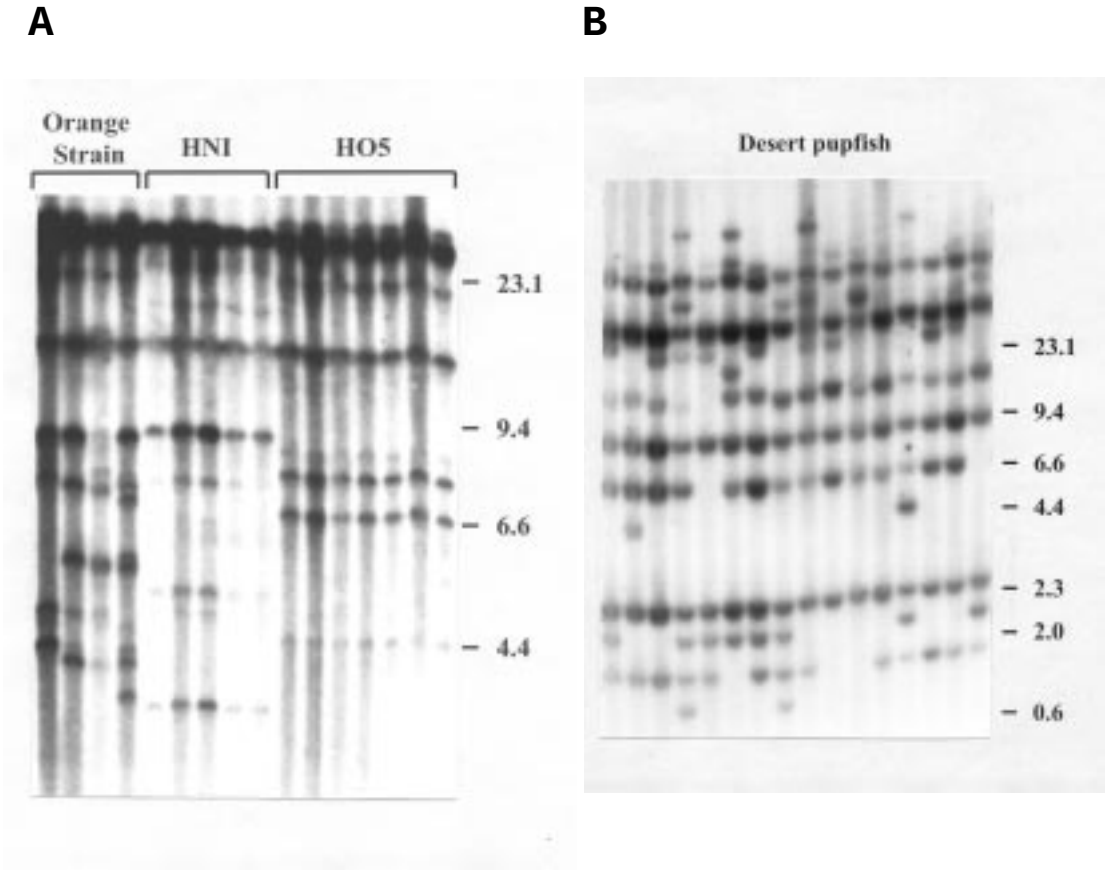

Fig. 2.5 Southern blot analyses of *(A)*, medaka laboratory strains and *(B)*, a population sample of Desert Pupfish (*C. macularius*). Genomic DNA digests were challenged with probes C and F (Fig. 2.1c) corresponding to a portion of the reverse transcriptase gene of SW1 elements from the corresponding species. The restriction enzymes chosen for this analysis apparently do not cut within the respective SW1 element sequences so that each electrophoretically distinct band is defined by the flanking genomic DNA of the corresponding insert and may represent a single, unique insertion element in the genomes.

the five probes, suggested that all SW1 copies detected were likely of full length.

An Insert in One Medaka Clone may be an LTR-containing Element

A 410 base pair presumptive insert in ORF II of clone λSW1*Ol*13 exhibited seven base pair inverted terminal repeats nested within six base pair short direct repeats which were derived from the SW1 sequence at the site of insertion. A Southern blot of medaka genomic DNA, probed with a PCR amplified fragment of this sequence (Fig. 2.1a), revealed that the sequence was repeated approximately 6 times in the medaka genome. This may be a low copy number LTR-containing transposable element with some features reminiscent of Miniature Inverted Repeat Transposable Elements (MITEs; Kidwell & Lisch, 1997).

Discussion

A new family of retrotransposable elements, named SW1, has been identified and its sequence characterized in the Japanese medaka and the Desert pupfish. The complete SW1 sequence, deduced from multiple copies isolated from the Japanese medaka, displays structural features which are typical of many non-LTR containing retrotransposons, including: i) two open reading frames with deduced amino acid sequence similarities to ORFs in other non-LTR retrotransposons, ii) 5' and 3' untranslated regions, iii) an overall length of 5.5 Kb, iv) a poly-A tail, and v) short target site duplications flanking the element.

Although a number of LTR-containing element families have now been identified in fishes (reviewed by Izsvák *et al.*, 1997), the only non-LTR-like element described thus far from teleost genomes is the RSg-1 family in salmonids (Winkfein *et al.*, 1988). None of the reported RSg-1 sequences contained the ORFs characteristic of other non-LTR elements (Winkfein *et al.*, 1988), and no discernible sequence similarity was found between the RSg-1 sequences and the SW1 family described here.

The SW1 element family exhibits several features which suggest it may be transpositionally active. First, extensive polymorphisms in Southern blot analyses of both pupfish and medaka populations (Fig. 2.5) suggested that few insertion sites have become fixed for the presence of the SW1 inserts. Second, a bias in favor of synonymous substitutions in both ORFs of the SW1*Ol* isolates suggested that those sequences, or the lineages they represent, have diverged under the

constraints of purifying selection. Third, two of the four inserts from the medaka genome exhibited intact ORFs.

A phylogenetic analysis of the reverse transcriptase amino acid residues of SW1 in medaka and pupfish suggests that the SW1 family is monophyletic and a sister group to the L1 family of mammals. The relationship between SW1 and L1 was further supported by deduced amino acid similarity in their respective ORF I sequences.

Deduced amino acid sequences are quite divergent (44.2%) between SW1 RT sequences isolated from medaka and pupfish (classified in separate Orders of the Division II Atherinomorpha, Parenti, 1993), exceeding that observed between RT sequences of mouse and human (38.0%). The reciprocal monophyly of these two element families combined with the level of sequence divergence among members within each family suggests that the SW1 and L1 retrotransposons are ancient components of their respective genomes. These observations are consistent with vertical transmission of the SW1 and L1 element families and do not support the notion of recent horizontal transmission between fish and mammals.

The evolutionary relationship of SW1 and L1 families suggests that SW1/L1-like elements may be a general feature of vertebrate genomes. The availability of SW1 sequences will allow the identification of conserved regions in these families of elements between teleost and mammalian groups, respectively, making it possible to search more efficiently for related element families in other vertebrates using a degenerate PCR primer approach like that applied in this study and others (Wichman & Van Den Bussche, 1992; Tristem, 1996; Ring *et al.*, 1996).

Despite the demonstrated phylogenetic relationship of SW1 and L1 sequences, the genomic composition of these two families differs in some respects. The most notable of these is the copy number of SW1 and L1 in their respective genomes. The L1 family occurs in mammals at 10^4 to 10^5 copies per genome (Hutchison *et al.*, 1989), and can account for as much as 10% of the genome (D'Ambrosio *et al.*, 1986). In contrast, the copy number of SW1 (≈10-20 per genome) is more like that of some other non-LTR containing elements such as Cin4 in maize (Schwarz-Sommer *et al.*, 1987), reported at 50-100 copies per genome, or G in *Drosophila* (Di Nocera, Graziani & Lavorgna, 1986), estimated at 20 copies per genome.

In addition to differences in copy number, the SW1 family in medaka also differs from typical L1 sequences in the apparent absence of inserts truncated at their 5' ends. Such truncation is widespread among non-LTR elements including L1 (but not L1*Rn*, see D'Ambrosio *et al.*, 1986), and apparently results from premature termination of reverse transcription during the insertion of element copies (Eickbush, 1994). Among some L1 families, it is estimated that 90% of all insertions may be truncated in this fashion (Hutchison *et al.*, 1989).

Numerous studies have been devoted to characterizing the complex subfamily dynamic relationships of L1 elements within mammalian genomes (Cabot *et al.*, 1997; Casavant & Hardies, 1994a; Casavant *et al.*, 1996; Clough *et al.*, 1996; Furano *et al.*, 1994; Jurka, 1989; Kass *et al.*, 1992; Smit *et al.*, 1995; Vanlerberghe *et al.*, 1993) and the regulatory mechanisms involved in controling L1 transpositional activity (Furano *et al.*, 1988; Hata & Sakaki, 1997; Howell & Usdin, 1997; Nur *et al.*,

1988; Swergold, 1990; Woodcock *et al.*, 1997). Difference between SW1 and L1 families in copy number and insert composition may indicate fundamental differences between the two related families in element-host genome regulatory interactions or in element transpositional activity, efficiency or insertional specificity.

Sequence Availability

The DNA sequences reported in this paper have been deposited in the GenBank database under accession numbers AF055639-AF055643.

Chapter 3: Variation and divergence of Death Valley pupfish populations at retrotransposon-defined loci

Introduction

The Death Valley pupfishes (Cyprinodontidae: *Cyprinodon* sp.) comprise a monophyletic assemblage of relict populations occupying the remnant aquatic habitats of Pleistocene Lake Manly and the Amargosa River drainage system in southern California and Nevada (Echelle & Dowling, 1992; Chapter 1). Isolation of several of these populations dates to the middle to late Pleistocene, as the region became more arid and this once-extensive system of lakes and rivers became fragmented (Soltz & Naiman, 1978; Miller, 1981). Several of the spring habitats were isolated only within the past several hundred to few thousand years (Soltz & Naiman, 1978).

Comparative studies have revealed apparently rapid and marked divergence in morphological features among many of the populations (Miller, 1948; LaBounty & Deacon, 1972). Accordingly, this group has been divided into three species, *C. nevadensis*, *C. diabolis*, and *C. salinus*, and nine subspecies, seven of which are extant. However, despite their phenotypic distinctiveness, knowledge of the levels of molecular genetic diversity and divergence among the populations is less complete. This information is important for our understanding of the processes involved in the divergence of the Death Valley populations, and is critical for the development of management strategies for these populations, many of which are threatened or endangered.

Studies of the genealogical relationships of mtDNA haplotypes, isolated in separate populations, suggested that population divergence with respect to the mtDNA genome has occurred largely through stochastic sorting of ancestral variation (Echelle & Dowling, 1992; Chapter 1). Overall, diversity of mtDNA haplotypes within populations was limited or absent. However, the level of divergence among haplotypes within some of the populations that were polymorphic suggested that they have maintained widely varying effective population sizes, and that ancestral populations were perhaps much larger than extant populations.

Less is known about the level of nuclear genetic diversity and divergence exhibited among the populations. Previous attempts at measuring divergence by allozyme techniques were confounded by a general lack of detectable variation at those loci (Turner, 1974, 1983; Echelle & Echelle, 1993). In general, the effective population size for the nuclear genome should be four-fold greater than that for the mtDNA genome as a consequence of the maternal haploid inheritance pattern of the later genome (Wilson *et al.*, 1985). Therefore, we might predict that population divergence, on average, should be more limited in the nuclear genome with a larger portion of the genetic variation retained within populations. Such patterns of variation would facilitate more detailed analyses of the relative historical effective population sizes of pupfishes in the various isolated habitats, and might reveal details regarding population relationships that are indiscernible by analysis of the mtDNA genome.

Transposable elements

The transposable elements, or transposons, are a diverse group of endogenous mobile genetic elements which constitute a prominent component of eukaryotic genomes (Finnegan, 1992; McDonald, 1993). When transposable elements are mobilized, they generate mutations. These mutations have been implicated as a potential source of novel genetic variation in evolutionary processes (see Kidwell & Lisch, 1997, for a recent review). Some studies have indicated that transposon mobilization may be facilitated by extrinsic factors often associated with small, peripheral populations, such as inbreeding and environmentally-induced stress (Fontdevila, 1993). Consequently, transposable element variation is of particular interest because of the role it may have played in the divergence of isolated populations such as those of the Death Valley pupfishes.

SW1 is a low copy number retrotransposable element family that has been identified and characterized in the genomes of the desert pupfish (*Cyprinodon macularius*) and the Japanese medaka (*Oryzias latipes*) (Chapter 2). A Southern blot of DNA samples from a single population of pupfish from the Colorado River basin (*C. macularius*), when probed with a portion of the desert pupfish derived-SW1 sequence, revealed extensive genetic variation manifested by numerous polymorphic bands. Similar levels of variation were detected in some Death Valley pupfish populations, suggesting that genetic variation associated with the SW1 element family would be informative for population studies in this ecosystem.

The goals of this study were two-fold:

i. To assess the level of genetic variation associated with the SW1 element family in natural populations, and to infer whether mutational processes associated with the SW1 family have generated novel genetic variation among populations in isolation.

ii. To use SW1-associated variation to assess the genetic distinctiveness of pupfish populations, and to attempt to evaluate the relative importance of duration of isolation and effective population sizes in driving population divergence.

Materials and Methods

The survey strategy employed was to identify restriction enzymes which apparently do not cleave within the SW1 element sequence. A Southern blot hybridization analysis of the resulting restriction endonuclease fragments using a portion of the SW1 element as the probe results in a series of bands, each composed of all, or a portion, of the SW1 insertion element and a variable length of unique, site specific, flanking genomic DNA at one or both ends of the insert (Fig. 3.1). The variation that is detected among individuals may result from a combination of the presence/absence of insertion elements at individual genomic sites or from restriction fragment length polymorphisms (RFLPs) in the flanking genomic DNA.

DNA samples were obtained from 16 populations distributed among three geographical regions (Salt Creek drainage, Amargosa River drainage and Ash Meadows). Collection locations and population identification are given in Figure 1.1. Collections were made in April, 1994 under permits listed in the acknowledgment section. All specimens were pithed, degutted and fixed in absolute methanol in the field. Fixative was changed at least twice in the 36-48 hrs following initial fixation.

DNA analysis

Ten-microgram aliquots of DNA, extracted from whole body muscle tissue, were digested with 15-20 units of the appropriate restriction enzyme (*Pvu* II or *Hin*d III) following supplier specifications. Digested DNAs were size fractionated on a 1% agarose gel (1X TAE) and alkaline-

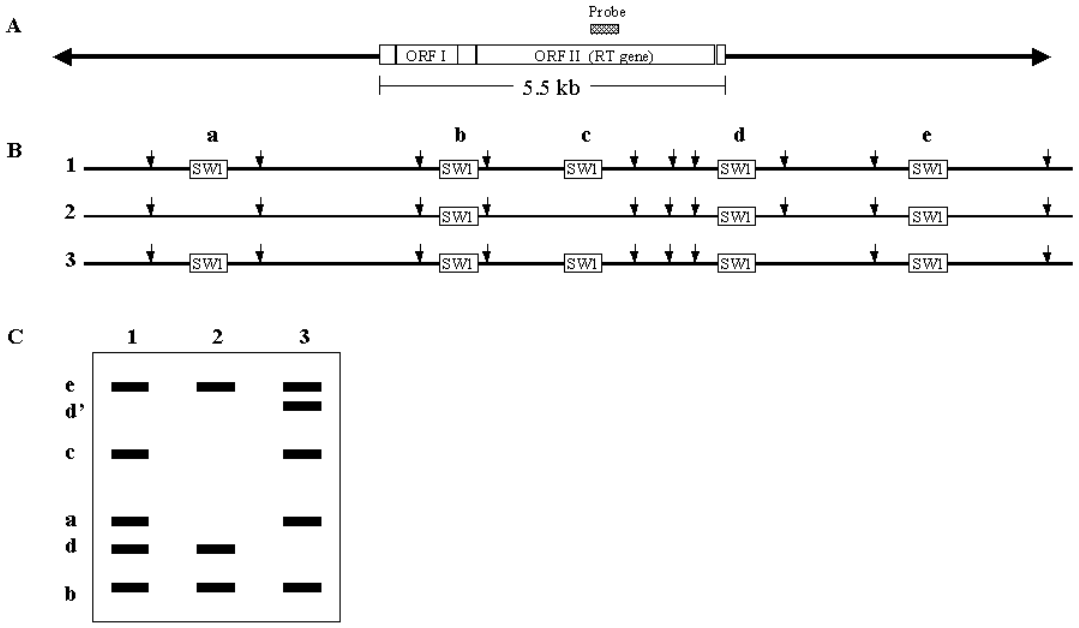

Fig. 3.1 Diagram of the restriction fragment length variation of the SW1 elements and their associated flanking genomic DNAs. A, Organization of the SW1 element sequence (ORF = open reading frame, RT = reverse transcriptase). The probe is derived from a portion of the RT gene. B, Illustration of variation among three hypothetical individuals (arrows indicate restriction sites). Individuals 1 and 2 differ by the presence/absence of inserts a and c; Individuals 1 and 3 differ by the presence/absence of a restriction site flanking insert d (absence indicated by an asterisk). C, Hypothetical band patterns from a Southern blot DNA hybridization of DNAs from individuals 1, 2 and 3.

transferred *in vacuo* to nylon membrane (Hybond-N, Amersham, Arlington Heights, IL) according to standard protocols (Ausubel *et al.*, 1989). Southern blots were hybridized to radiolabeled probes (Random Primers DNA Labeling System, Gibco BRL, Gaithersburg, MD) derived from a 309 bp region of the reverse transcriptase gene (Chapter 2). Nylon membranes were hybridized overnight at 65°C following the procedures of Church and Gilbert (1984), and then washed three times with 2X SSC, 0.1% SDS at 65°C. Hybridized membranes were exposed to Kodak BioMax films (Rochester, NY) for 48 hours at -80°C with intensifying screens.

Typically, thirty-two specimens were analyzed on each gel, accompanied by an *Eco*R I-digested lambda DNA size marker on the outer lanes. In the first round of analysis, samples from two or more populations were surveyed on each gel. Once all populations were assayed, composite gels were constructed which contained individuals from multiple populations so that cross-gel comparisons could be made more readily. Three fixed bands (bands A, V, & X, Table 3.1), which were present in virtually every individual from all populations, served as internal controls for detecting any electrophoretic anomalies.

Data analysis

The electrophoretic conditions gave maximum resolution of bands ranging in size from approximately 0.5 to 35 kb. Bands within this size range were scored as present [1] or absent [0] for each individual. A similarity matrix of all individuals then was constructed with GELSTATS (ver. 2.61, Pelikan & Rogstad, 1996; Rogstadt & Pelikan, 1996) using a band-sharing index of similarity [$s=(2*N_{XY})/(N_X+N_Y)$, where N_{XY} is the

Table 3.1 Population frequencies for twenty-six scored SW1 hybridization bands[1], and sample sizes for sixteen populations[2] of Death Valley Pupfishes.

	Population number															
Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
B	1.000	1.000	1.000		0.068	0.091	0.141	0.345				0.111	0.333	0.647	0.091	
C					0.023	0.091	0.107			0.452	0.462	0.111	0.250			
D				0.324	0.182	0.333	0.250	0.793	0.481	0.419		0.444	0.083		0.364	
E			0.240		0.091	0.061	0.071		0.037			0.222	0.500		0.091	
F												0.222	0.083			
G			0.440	0.297	0.023	0.121	0.143		0.037			0.444	0.167	0.588	0.091	
H				0.162		0.061	0.036					0.111	0.083			
I	1.000	1.000	1.000										0.167			
J				0.135	0.159	0.091	0.107	0.793	0.074			0.333	0.250			
K				0.027	0.023											
L				0.351	0.045						0.115				0.091	
M				0.081	0.227	0.091	0.143		0.185	0.097	0.423	0.333	0.500		0.273	
N	1.000	1.000	1.000	0.486	0.136	0.152	0.107	0.690	0.148	0.613	0.692	0.111		0.647	0.182	
O				0.189	0.182	0.152	0.107		0.037			0.222	0.333			1.000
P				0.541	0.636	0.576	0.571	0.931	0.593			0.111	0.500		0.455	
Q			1.000									0.333				
R					0.023	0.030										
S				0.135	0.114	0.212	0.143		0.037			0.555	0.500			
T					0.023											
U									0.037							
V	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
W			0.280							0.032						
X	1.000	1.000	0.880	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Y				0.162	0.182	0.212	0.321	0.724	0.148	0.387	0.231	0.555	1.000		1.000	1.000
Z				0.243												
Sample Size	15	16	25	37	44	33	28	29	27	31	26	9	12	17	11	10

[1]**The restriction enzyme used was** *Pvu* **II**

[2]**Populations are identified in Figure 1.1.**

number of bands shared between individuals X and Y, and N_X and N_Y are the number of bands in individuals X and Y respectively; Nei & Li, 1979; Lynch, 1990, 1991; Rogstad & Pelikan, 1996]. A measure of pairwise genetic distance among individuals was calculated as 1-s.

Relative genetic diversity among populations was estimated by pairwise similarity tests using GELSTATS. These tests determined whether paired populations differ in the average within-population, inter-individual similarity (Rogstad & Pelikan, 1996). The significance of these comparisons were determined using a non-parametric permutational test (1000 randomizations) performed by GELSTATS.

An analysis of molecular variance (AMOVA, ver. 1.55, Excoffier *et al.*, 1992) was applied to the distance matrix to estimate variance components and population-pairwise distance measures (Φ_{st}). The AMOVA procedure utilizes an analysis of variance format to reveal the patterns of intra-specific genetic structure among populations from distance metrics derived from molecular data (Excoffier *et al.*, 1992). The significance of all population statistics was determined using a non-parametric permutational procedure (500 randomizations) performed by AMOVA. The utility of AMOVA for analyzing phenotypic patterns from multi-locus markers, such as RAPDs, has been well demonstrated by others (e.g., Huff *et al.*, 1993; Stewart & Excoffier, 1996; Gillies *et al.*, 1997).

A multidimensional scaling (MDS, Kruskal & Wish, 1978) analysis was conducted on the pairwise population estimates of Φ_{st} (analogs of F_{st}, estimated by AMOVA) for both the SW1 data presented here and

previously reported mtDNA D-loop sequence data (Chapter 1). The MDS analysis provides a means of spatially arranging the matrix of Φ_{st} values in a configuration of points which visualizes the relative genetic distances among populations. The "Stress" value associated with each MDS analysis is the square root of the normalized "residual sum of squares" (Kruskal & Wish, 1978) and provides a measure of the "goodness-of-fit" for each analysis. The MDS analyses were conducted using the MDS module of STATISTICA (release 4.3, StatSoft, 1994) with the "Standard Guttman-Lingoes" starting configuration. Populations which were indistinguishably monomorphic were pooled for the MDS analyses.

Results

A Southern blot DNA hybridization analysis of 370 individuals from sixteen Death Valley pupfish populations, using *Pvu* II as the restriction enzyme, uncovered approximately 35 to 40 distinguishable bands in total. These bands varied from about 0.5 to roughly 75 kb in length. The number of bands observed in each individual ranged from 4 to 15 with the mean number of bands per population ranging from 6 (populations 1, 2, 15 & 16) to 9.5 (population 14). The hybridizing DNA fragments revealed extensive genetic variation in many of the populations, and a complete absence of genetic variation in others (Fig. 3.2a,b); nearly all bands were polymorphic in the majority of the populations where they were found, with only three bands appearing fixed or nearly fixed in all sixteen population samples. A more limited survey of some populations (populations 1, 2, 3, 10, 14 and 15), using *Hin*d III as the restriction enzyme, revealed equivalent banding patterns with similar levels of genetic variation.

Distribution and frequency of Bands

Twenty-six of the *Pvu* II-generated bands (ranging in size from ≈ 0.5 to 35 kb) were scored; their population frequencies are indicated in Table 3.1. Fixed bands A and V were uninformative for discriminating population structure, and consequently have been excluded from the analyses that follow. Most polymorphic bands were distributed among multiple populations; only three bands (T, U & Z) were unique to single samples, and three others (F, K, & R) appeared localized in pairs of

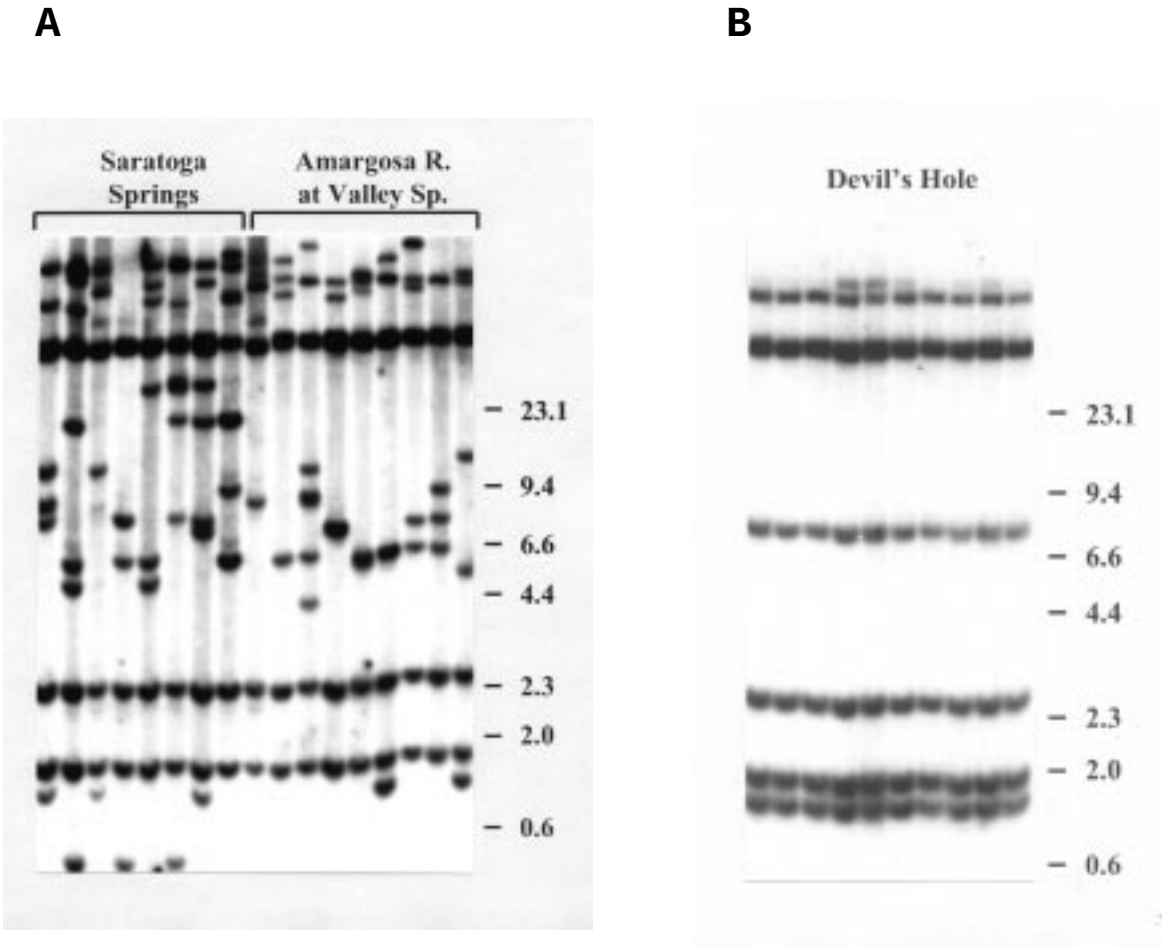

Fig. 3.2 Southern blot hybridization surveys of pupfish specimens from A, Saratoga Springs (4) and the Amargosa River at Valley Springs (5) and B, Devil's Hole (16). The extensive variation revealed among individuals from Saratoga Springs and the Amargosa River contrasts sharply with the absolute band identity detected among individuals from Devil's Hole. Samples were prepared with enzyme *Pvu* II and hybridized with a 309 bp probe derived from the SW1 reverse transcriptase gene.

geographically adjacent samples. Fixed differences between populations were rare, though notable fixed differences were identified between Salt Creek/McLean Spring (1,2) and Cottonball Marsh (3) populations (band Q)(Fig. 3.3a), and between adjacent School (14) and Indian (15) Springs populations (band Y). The fixed difference between Salt Creek/McLean Spring and Cottonball Marsh populations was also demonstrable using *Hin*d III as the restriction enzyme (Fig. 3.3b).

Genetic Diversity

The number of polymorphic bands present within populations varied markedly among many of the populations. At one extreme were populations at Devil's Hole (16) and Salt Creek/McLean Spring (1,2), in which all individuals were identical. At the other extreme were populations from Saratoga Springs (4), Big Spring (12), and some of the Amargosa River sites (5,6, & 7 in particular), in which banding patterns were highly complex and virtually individual-specific. These differences were quantified based on pairwise individual similarities and on probabilities of identity among individuals within populations (Rogstadt & Pelikan, 1996). Both measures of band complexity amount to estimates of relative genetic diversity among the respective populations. Permutational tests of pairwise similarities among individuals within populations revealed that population band complexity estimates could be separated into several distinct clusters that were significantly divergent ($P < 0.05$, Table 3.2).

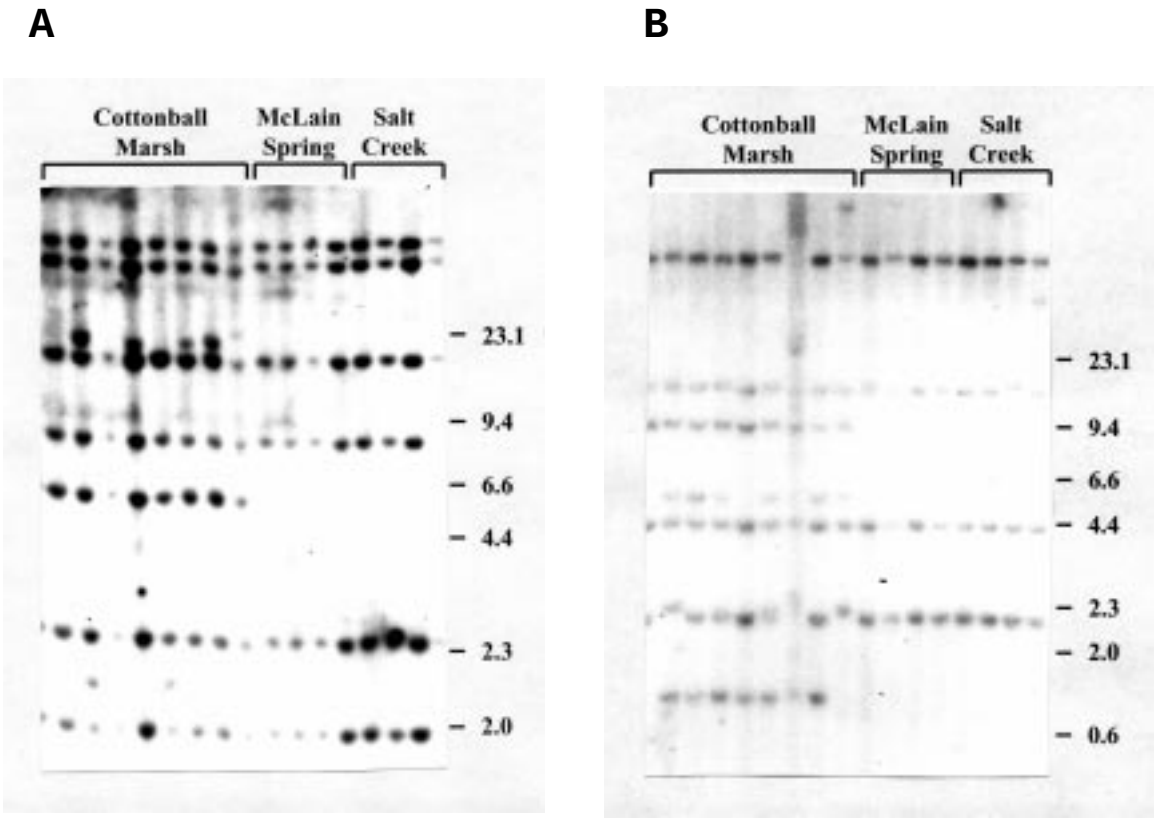

Fig. 3.3 Southern blot hybridization surveys of pupfish specimens from Cottonball Marsh (3), Salt Creek (2) and McLain Spring (1) using A, *Pvu* II, and B, *Hin*d III restriction enzymes. Fixed differences between Cottonball Marsh (*C. s. milleri*) and Salt Creek/McLain Spring (*C. s. salinus*) are observed with both restriction enzymes. Both blots were hybridized with a 309 bp probe derived from the SW1 reverse transcriptase gene.

Table 3.2 Rank order of genetic diversity among Death Valley pupfish populations[1] based on intrapopulation interindividual similarity. Statistical significance of differences in similarity measures, as determined by permutational tests (1000 iterations, $P < 0.05$), are indicated by bars connecting population numbers.

Population	16	1	2	3	8	14	11	15	10	9	5	7	6	13	4	12
Mean Sim.	1.00	1.00	1.00	0.90	0.85	0.84	0.79	0.78	0.77	0.76	0.71	0.69	0.68	0.67	0.65	0.59
Prob. Id.	1.00	1.00	1.00	0.45	0.30	0.44	0.32	0.26	0.28	0.26	0.17	0.14	0.13	0.05	0.07	0.02

[1]**Population numbers are as described in Figure 1.1.**

Population Structure

Estimates of variance components within populations, among populations within drainage areas (ie. Salt Creek, Amargosa River, and Ash Meadows), and among grouped populations among drainage areas, revealed that most of the variation was distributed within populations (54%; P < 0.002). The next largest portion of the variance was attributed to variation among drainage areas (31%; P < 0.002), with the smallest portion of the variance occurring among populations within drainage areas (15%; P < 0.002). These results contrasted with those for mtDNA D-loop sequences (Chapter 1) where the majority of the variance (62%; P < 0.002) was partitioned among drainage areas, with the smallest portion of the variance occurring within populations (12%; P < 0.002).

Pairwise estimates of Φ_{st} (Table 3.3) were employed to resolve regional and population genetic structure from the background of within-population variation. An MDS analysis of these distance measures revealed extensive fine scale genetic structure. For comparative purposes, an MDS analysis was also performed on pairwise Φ_{st} estimates reported for mtDNA D-loop sequence data from the same population samples (Table 1.2).

MDS analyses of both SW1 and mtDNA data sets (Fig. 3.4) revealed that populations within drainage regions form distinct, non-overlapping clusters. Genetic structure among many populations within regions was also readily apparent with both genetic markers. However, the SW1 nuclear marker analysis revealed genetic structure among several populations that proved indistinguishably monomorphic with respect to

mtDNA haplotypes. These included populations in the Salt Creek drainage (1, 2 & 3), Ash Meadows populations at Indian Spring (14) and School Spring (15), and the Shoshone Spring (10) and adjacent Amargosa River populations (11). Also noteworthy was the *C. n. amargosae* **population**

Table 3.3 Pairwise estimates of Φ_{st} among Death Valley pupfish populations[1].

Populations:	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1,2	0.638	0.496	0.533	0.526	0.556	0.761	0.624	0.666	0.689	0.690	0.714	0.633	0.787	1.000
3		0.506	0.553	0.543	0.566	0.731	0.630	0.665	0.683	0.581	0.661	0.578	0.723	0.885
4			0.078	0.052	0.073	0.261	0.097	0.181	0.187	0.146	0.246	0.315	0.186	0.446
5				NS	NS	0.287	0.028†	0.104	0.089	0.175	0.202	0.390	0.188	0.419
6					NS	0.258	NS	0.062	0.103	0.110†	0.181	0.357	0.151	0.417
7						0.278	NS	0.048†	0.093	0.122	0.167	0.370	0.125	0.424
8							0.304	0.370	0.470	0.397	0.437	0.615	0.353	0.739
9								0.080	0.179	0.210	0.300	0.459	0.189	0.532
10									0.092	0.252	0.286	0.523	0.213	0.506
11										0.323	0.319	0.533	0.348	0.501
12											0.076†	0.411	0.169	0.504
13												0.517	0.164	0.481
14													0.543	0.795
15														0.582

[1]Population numbers are as indicated in Figure 1.1.
NS, estimates not significantly different from zero in randomization tests ($p < 0.05$).
† estimates not significantly different from zero when critical values were adjusted (overall $\alpha = 0.05$) using the sequential Bonferroni correction (Rice, 1989) for multiple tests.

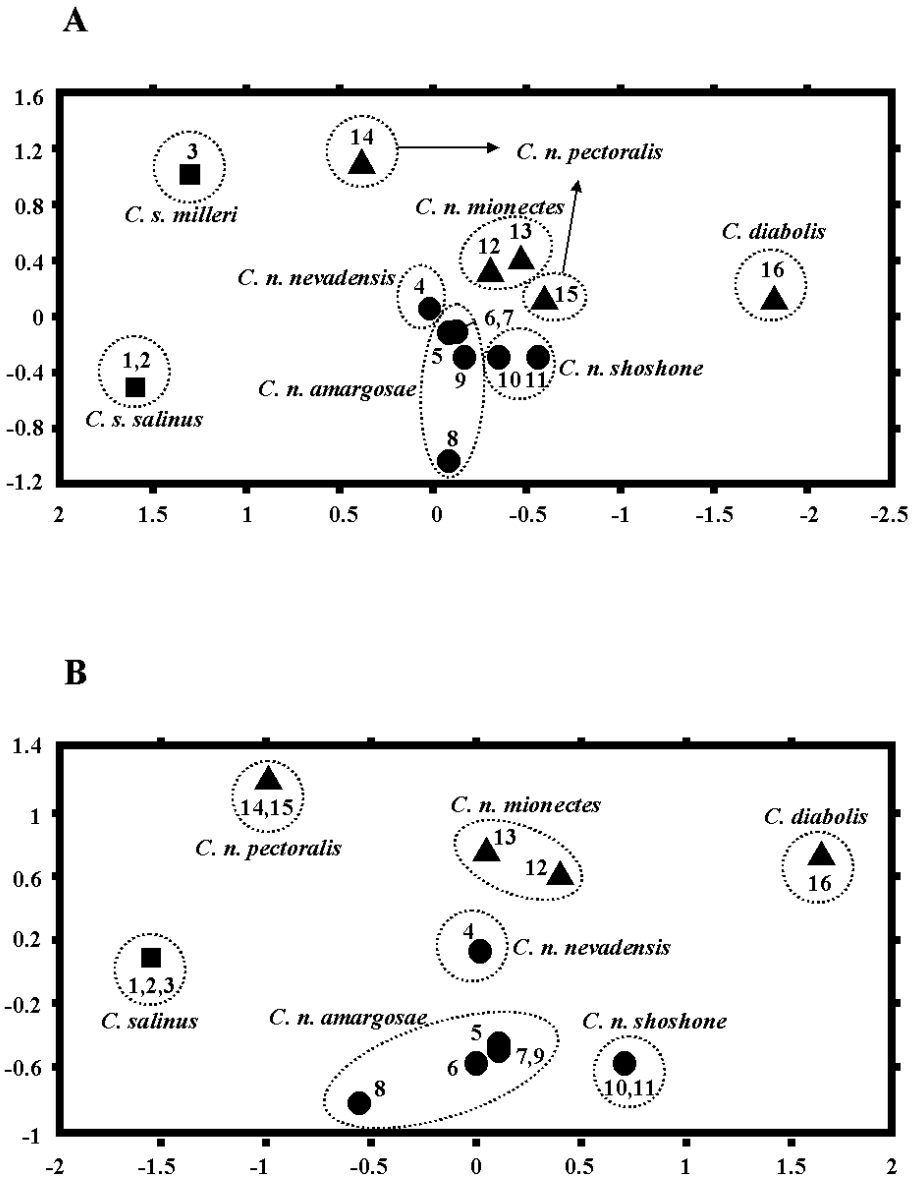

Fig. 3.4 A Multi-dimensional Scaling analysis of Φ_{st} values generated for, A, SW1 banding patterns (stress = 0.117), and B, mtDNA D loop nucleotide divergence (stress = 0.171). Populations which were indistinguishably monomorphic were pooled in both analyses. Symbols are as follows: Square, Salt Creek drainage; circle, Amargosa River drainage; triangle, Ash Meadows. The X-axis was reversed for the SW1 MDS analysis in order to facilitate comparisons between the two data sets. The magnitudes and significance of SW1 generated Φ_{st} values are given in Table 3.3; Φ_{st} values for mtDNA sequences are from Table 1.2.

isolated in the unnamed spring north of Tecopa (8), that appeared more divergent from the remaining *C. n. amargosae* **populations than were populations of either** *C. n. nevadensis* **(4) or** *C. n. shoshone* **(10,11).**

Discussion

The DNA hybridization banding patterns of the SW1 element family revealed extensive genetic variation within and among the pupfish populations of Death Valley. These patterns appeared to be highly individual-specific within a number of the populations, with levels of band complexity rivaling those typically observed when other multilocus DNA screening techniques, such as mini- (Jeffreys *et al.*, 1985) or micro-satellite (Laughlin and Turner, 1994, 1996) "VNTR fingerprints", are employed. An analysis of molecular variance revealed that most of the SW1 variation was distributed within populations. Population or region-specific bands, either fixed or polymorphic, appeared rare. Consequently, much of the SW1 variation observed in extant pupfish populations may have been ancestral in origin.

Variation in Effective Population Sizes

Death Valley pupfish populations varied most dramatically in the levels of genetic diversity exhibited in their respective SW1 hybridization banding patterns. The striking contrasts among some populations in band complexity are readily apparent upon casual inspection of the respective Southern blots (Fig. 3.3a,b). Given that most of the variation is distributed within populations, it appears that the differences in intrapopulation diversity have resulted from variable rates of genetic drift at the SW1-defined loci as mediated by contemporary or historical differences in effective population sizes, and less so as a result of mutational events within isolated populations. The possibility that selection has operated at these SW1-

defined sites is not ruled out. However, it is assumed that, on average, any selective forces imposed at insertion element sites should be equivalent across populations.

Comparisons of the level of genetic diversity present within populations revealed that the highest SW1-associated genetic diversity was found in the relatively large spring populations at Ash Meadows and Saratoga Springs, and in the much larger but highly fluctuating riverine populations of the Amargosa River. Such comparable levels of genetic diversity in these contrasting habitats may indicate that long term effective population sizes have been similar. These spring and river populations also exhibited mtDNA polymorphisms (Chapter 1). In contrast, other populations that were monomorphic for mtDNA haplotypes exhibited lower levels of SW1 variation as well. At the extreme were those populations that lacked variation all together. Among those was the Devil's Hole pupfish (*C. diabolis*), found in an isolated fault depression elevated above Ash Meadows. This population generally has fewer than 500 individuals and has been isolated for longer than most other Death Valley populations (≈10-20,000 years, Soltz & Naiman, 1978). The only other completely invariant population samples were of *C. s. salinus* at McLean Spring and its associated outflow at Salt Creek. These are discussed below.

Genetic structure and population divergence

Analysis of the SW1 variation described here, in combination with recently reported mtDNA data (Chapter 1), provides the most detailed description of molecular genetic diversity and structure among Death

Valley pupfish populations thus far available. In addition to providing insights to the influences involved in the divergence of Death Valley pupfishes, these data are critical for discerning demographically independent units which are important for establishing conservation priorities (Moritz, 1995) in a group containing several threatened and endangered species (Minckley & Deacon, 1991).

The SW1 data has resolved the presence and magnitude of genetic structure among some populations where mtDNA sequence data proved uninformative due to limited or absent genetic variation. For example, populations from McLean Spring, Salt Creek and Cottonball Marsh are indistinguishably monomorphic for a diagnostic mtDNA D-loop haplotype (Chapter 1). Although population samples from McLean Spring and Salt Creek (*C. s. salinus*) were also indistinguishable by their invariant SW1 hybridization band profile, these data demonstrated extensive population structure between Salt Creek/McClain Spring and Cottonball Marsh. A fixed difference in DNA hybridization banding patterns (Fig. 3.4) is the first molecular evidence of the possible genetic distinctiveness of these populations, which have been isolated for only a few thousand years (LaBounty & Deacon, 1972). The difference is demonstrable with both restriction enzymes, suggesting that it represents the presence/absence of an SW1 copy at a specific genomic site, and not simply restriction site divergence in flanking genomic DNAs.

The Salt Creek/McLean Spring populations were the only samples, besides that from Devil's Hole, to exhibit a complete absence of variation in SW1 banding patterns using the *Pvu* II enzyme. Genetic

variation was also limited in the Cottonball Marsh population (Table 3.2). These results are quite striking since the populations in McLean Spring and Cottonball Marsh may be among the largest of the Death Valley pupfishes. The lack of genetic diversity in SW1 banding patterns, in addition to the mtDNA monomorphism, may point to one or more historical bottleneck events.

The Indian and School Springs populations (*C. n. pectoralis*) were also quite divergent with respect to SW1-associated banding patterns (Φ_{st} = 0.73, Table 3.3) despite being monomorphic for identical mtDNA haplotypes (Chapter 1). These populations have probably been isolated for no more than a few hundred years and are among the smallest in Death Valley (fewer than two-hundred individuals each) (Soltz & Naiman, 1978). This substantial genetic divergence over a brief time interval was likely facilitated by rapid genetic drift as a consequence of the small effective population sizes imposed in these habitats. In contrast, the much larger populations at Big and Point of Rocks Springs (census sizes ≈ 10,000 individuals) were not significantly divergent for SW1-associated banding patterns though they likely have been isolated for at least as long as School and Indian Springs (Soltz & Naiman, 1978).

There is generally less divergence among populations in the Amargosa River drainage, particularly among river populations that may undergo occasional genetic exchange. Perhaps surprisingly, the most extensive divergence in this region was not among the nominal subspecies. Instead, the most divergent population was one of *C. n. amargosae* in a small, unnamed spring only partially isolated from the

adjacent river population. Concordantly, the estimated relative genetic diversity in this small spring population was among the lowest in Death Valley. This observation supports the supposition that genetic divergence among these populations has been driven primarily through loss of genetic variation via drift.

The MDS analysis of population structure with respect to SW1 variation (Fig. 3.4a) revealed that the most divergent populations were always those with the lowest genetic diversity (specifically populations 1, 2, 3, 8, 14 & 16), suggesting that effective population size may be the most critical factor in driving molecular divergence at these loci. Populations with intermediate to high levels of genetic variation were generally less divergent, regardless of the hydrological relationships of the habitats or presumed length of isolation time. However, despite the more limited divergence among the later populations, spatial relationships on the MDS configuration bear some resemblance to the hydrological relationships of the habitats (Fig. 1.1) and the inferred history of their connections. The Amargosa River drainage populations and Ash Meadows populations display more or less distinct clusters, and the spatial relationships of populations from the Amargosa River reflect their linear hydrological relationships along the river channel consistent with isolation by distance.

The MDS analysis of population structure with respect to mtDNA D-loop sequence variation (Fig 3.4b) reveals evidence of much greater genetic structure among these later populations. This may be attributed to the lower effective population size of the mtDNA genome resulting in a higher extinction rate for the mtDNA lineages, and the consequent

larger portion of variance distributed among, rather than within, populations.

The Role of Retrotransposable Elements in Generating Variation

This study provides little, if any, evidence that novel genetic changes have accrued through retrotransposition or any other mutational processes in any of the populations in the Death Valley drainage system subsequent to their isolation. Nonetheless, extensive variation at these SW1-defined loci may indicate that (1) retrotranspositional activity before the isolation of extant populations has been a source of genetic variation in this system, or (2) SW1 elements are found in regions of the pupfish genome that are highly variable for other reasons. There is a repeated suggestion in the literature (e.g., Fontdevila, 1993) that transposable elements may serve an important function as mutation generators in natural populations (reviewed by Kidwell and Lisch, 1997). This study has revealed that informative population level genetic variation can be found by surveying DNA regions associated with inserted transposable elements such as SW1.

Conservation Genetics of the Death Valley Pupfishes

The primary goals of conservation biology include the preservation of population genetic diversity and the maintenance of natural evolutionary processes (Avise, 1994). These goals are best accomplished when management practices are guided by a detailed knowledge of the geographic patterns of genetic variation in the populations that are being managed (Echelle, 1991). Unfortunately, this basic information is often unavailable. Such has been the case for the pupfish populations of Death Valley.

Comparative studies have provided a detailed knowledge of the morphological variation (Miller, 1948; LaBounty & Deacon, 1972), courtship behaviors (Liu, 1969), and heritable differences in physiological tolerances (reviewed by Feldmeth, 1981, and Soltz & Hirshfield, 1981) among these populations. However, basic information regarding the distribution and levels of genetic variation within and among the same populations repeatedly has eluded researchers (Turner, 1974, 1982; Echelle & Dowling, 1992; Echelle & Echelle, 1993).

What significance is this study to the conservation biology of the pupfishes of Death Valley?

The mitochondrial DNA and the SW1 Southern blot hybridization banding pattern data presented here provide two critical contributions to the conservation biology of the Death Valley pupfishes. First, these data have revealed the genetic relationships and extent of divergence that exists among the Death Valley pupfish populations. Secondly, these

data reveal the relative levels of genetic variation present within the populations.

This new molecular perspective provides a resolution to several long standing issues:

i. Ash Meadows Populations: **The mtDNA and SW1 data, in combination, establish that each of the Ash Meadows populations exhibits a unique, divergent gene pool. The two** *C. n. mionectes* **populations each contained a unique complement of mtDNA haplotypes which distinguished them, and the two** *C. n. pectoralis* **populations were distinguishable by a single fixed molecular hybridization band difference in their SW1 patterns. Moreover, these two data sets illustrate that genetic diversities vary markedly among the Ash Meadows populations and are consistent with variation in contemporary population sizes as dictated by the sizes of their respective habitats.**

ii. Amargosa River Populations: **The extent of gene flow among populations along the Amargosa River and its tributaries has been in question for some time (Miller, 1948), particularly where habitats are connected only intermittently during occasional periods of flash flooding. The molecular data presented here provide only marginal evidence of genetic structure among most of the** *C. n. amargosae* **populations with the exception of a single population isolated in a small unnamed spring located north of Tecopa. This population apparently has diverged from neighboring Amargosa River populations via drift as a consequence of its small population size.**

iii. Salt Creek drainage: **The first molecular evidence for the distinctiveness of the McLean Spring/Salt Creek (***C. s. salinus***) and**

Cottonball Marsh (*C. s. milleri*) populations is provided by the fixed difference in molecular hybridization banding patterns revealed by SW1 analyses. Additionally, molecular data demonstrated a surprisingly low level of genetic variation in all populations of the Salt Creek drainage. The most parsimonious explanation for this lack of genetic variation is that these populations have undergone one or more intense bottleneck events.

The Shoshone Spring Pupfish.

An important question that is not addressed by the mtDNA and SW1 data is the origin of extant pupfish populations in the historic *C. n. shoshone* habitats. Populations of the nominal subspecies *C. n. shoshone* (Miller, 1948) were considered extinct in the late 1960s (Pister, 1974). However, in 1986 pupfishes were found to be quite common in these habitats. A morphological assessment of the rediscovered populations indicated that they were more similar to the *C. n. shoshone* subspecies that once occupied the Shoshone Spring habitat than to adjacent *C. n. amargosae* populations occupying the Amargosa River at Tecopa (Taylor *et al.*, 1988; Taylor & Pedretti, 1994). It was concluded that *C. n. shoshone* populations were never extinct but simply survived for a time at a reduced population density. However, there were some discrepancies in the morphological comparisons between the extant population and museum specimens that pre-dated the disappearance of the populations. Furthermore, the possibility that these populations are subject to ecophenotypic plasticity cannot be ruled out. Consequently, the taxonomic designation of extant populations at Shoshone Spring remains in contention.

This issue might be resolvable if a diagnostic molecular marker could be identified that would differentiate the Shoshone populations from all other Amargosa River populations. Unfortunately, no diagnostic mtDNA haplotypes or SW1 band profiles were identified in any of the population samples from the Amargosa River drainage, let alone those from Shoshone Springs. These analyses did reveal significant differences in haplotype frequencies and SW1 molecular band profiles between Shoshone populations and neighboring Tecopa and Tecopa Bore populations, and the Shoshone populations also exhibited significantly lower genetic variation than the neighboring populations with both markers. However, these differences were not diagnostic. Whether they are indicative of extended isolation between populations nominally designated as separate subspecies or simply have resulted from a founder event associated with the reinvasion of the Shoshone habitat by a neighboring *C. n. amargosae* **population remains unresolvable.**

Application of molecular data to the delineation of Evolutionarily Significant Units

Understanding of some of the most interesting, and perhaps the most important components of the genetic distinctiveness of Death Valley pupfish populations remain elusive. These components include the genetic variation that underlies the potentially adaptive phenotypic differences that exist among populations. At present, we are largely limited in our ability to identify and interpret patterns of variation at the molecular level to those that have resulted through the stochastic processes of drift and mutation (Allendorf, 1995).

Among the Death Valley pupfish populations, the patterns of divergence seem to have resulted predominantly from drift, with little evidence for the acquisition of novel genetic variation that may have been generated through mutational processes in the isolated populations. One consequence of this pattern of divergence is that diagnostic differences among populations appear few in number. Additionally, although it is true that some of the populations exhibited unique mtDNA haplotypes, a closer analysis revealed that these differences resulted simply from the complete stochastic sorting of apparent ancestral variation.

In the context of conservation biology, we must use all available data to identify biologically meaningful units for the development of species management plans (Waples, 1991, 1995; Bernatchez, 1995). The molecular data developed in this study provide ample evidence that nearly all of the Death Valley pupfish populations are demographically isolated and must be managed independently. Long-term conservation goals are increasingly based on the resolution of Evolutionarily Significant Units (ESUs, Waples, 1991).

Molecular data may play a key role in the delineation of ESUs by providing an assessment of the extent of genetic distinctiveness demonstrated among populations, and providing an indirect test of their reproductive isolation. The reproductive isolation of most pupfish populations inhabiting spring habitats is clearly demonstrated by their geographic isolation. However, the direct assessment of reproductive isolation among populations that are located in riverine habitats is more problematic. These populations may retain the potential for occasional genetic exchange through migration. This study did not

uncover evidence of substantial population subdivision among any of the *C. n. amargosae* populations occupying the Amargosa River or any of its sampled tributaries. However, the significant reduction in genetic diversity detected in the small spring population north of Tecopa (site 8) did indicate that this population was at least partially reproductively isolated.

This study has largely confirmed the genetic distinctiveness of pupfish populations as originally described on the basis of their morphological features (Miller, 1948). Consequently, the data presented here add to the body of evidence which would suggest that all populations designated as distinct subspecies should be given separate ESU status. Additionally, each of the populations from the Ash Meadows region was demonstrably genetically distinctive and consequently qualified as a separate ESU according to the definition first proposed by Waples (1991).

References

Adey, N. B., S. A. Schichman, D. K. Graham, S. N. Peterson, M. H. Edgell, and C. A. Hutchison III. 1994. Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. Mol. Biol. Evol. 11:778-789.

Allendorf, F. W. 1995. Gentetics: defining the units of conservation. pp. 247-248 *in* **J. Nielsen, ed. Evolution and the Aquatic Ecosystem. Defining Unique Units in Population Conservation. American Fisheries Society Symposium, 17.**

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, K. Struhl, eds. 1987. Current protocols in molecular biology. John Wiley & Sons, New York.

Avise, J. C. 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. Phil. Trans. Roy. Soc. London B312: 325-342.

Avise, J. C., J. E. Neigel, and J. Arnold. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. 20: 99-105.

Avise, J. C., J. Arnold, R. M. Ball, *et al*. **1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18: 489-522.**

Avise, J. C. 1994. *Molecular Markers, Natural History and Evolution*. **Chapman and Hall, New York.**

Bermingham, E. and J. C. Avise. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113: 939-965.

Bernatchez, L. 1995. A role for molecular systematics in defining evolutionary significant units in fishes. pp. 114-132 *in* **J. Nielsen, ed. Evolution and the Aquatic Ecosystem. Defining Unique Units in Population Conservation. American Fisheries Society Symposium, 17.**

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

Cabot, E. L., B. Angeletti, K. Usdin, and A. V. Furano. 1997. Rapid evolution of a young L1 (LINE-1) clade in recently speciated *Rattus* **taxa. J. Mol. Evol. 45:412-423.**

Casavant, N. C., and S. C. Hardies. 1994a. The dynamics of murine LINE-1 subfamily amplification. J. Mol. Biol. 241:390-397.

Casavant, N. C., A. N. Sherman, and H. A. Wichman. 1996. Two persistent LINE-1 lineages in *Peromyscus* have unequal rates of evolution. Genetics 142:1289-1298.

Church, G. M. and W. Gilbert. 1984. Genomic Sequencing. Proc. Nat. Acad. Sci. USA 81:1991-1995.

Clough, J. E., J. A. Foster, M. Barnett, and H. A. Wichman. 1996. Computer simulation of transposable element evolution: random template and strict master models. J. Mol. Evol. 42:52-58.

D'Ambrosio, E., S. D. Waitzkin, F. R. Witney, A. Salemme and A. V. Furano. 1986. Structure of the highly repeated, long interspersed DNA family (LINE or L1RN) of the rat. Mol. Cell. Biol. 6:411-424.

Demers, G. W., M. J. Matunis, and R. C. Hardison. 1989. The L1 family of long interspersed repetitive DNA in rabbits: sequence, copy number, conserved open reading frames, and similarity to keratin. J. Mol. Evol. 29, 3-19.

Di Nocera, P. P., M. E. Digan, and I. B. Dawid. 1983. Genomic and structural organization of *Drosophila melanogaster* G elements. Nucleic Acids Res. 14:675-691.

Echelle, A. A. 1991. Conservation genetics and genic diversity in freshwater fishes of western North America. pp. 141-154 *in* W. L. Minckley and J. E. Deacon, Eds. Battle Against Extinction. Native Fish Management in the American West. The University of Arizona Press, Tucson.

Echelle, A. A. and T. E. Dowling. 1992. Mitochondrial DNA variation and evolution of the Death Valley pupfishes (Cyprinodon, Cyprinodontidae). Evolution 46: 193-206.

Echelle, A. A. and A. F. Echelle. 1993. Allozyme perspective on mitochondrial DNA variation and evolution of the Death Valley pupfishes (Cyprinodontidae: Cyprinodon). Copeia 1993: 275-287.

Egilmez, N. K., R. H. Ebert II, R. J. S. Reis. 1995. Strain evolution in *Caenorhabditis elegans*: transposable elements as markers of interstrain evolutionary history. J. Mol. Evol. 40:372-381.

Eickbush, T. H. 1992. The non-LTR class of retrotransposable elements. New Biol. 4:430-440.

------. 1994. Origin and evolutionary relationships of retroelements. Pp. 121-157 *in* S. S. Morse, ed. The Evolutionary Biology of Viruses. Raven Press, New York.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479-491.

Fajen, A. and F. Breden. 1992. Mitochondrial DNA sequence variation among natural populations of the Trinidad Guppy, *Poecilia reticulata*. **Evolution 46: 1457-1465.**

Fan, E., D. B. Levin, B. W. Glickman, and D. M. Logan. 1993. Limitations in the use of SSCP analysis. Mutation Res. 288: 85-92.

Feldmeth, C. R. 1981. The evolution of thermal tolerance in desert pupfish (genus Cyprinodon). pp. 335-356 *in* **R. J. Naiman and D. L. Soltz, eds. Fishes in North American Deserts. Wiley, New York.**

Finnegan, D. J. 1992. Transposable elements. Curr. Opinion Genet. Develop. 2: 861-867.

Fontdevila, A. 1993. Genetic instability and rapid speciation: are they coupled? pp. 242-253 *in* **J. F. McDonald, ed. Transposable Elements and Evolution. Kluwer Academic Publishers, Dordrecht.**

Furano, A. V., S. M. Robb, and F. T. Robb. 1988. The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. Nucleic Acids Res. 16:9215-9231.

Furano, A. V., B. E. Hayward, P. Chevret, F. Catzeflis, and K. Usdin. 1994. Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation. J. Mol. Evol. 38:18-27.

Garrett, J. E., D. S. Knutzon and D. Carroll. 1989. Composite transposable elements in the *Xenopus laevis* **genome. Mol. Cell. Biol. 9:3018-3027.**

Gillies, A. C. M., J. P. Cornelius, A. C. Newton, *et al.* **1997. Genetic variation in Costa Rican populations of the tropical timber species** *Cedrela odorata* **L., assessed using RAPDs. Mol. Ecol. 6: 1133-1145.**

Grossberger, D. 1987. Minipreps of DNA from bacteriophage lambda. Nucleic Acids Res. 15:6737.

Hard, J. J. 1995. A quantitative genetic perspective on the conservation of intraspecific diversity. pp. 304-326 *in* **J. Nielsen, ed. Evolution and the Aquatic Ecosystem. Defining Unique Units in Population Conservation. American Fisheries Society Symposium, 17.**

Hata, K., and Y. Sakaki. 1997. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. Gene 189:227-234.

Higgins, D. G., A. J. Bleasby, and R. Fuchs. 1991. CLUSTAL V:: improved software for multiple sequence alignment. Cabios 8: 189-191.

Hohjoh, H., and M. F. Singer. 1996. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J. 15:630-639.

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

Howell, R., and K. Usdin. 1997. The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons. Mol. Biol. Evol. 14:144-155.

Hudson, R. R., D. D. Boos, and N. L. Kaplan .1992. A statistical test for detecting geographic subdivision. Molecular Biology and Evolution 9: 138-151.

Huff, D. R., R. Peakall, and P. E. Smouse. 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm.]. Theor. Appl. Genet. 86: 927-934.

Hutchison, C. A. III, S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell. 1989. LINEs and related retroposons. Pp. 593-617 *in* D. E. Berg and M. M. Howe, eds. Mobile DNA. Am. Soc. Microbiol., Washington.

Ilves, H., O. Kahre and M. Speek. 1992. Translation of the rat LINE bicistronic RNAs in vitro involves ribosomal reinitiation instead of frameshifting. Mol. Cell. Biol. 12:4242-4248.

Izsvák, Z., Z. Ivics and P. B. Hackett. 1997. Repetitive elements and their genetic applications in zebrafish. Biochem. Cell. Biol. 75:507-523

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. Nature 314: 67-73.

Jurka, J. 1989. Subfamily structure and evolution of the human L1 family of repetitive sequences. J. Mol. Evol. 29:496-503.

Kass, D. H., F. G. Berger, and W. D. Dawson. 1992. The evolution of coexisting highly divergent LINE-1 subfamilies within the rodent genus *Peromyscus*. J. Mol. Evol. 35:472-485.

Kidwell M. G. and D. Lisch. 1997. Transposable elements as sources of variation in animals and plants. Proc. Nat. Acad. Sci. USA 94:7704-7711.

Kimura, M. 1968a. Evolutionary rate at the molecular level. Nature 217: 624-626.

-----. 1968b. Benetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. 11: 247-269.

King, J. L. and T. H. Jukes. 1969. Non-Darwinian evolution: random fixation of selectively neutral mutations. Science 164: 788-798.

Kocher, T. D., K. W. Thomas, A. Meyer, *et al*. **1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Nat. Acad. Sci. USA 86: 6196-6200.**

Kruskal, J. B. and M. Wish. 1978. *Multidimensional Scaling*. **SAGE Publications LTD, London.**

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.02. The Pennsylvania State University, University Park, PA.

LaBounty, J. F. and J. E. Deacon. 1972. *Cyprinodon milleri*, **a new species of pupfish (family Cyprinodontidae) from Death Valley, California. Copeia 1972: 769-780.**

Laughlin, T. F. and B. J. Turner. 1994. Multilocus DNA fingerprinting detects population differentiation in the outbred and abundant fish species *Poecilia latipinna*. **Mol. Ecol. 3: 263-266.**

-----. 1996. Hypervariable DNA markers reveal high genetic variability within striped bass populations of the lower Chesapeake Bay. Trans. Amer. Fish. Soc. 125: 49-55.

Lee, W-J, J. Conroy, W. H. Howell, and T. D. Kocher. 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. 41:54-66.

Liu, R. K. 1969. The comparative behavior of allopatric species (Teleostei: Cyprinodontidae: Cyprinodon). PhD thesis. University of California, Los Angeles.

Luan, D. D., and T. H. Eickbush. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. 15:3882-3891.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Mol. Biol. Evol. 7: 478-484.

-----. 1991. Analysis of population genetic structure by DNA fingerprinting. pp. 113-126 *in* **T. Burke** *et al.*, **eds. DNA Fingerprinting Approaches and Applications. Birkhauser.**

Marschalek, R., J. Hofmann, G. Schumann, R. Gosseringer, and T. Dingermann. 1992. Structure of DRE, a retrotransposable element which integrates with position specificity upstream of *Dictyostelium discoideum* **tRNA genes. Mol. Cell. Biol. 12:229-239.**

McDonald, J. F. 1993. Evolution and consequences of transposable elements. Current Opinion in Genetics and Development 3:855-864.

McElroy, D., P. Moran, E. Bermingham, and I. Kornfield. 1992. REAP: An integrated environment for the manipulation and phylogenetic analysis of restriction data. J. Heredity 83: 157-158.

Meyer, A., T. D. Kocher, P. Basasibwaki, and A. C. Wilson. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347: 550-553.

Miller, R. R. 1943. Cyprinodon salinus, a new species of fish from Death Valley, California. Copeia 1943: 70-78.

-----. 1948. The cyprinodont fishes of the Death Valley system of eastern California and southwestern Nevada. Miscellaneous Publications, Museum of Zoology, University of Michigan 68: 1-155.

-----. 1950. Speciation in fishes of the genera Cyprinodon and Empetrichthys inhabiting the Death Valley region. Evolution 4: 155-162.

-----. 1981. Coevolution of deserts and pupfishes (genus *Cyprinodon*) **in the American southwest. pp. 39-94** *in* **R. J. Naiman and D. L. Soltz, eds. Fishes in North American Deserts. Wiley, New York.**

Minckley, W.L. and J. E. Deacon, eds. 1991. Battle Against Extinction: Native Fish Management in the American West. University of Arizona Press, Tucson.

Moran, P. and I. Kornfield. 1993. Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. Mol. Biol. Evol. 10: 1015-1029.

Moran, P. and I. Kornfield. 1995. Were population bottlenecks associated with the radiation of the *Mbuna* **species flock (Teleostei: Cichlidae) of Lake Malawi? Mol. Biol. Evol. 12: 1085-1093.**

Moritz, C. 1994a. Application of mitochondrial DNA analysis in conservation: a critical review. Mol. Ecol. 3: 401-411.

-----. 1994b. Defining 'evolutionarily significant units' for conservation. Trends Ecol. Evol. 9: 373-375.

-----. 1995. Uses of molecular phylogenies for conservation. Phil. Trans. Roy. Soc. London B349: 113-118.

Moritz, C., T. E. Dowling, W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18: 269-292.

Moritz, C., S. Lavery, and R. Slade. 1995. Using allele frequency and phylogeny to define units for conservation and management. pp. 249-262 *in* **J. Nielsen, ed. Evolution and the Aquatic Ecosystem. Defining Unique Units in Population Conservation. American Fisheries Society Symposium, 17.**

Murata, S., N. Takasaki, M. Saitoh, and N. Okada. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc. Nat. Acad. Sci. USA 90:6995-6999.

Murata, S., N. Takasaki, M. Saitoh, *et al***. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the Pacific salmon and trout (***Oncorhynchus***) from the Atlantic salmon and trout (***Salmo***). Genetics 142: 915-926.**

Nei, M. and W-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Nat. Acad. Sci. USA 76: 5269-5273.

Nei, M. and T. Gojobori. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418-426.

Nur, I., E. Pascale, and A. V. Furano. 1988. The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. Nucleic Acids Res. 16:9233-9251.

Okada, N. and M. Hamada. 1997. The 3' ends of tRNA-derived SINEs originated from the 3' ends of LINEs: a new example from the bovine genome. J. Mol. Evol. 44:S52-S56.

Okada, N. M. Hamada, I. Ogiwara, and K. Ohshima. in press. SINEs and LINEs share common 3' sequences: a review. Gene.

Oshima, K., M. Hamada, Y. Terai, and N. Okada. 1996. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. Mol. Cell. Biol. 16:3756-3764.

Parenti, L. R. 1993. Relationships of Atherinomorph fishes (Teleostei). Bull. Mar. Sci. 52:170-196.

Parker, A. and I. Kornfield. 1997. Evolution of the mitochondrial DNA control region in the *mbuna* **(Cichlidae) species flock of Lake Malawi, East Africa. J. Mol. Evol. 45: 70-83.**

Pelikan, S. and S. H. Rogstad. 1996. *GELSTATS version 2.6*. **University of Cincinnati, Cincinnati.**

Pister, E. P. 1991. The desert fishes council: catalyst for change. Pp. 55-68 *in* **W. L. Minckley and J. E. Deacon, Eds. Battle Against Extinction. Native Fish Management in the American West. The University of Arizona Press, Tucson.**

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43: 223-225.

Ring, M., T. A. Pfeifer, and T. A. Grigliatti. 1996. Identification of a 5' truncated non-LTR-retrotransposon, YAKPs1, from the Variegated Cutworm, *Peridroma saucia*, **using PCR. Insect Biochem. Molec. Biol. 26:511-518.**

Roff, D. A. and P. Bentzen. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 **and the problem of small samples. Mol. Biol. Evol. 6: 539-545.**

Rogstad, S. H. and S. Pelikan. 1996. GELSTATS: A computer program for population genetics analyses using VNTR multilocus probe data. Biotechniques 21: 1128-1131.

Rosel, P. E. and B. A. Block. 1996. Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. **Marine Biol. 125: 11-22.**

Rozas, J. and R. Rozas. 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Computer Appl. Biosci. 13: 307-311.

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

Sassaman, D. M., B. A. Dombroski, J. V. Moran, M. L. Kimberland, T. P. Naas, R. J. DeBerardinis, A. Gabriel, G. D. Swergold and H. H. Kazazian Jr. 1997. Many human L1 elements are capable of retrotransposition. Nature Genet. 16:37-44.

Schwarz-Sommer, Z., L. Leclercq, E. Gobel, and H. Saedler. 1987. Cin4, an insert altering the structure of the *A1* **gene in** *Zea mays*, **exhibits properties of nonviral retrotransposons. EMBO J. 6:3873-3880.**

Shehee, W. R., S.-F. Chao, D. D. Loeb, M. B. Comer, C. A. Hutchison III, and M. H. Edgell. 1987. Determination of a functional ancestral sequence and definition of the 5′ end of A-type mouse L1 elements. J. Mol. Biol. 196:757-767.

Shimamura, M., H. Yasue, K. Ohshima, *et al*. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388:666-670.

Slatkin, M. and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555-562.

Smit, A. F. A., G. Tóth, A. D. Riggs, and J. Jurka. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J. Mol. Biol. 246:401-417.

Smith, G. R. 1981. Effects of habitat size on species richness and adult body sizes of desert fishes. pp. 125-172. *in* R. J. Naiman and D. L. Soltz, eds. Fishes in North American Deserts. Wiley, New York.

Soltz, D. L. and M. F. Hirshfield. 1981. Genetic differentiation of pupfishes (genus Cyprinodon) in the American southwest. pp. 291-333 *in* R. J. Naiman and D. L. Soltz, eds. Fishes in North American Deserts. Wiley, New York.

Soltz, D. L. and R. J. Naiman. 1978. The natural history of native fishes in the Death Valley system. Natural History Museum of Los Angeles County, Science Series 30: 1-76.

StatSoft. 1994. *Statistica release 4.3*. Statsoft Inc., Tulsa.

Stewart, C. N. Jr. and L. Excoffier. 1996. Assessing population genetic structure and variability with RAPD data: Application to *Vaccinium macrocarpon* (American Cranberry). J. Evol. Biol. 9: 153-171.

Strecker, U., C. G. Meyer, C. Sturmbauer, and H. Wilkens. 1996. Genetic divergence and speciation in an extremely young species flock in Mexico formed by the genus Cyprinodon (Cyprinodontidae, Teleostei). Mol. Phylogenet. Evol. 6: 143-149.

Swergold, G. D. 1990. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. 10:6718-6729.

Swofford, D. L. 1993. "Phylogenetic Analysis Using Parsimony (PAUP), Version 3.1.1," Illinois Natural History Survey, Champaign.

Tajima, F. and M. Nei. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1: 269-285.

Takasaki, N., T, Yamaki, M, Hamada, *et al.* 1997. The salmon *Sma*I family of short interspersed repetitive elements (SINEs): interspecific and intraspecific variation of the insertion of SINEs in the genomes of chum and pink salmon. Genetics 146: 369-380.

Taylor, F. R., R. R. Miller, J. W. Pedretti, *et al.* 1988. Rediscovery of the Shoshone pupfish, *Cyprinodon nevadensis shoshone*, (Cyprinodontidae), at Shoshone Springs, Inyo County, California. Bull. Southern California Acad. Sci. 87: 67-73.

Taylor, F. R. and J. W. Pedretti. 1994. Morphometric comparisons of pupfish populations, Cyprinodon nevadensis, at Shoshone and Tecopa, California. The Southwestern Naturalist 39: 300-303.

Thompson, C. E., E. B. Taylor, and J. D. McPhail. 1997. Parallel evolution of lake-stream pairs of threespine sticklebacks (*Gasterosteus*) inferred from mitochondrial DNA variation. Evolution 51: 1955-1965.

Toline, C. A. and A. J. Baker. 1995. Mitochondrial DNA variation and population genetic structure of the northern redbelly dace (*Phoxinus eos*). Mol. Ecol. 4: 745-753.

Tristem, M. 1996. Amplification of divergent retroelements by PCR. BioTechniques 20:608-612.

Turner, B. J. 1974. Genetic divergence of Death Valley pupfish species: Biochemical versus morphological evidence. Evolution 28: 281-294.

-----. 1983. Genic variation and differentiation of remnant populations of the desert pupfish, Cyprinodon macularius. Evolution 37: 690-700.

Turner, B. J. and R. K. Liu. 1977. Extensive interspecific genetic compatibility in the new world Killifish genus *Cyprinodon*. Copeia 1977:259-269.

Vanlerberghe, F., F. Bonhomme, C. A. Hutchison III, and M. H. Edgell. 1993. A major difference between the divergence patterns within the Lines-1 families in mice and voles. Mol. Biol. Evol. 10:719-731.

Waples, R. S. 1991. Definition of "species" under the endangered species act: application to Pacific Salmon. NOAA (National Oceanic and Atmospheric Administration) technical memorandum NMFS (National Marine Fisheries Service) F/NWC-194, Northwest Fisheries Science Center, Seattle.

Wichman, H. A. and R. A. Van Den Bussche. 1992. In search of retrotransposons: exploring the potential of the PCR. BioTechniques 13:258-264.

Wildekamp, R. H. 1995. *A World of Killies Atlas of the Oviparous Cyprinodontiform Fishes of the World*, Volume II. The American Killifish Association, Inc., Mishawaka.

Wilson, A. C., R. L. Cann, S. M. Carr, *et al.* 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. 26: 375-400.

Winkfein, R. J., R. D. Moir, S. A. Krawetz, J. Blanco, J. C. States, and G. H. Dixon. 1988. A new family of repetitive, retroposon-like sequences in the genome of the rainbow trout. J. Biochem. 176:255-264.

Woodcock, D. M., C. B. Lawler, M. E. Linsenmeyer, J. P. Doherty, and W. D. Warren. 1997. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. J. Biol. Chem. 272:7810-7816.

Xiong, Y., and Eickbush, T. H. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. 5: 675-690.

-----. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO. 9:3353-3362.Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, K. Struhl, eds. 1987. Current protocols in molecular biology. John Wiley & Sons, New York.

Curriculum Vitae

David Douglas Duvernell
Department of Biology
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0406

General Biography

1. **Born 17 April, 1970, at St. Louis, Missouri.**

2. **Married 15 June, 1996, to Christine L. Consiglio.**

Education

1. **Bachelor of Arts; Drury College, Springfield, MO, 1992.**

2. **Master of Science; Saint Louis University, 1994.**

3. **Doctor of Philosophy; Virginia Polytechnic Institute and State University, 1998**

Scholarships

1. **Missouri Bright Flite Scholarship; Sept., 1988 to May, 1992.**

2. **Drury College Academic Scholarship; Sept., 1988 to May, 1992.**

Professional Experience and Teaching

1. **Teaching Assistant; Saint Louis University, August 1992 to July, 1993.**

2. **Research Assistant; Saint Louis University, August, 1993 to July, 1994.**

3. **Graduate Teaching Assistant; VPI&SU, August, 1994 to May, 1998.**

4. **University Writing Program Summer Seminar; VPI&SU, May 12-16, 1997.**

Peer-Reviewed Publications

1. **Duvernell, D. D. & N. Aspinwall. 1995. Introgression of** *Luxilus cornutus* **mtDNA into allopatric populations of** *Luxilus chrysocephalus* **(Teleostei: Cyprinidae) in Missouri and Arkansas. Molecular Ecology, 4:173-181.**

2. **Duvernell D.D. & B.J. Turner. Evolutionary genetics of Death Valley pupfish populations: mitochondrial DNA sequence variation and population structure. Molecular Ecology. (in press).**

Papers Presented at Professional Meetings

1. **Independent vs. concerted evolution of a pupfish Hind III satellite DNA. [with B. Turner] Society for the Study of Evolution, Montreal, 1995.**

2. **An active L1-like LINE mobile sequence family in fish genomes: Its variation and divergence in Death Valley Pupfishes. [with B. Turner & H. Wichman] Society for the Study of Evolution, St. Louis, MO. June, 1996. also American Society of Ichthyologists and Herpetologists, New Orleans, LA, June, 1996.**

3. **An active L1-like mobile sequence family in fish genomes: Its variation and divergence in Death Valley Pupfishes. [with B. Turner] Southeastern Population Ecology and Genetics Group. Mountain Lake Biological Station. Pembroke, VA. September 13-15, 1996.**

4. **A family of L1-like retrotransposons in teleost genomes: molecular studies and population variation in Death Valley pupfishes. [with B. Turner] American Society of Ichthyologists & Herpetologists Symposium: Nuclear DNA and Evolutionary Genetics of Fishes, Amphibians, and Reptiles. Seattle, WA. June 26-July 2, 1997.**

5. **Mitochondrial DNA variation and population divergence in the pupfishes of Death Valley. American Society of Ichthyologists & Herpetologists. Seattle, WA. June 26-July 2, 1997.**

6. **Microgeographic variation of mtDNA haplotypes in the striped shiner, Luxilus chrysocephalus (Teleostei: Cyprinidae) from the Meramec River drainage (Missouri, USA). [with N. Aspinwall, K. Kohlberg, & R. Kube II] American Society of Ichthyologists & Herpetologists. Seattle, WA. June 26-July 2, 1997.**

7. **Population variation and divergence in Death Valley pupfishes delineated by a new family of retrotransposable elements. [with B. Turner]. Desert Fishes Council, Death Valley National Park, CA. November 20-23, 1997.**

8. **Evolutionary genetics of Death Valley pupfish populations: mitochondrial DNA sequence variation and population structure. [with B. Turner]. Desert Fishes Council, Death Valley National Park, CA. November 20-23, 1997.**

Invited Seminar

1. A family of L1-like retrotransposons in teleost genomes: molecular and population studies. National Institutes of Health. Bethesda, MA. May 19, 1997.

Grant Support

1995 Natural History Museum Roosevelt Fund, Satellite DNA sequence variation in the pupfish of Death Valley. One year, $1200.

1996 Sigma Xi Grant-in-Aid of Research, Population genetics of a retrotransposable element in the pupfishes of Death Valley. One year, $600.