Microsatellite variation and identification of a Y-chromosome marker in Atlantic salmon

Nick Elliott and Anne Reilly

FRDC Projects 95/80 and 96/347

Elliott, N. G. (Nicholas Grant), 1953-.

Microsatellite variation and indentification of a
Y-chromosome marker in Atlantic salmon.

ISBN 0 643 06187 8.

1. Atlantic salmon - Tasmania - Variation.
2. Atlantic salmon - Tasmania - Molecular genetics.
3. Atlantic salmon - Breeding - Tasmania.
   I. Reilly, Anne, 1968-.
   II. CSIRO. Division of Marine Research.
   III. Title.

572.86175609946

TABLE OF CONTENTS

1. SUMMARY

The aquaculture of Atlantic salmon began in Australia in the mid-1960's with an importation of ova from Canada to New South Wales. Anecdotal evidence suggested that the Australian population went through a severe bottleneck event during the early years due to poor survival and subsequent small broodstock numbers. The population, however, survived and numbers increased such that a few hundred broodstock were available each year from the early 1970's. The now flourishing Tasmanian Atlantic salmon industry was founded from the New South Wales population in the mid 1980's. About 115 000, 180 000 and 275 000 ova were brought to Tasmania in 1984, 1985 and 1986, respectively. The Tasmanian industry now produces over 7 000 tonnes annually, with the majority of smolt being supplied by the Salmon Enterprises of Tasmania Pty Ltd (SALTAS) commercial hatchery. SALTAS has for the past ten years maintained the Tasmanian broodstock, using several hundred females and males from two-year classes to produce each year's supply of smolt.

An important task for hatchery managers is the maintenance of genetic variation in their broodstock, and molecular geneticists can accurately assess genetic variation and detect meaningful changes. In recent years a new class of DNA marker - microsatellites - has shown promise for studies of genetic variation, progeny testing, genome mapping and quantitative trait loci (trait markers). We constructed an Atlantic salmon genomic DNA library which contains over 9 000 clones and several hundred potential microsatellite DNA markers. The DNA sequences of 24 of these have been determined and primers developed for eight markers for use in genetic studies. Three of these, plus five obtained from overseas contacts, were used to examine genetic variation in the Tasmanian population. As these markers were applied to the same individual fish that had earlier been examined for allozyme and mitochondrial DNA (mtDNA) variation, the relative abilities of these three techniques to detect genetic changes in hatchery populations could be assessed.

The genetic variation present in the Tasmanian and Canadian populations in 1993 was compared using these three molecular genetic methods - allozymes, mitochondrial DNA and nuclear DNA microsatellites. Some small but statistically significant allele frequency differences between the two populations were observed for one of seven polymorphic allozyme loci and for mitochondrial DNA haplotypes. However, there was no evidence of reduced genetic variability in the Tasmanian population. There were small but significant allele frequency differences observed at four of the eight microsatellite loci, and this analysis did show evidence of a small overall loss of genetic variation (loss of alleles and heterozygosity) in the Tasmanian population.

To investigate this suggested loss of genetic variation, and whether this occurred since the introduction to Tasmania or, as anecdotally suggested, during the early years of domestication in New South Wales, the genetic variation (microsatellites and allozymes) present in the Tasmanian and New South Wales populations in 1997 was then compared with that in the two previous samples.

While not conclusive, a loss of alleles and consequently low estimated effective population sizes for the Australian population does support the anecdotal evidence of a bottleneck in the Australian population of Atlantic salmon early in its introduction. However, low estimates of effective population sizes and losses of alleles in the Tasmanian population compared with the New South Wales population, and significant genetic differences between two Tasmanian samples, suggest further investigation of the Tasmanian population is warranted. In particular, time series data on genetic variation and more refined estimations of effective population sizes should be carried out.

A major activity for SALTAS (and salmonid culture world wide) is the production of all-female smolt for the industry. The precocious sexual maturity of males and their decrease in growth on maturity are cited as the main reasons for female preference. The production of XX males is standard practice; however, hatchery managers cannot be completely confident in discriminating between XX and XY males by morphological characters. A genetic marker for Y-chromosome carrying individuals is needed by the industry.

Other researchers have attempted to identify chromosome markers by applying a number of techniques including standard subtractive hybridization and RAPDs (random amplification of polymorphic DNA), but have failed to provide a sex marker. In our study, the application of microsatellite DNA markers and initial results from a modified subtractive hybridization technique, also unfortunately failed to identify a specific male-marker for Atlantic salmon. There are, however, additional difference product clones from our subtractive hybridization experiment that may yet provide a sex difference.

Future molecular genetics research opportunities with Atlantic salmon include continued application of the modified subtractive hybridization technique, and other techniques, for identifying a sex marker; such a marker remains an important international need. In addition, development of more Atlantic salmon microsatellite DNA markers is required for use in genome mapping and selective breeding projects. The 'genetic-health' of the Tasmanian population should continue to be monitored with the application of microsatellite makers for examining genetic variation and for progeny testing to refine the estimates of the effective breeding population. In addition, a better assessment of the genetic variation present in the progenitor Canadian population closer to the time of importation to Australia should be made through the use of DNA extracted from historical collections of scales.

In summary, the findings of this study include:

- There has been a small overall loss of genetic variation (loss of alleles and heterozygosity at eight microsatellite loci) in the Tasmanian Atlantic salmon population.
- A loss of alleles and consequently low estimated effective population sizes do support the anecdotal evidence of a bottleneck in the Australian Atlantic salmon population early in its introduction.
- Microsatellite loci, with higher numbers of alleles and higher heterozygosities, are more sensitive than allozyme or mtDNA loci to changes in genetic variation within a hatchery population.
- Suitable quality DNA for molecular genetic research was obatined from non-lethal fin clips.
- A Y-chromosome marker for Atlantic salmon was not identified

2. BACKGROUND

Genetic variation

SALTAS commissioned CSIRO in 1992 to study the genetic diversity in the Tasmanian Atlantic salmon population. The project was subsequently funded by FRDC as project 92/152. Allozyme and mitochondrial DNA (mtDNA) variation in the Tasmanian broodstock was assessed in 1993, and compared with levels in the parental River Philip, Nova Scotia, population (Ward et al. 1994). That initial study examined allozyme variation at seven variable loci and mtDNA variation with eighteen restriction enzymes. Small but significant changes in allele frequencies at one of the seven allozyme loci were observed, but there was no evidence of any overall reduction in genetic diversity. Only one of the 18 restriction enzymes applied to the mtDNA showed polymorphism, the two haplotypes having small but significant frequency differences

between the samples, with the Tasmanian sample being slightly more diverse than the Canadian. There was no evidence that allozyme or mtDNA variability in the Tasmanian stock was less than in its ancestral stock.

Loss of genetic variation in a cultured population can indicate inbreeding, which could have grave consequences as deleterious recessive genes are exposed and stocks lose vigour which is dependent on genetic variance. On the other hand, changes in genetic variation would be advantageous to a hatchery if they are associated with increased production or desirable characteristics. There has been no systematic selective breeding program in the Tasmanian Atlantic salmon industry, so any perceived advantageous changes have occurred by chance.

Genetic markers - microsatellites
A new class of nuclear DNA markers termed microsatellites has recently been developed, and is now well established in fisheries and in salmonids in particular (see O'Connell and Wright, 1997). Microsatellites are typically highly polymorphic and are expected to provide a more sensitive indicator of inbreeding effects than either allozymes or mtDNA. A microsatellite is a simple DNA sequence (e.g. CA, TTC, or TAAG) that is repeated several times, and the number of repeat units at any particular chromosomal location is often highly variable between individuals. There are many thousands of different microsatellite locations in the genome of higher organisms, and this along with their high variability makes them ideal for many genetic studies, including examination of genetic variation in populations.

The use of molecular genetic markers in broodstock management programs also includes their use in progeny or pedigree testing and as markers for desirable traits in selective breeding programs. Markers can be used to identify particular strains or cohorts, and also contribute to the development of linkage maps and identification of heritable traits such as growth, age at maturity, pigmentation and survival. In basic terms, they can identify regions of the genome (or individual chromosomes) that are associated with particular functions or characteristics.

Male-sex (Y-chromosome) marker
The ability to provide all-female smolt to industry is of major importance to SALTAS and other hatcheries. Males have a precocious sexual maturity compared to females, and on maturity the growth of males slows or ceases. Single sex farming in salmonids is recognized world-wide to be of major importance to the growth and continued success of the industry.

Standard breeding protocols to produce all-female smolt require, initially, the masculinisation (with androgens) of fry to produce all-male progeny (both XX males and normal XY males). Crossing the XX males with normal (XX) females will produce an all-female generation. Such methods are being applied by SALTAS. The selection of the XX males by this procedure is not 100% effective. A major problem with this procedure is the need to sacrifice males in order to obtain milt. It is believed that generally the sperm ducts in the XX males are incomplete and that normal running males are true XY males.

Researchers in Canada isolated a Y-chromosome marker for Chinook salmon (*Oncorhynchus tshawytscha*, Devlin *et al.* 1991), using a subtractive hybridization technique. They were unable to locate a similar marker for Atlantic salmon, but are confident that such a marker can be isolated (Devlin pers. com.). Atlantic salmon have 58 chromosomes and it is likely that only part of one will be Y-specific. A Y-chromosome probe has also been developed for the lake trout (*Salvelinus namaycush*), however this was made possible through microdissection of the Y-chromosome, as the lake trout is one of the few salmonids with morphologically differentiated sex chromosomes (Reed *et al.* 1995).

In addition to the use of microsatellite markers, a number of molecular genetic techniques for trait or marker screening have been developed in recent years. Some of these may be suitable for screening Atlantic salmon DNA for a potential Y-chromosome marker. Of particular interest is the application of modified subtractive hybridization techniques designed for isolating rare differentially expressed genes.

3. NEED

Genetic variation
In the first year of the project (95/80), microsatellites were applied to the same individuals (from Tasmania and River Philip) examined in 1993 for allozymes and mtDNA variation, and the results compared to confirm or refute the earlier suggestion of little loss of genetic variation. However, any long term trend in loss of genetic variation and so loss of diversity will need additional data points as the time since isolation of the local population increases. The initial sample, collected in January 1993, consisted of the progeny of 1989 year-class fish, and so represented the population after two generations of isolation. A sample collected in January 1997 (1993 year-class parents) provided the opportunity to examine whether the results obtained from the analysis of the 1993 samples were real, a sampling artifact, or indicative of a long term trend. The analysis of the 1997 sample was the second of a proposed regular 4 to 5 year assessment of the genetic "health" of the Tasmanian population, aimed at describing the nature and speed of any long term trends.

SALTAS, as the principal Atlantic salmon hatchery in Australia, has a long-term requirement to maintain industry and investor confidence in their product. Its ability to confirm the reliability of its breeding practices is important for the sustainability of the industry.

Any loss of genetic variation in Tasmanian Atlantic salmon could be difficult or impossible to recover due to the restrictions on importation of new broodstock.

The Tasmanian Atlantic salmon industry relies on captive spawners and as yet there is no quantitative selection of broodstock. The industry needs to have molecular genetic information and genetic guidelines to facilitate future selective breeding programs. Genetic markers (microsatellites) produced in the current study will be beneficial for future marker assisted selection trials and genetic mapping projects.

Y-chromosome marker
The ability to identify individual fish as Y-chromosome males without sacrificing animals would be a significant advance for both SALTAS and the international Atlantic salmon industry. For SALTAS alone such a marker has the potential to result in improvements in production efficiencies of 1-2% which, in terms of the current value of the industry production, represent a $1-1.5 million improvement in returns per annum.

While there was no guarantee that any of the markers developed in the project would be located on the Y-chromosome, these markers have the potential for identifying other quantitative traits required by the industry in the future, and can be combined with other microsatellite markers developed overseas for a potential Atlantic salmon genome mapping project.

4. OBJECTIVES

Project 95/80 was funded for 12 months (January to December 1996) with the objectives:

1. To develop DNA microsatellite techniques in Atlantic salmon, and to apply these to:

    i. a comparison of levels of microsatellite variation in farmed Tasmanian salmon and the parent stock in Nova Scotia; and

    ii. locating a male-sex (Y-chromosome) marker for broodstock management.

2. To develop non-lethal and non-destructive DNA extraction techniques to enable genetic analysis of valuable and non-replaceable individuals.

Project 96/347 was funded for a further 12 months (January to December 1997) with the objectives:

1. To locate a Y-chromosome marker in Atlantic salmon by applying a range of molecular genetic techniques.

2. To establish the rate of change in genetic variation in Tasmanian Atlantic salmon by comparing the genetic (microsatellite and allozyme) variation in progeny from 1993 year-class parents with that present in 1989 year-class parents and the parental Nova Scotia population.

5. MICROSATELLITE DNA MARKERS

5.1. CONSTRUCTION OF MICROSATELLITE LIBRARY

An Atlantic salmon genomic DNA library was constructed from the DNA of a male fish derived from the SALTAS 1993 year-class.

DNA was extracted from 50 mg of muscle tissue using a modified CTAB (hexadecyltrimethylammonium bromide) protocol (Grewe et al, 1993). DNA was treated with RNAase to destroy contaminating RNA. Approximately 30 µg of genomic DNA was digested to completion with 100 units of the restriction enzyme *Sau*3AI to generate small DNA fragments. These fragments were size-fractionated on a 1% TAE (Tris-Acetate) agarose gel, with ethidium bromide stain; size standards were included on either side of the sample lane. Exposure time of the gel to UV light was minimized to prevent DNA degradation and for good recovery of DNA from the gel. DNA was shielded from the UV light by placement of foil strips over the sample area. The gel was photographed with size standards visible and with a ruler alongside to assist in sizing.

The 500-750 bp (base pair) fraction was excised from the gel and the DNA extracted and purified by a Qiaex II Gel extraction kit (Qiagen). The fragments were ligated into the dephosphorylated *Bam*HI site of the plasmid pGEM™-3Zf(+) (Promega) by T4 DNA ligase (Pharmacia), in accordance with the manufacturer's recommendations. The plasmid vector was prepared by digesting approximately 19 µg of plasmid with 80 units of *Bam*HI, followed by treatment with 1 unit of calf intestinal phosphatase (Pharmacia) to remove the 5' phosphate groups. The enzyme was heat inactivated at 75˚C and the vector purified by phenol/chloroform extraction, followed by ethanol precipitation of the DNA.

Ligations were set up as 20 µl reactions to which was added 1 mM ATP and 1 Weiss unit of T4 DNA ligase (Pharmacia). Reactions were incubated overnight at 16˚C. The optimum molar ratio of vector to insert cohesive ends was found to be 1:1, providing the highest proportion of clones with an insert. This ratio was achieved using 400 ng vector and 78 ng insert. Half of the ligation reaction was transformed into Stratagene JM109 supercompetent cells by heat shock treatment, according to the recommended protocol; the remaining ligation was stored at -20˚C. Aliquots of 100 µl of transformation mix were spread onto LB (Luria-Bertani) agar plates, containing ampicillin, and X-gal (5-bromo-4-chloro-3-indoyl-ß-D-galactoside) and IPTG (isopropylthio-ß-D-galactoside) for blue/white color selection. Following overnight incubation at 37˚C, recombinant clones were detected as white colonies. The ratio of transformants to non-transformants was 2.5:1, as demonstrated by blue/white color selection.

The library, consisting of about 9000 recombinant clones, was screened for the presence of dinucleotide (CA) microsatellites, by the method outlined below.

5.2 OPTIMIZATION FOR NON-RADIOACTIVE SCREENING

A non-radioactive labeling and detection system was selected for screening microsatellite containing clones, as it is convenient and safe. A chemiluminescent alkaline substrate CSPD® was adopted, providing rapid and sensitive detection.

Oligonucleotide probes were prepared by 3' end labeling with DIG (digoxigenin) molecules (Boehringer Mannheim). Compared with other labeling methods, 3' end labeling gives high probe specificity which is necessary for colony hybridization (Boehringer Mannheim, 1995). Conditions recommended in the DIG users kit were adhered to for colony screening, with some modifications:

- UV fixation time of the DNA to the membrane was found to be important, and tests of cross-linking times with the particular light box (Ultra-Lum Transilluminator, 300 nm) gave an optimal 10 second fixation on each side;
- the volume of hybridization solution was increased approximately 5-fold to eliminate any risk of the membranes drying out and to ensure even distribution of the probe;
- Detection Buffer 1 was filtered through a 0.45 µm filter to remove particulate matter responsible for background spotting on the filter;
- magnesium chloride was omitted from Detection Buffer 3 as this also caused background spotting; and
- hybridization was optimal at the calculated Tm (52°C), not at 10˚C below as recommended in the kit.

5.3. END-LABELING OF OLIGONULEOTIDE DNA PROBE

The library of 9000 clones was screened for dinucleotide (CA) repeats with a $(CA)_8RT$ oligonucleotide 3' end-labelled with DIG molecules.

Briefly, 100 pmol of $(CA)_8RT$ oligonucleotide was labelled with Digoxigenin-11-ddUTP (2',3'-dideoxyuridine-5'-triphosphate, coupled to digoxigenin) using terminal transferase, in a total volume of 20 µl. The mixture was incubated at 37˚C for 15 minutes. The reaction was stopped in the presence of 0.4M EDTA (disodium ethylenediaminetetracetic acid) and ethanol precipitated in the presence of glycogen and lithium chloride. The probe was then dissolved in 20 µl sterile water, ready for quantitation. The test oligo probe was quantitated against dilutions of a DIG labelled oligo control provided by Boehringer Mannheim.

5.4. COLONY HYBRIDIZATION

The plating of 100 µl of transformed library mix provided a density of between 100-200 colonies per 80 mm diameter LB agar plate, which was ideal for screening. Plates were then refrigerated for about 1 hr before replica plating the colonies to nylon uncharged membrane filters (Boehringer Mannheim). Colonies on the membranes were grown on fresh LB ampicillin plates to approximately 1-2 mm in size, and master plates were incubated for 2 to 3 hr to regenerate colonies.

Membranes were soaked for 15 minutes in a solution containing 0.5M NaOH, 1.5 M NaCl, 0.1% SDS (sodium dodecyl sulfate) to denature the double-stranded DNA to single-stranded. Membranes were then neutralised by soaking in 1M TrisCl, pH 7.5, 1.5 M NaCl for 15 minutes, and then finally washed in 2 x SSC (sodium chloride, sodium citrate solution). DNA was UV fixed to the membrane by exposing each side of the membrane to UV light for 10 seconds. Bacterial proteins were then removed by proteinase K treatment. Cellular debris was removed by blotting between damp filter papers. Membranes were prehybridised at 52˚C in standard hybridization buffer containing 5 x SSC, 1% block solution, 0.1% N-lauroyl sarcosine, 0.02% SDS. The CA-repeat probe was added at about 60 pmol in 50 ml hybridization solution and allowed to hybridise at 52˚C for between 2 hrs to overnight. The membranes were then washed in 2 x SSC, 0.1% SDS at room temperature with gentle rocking. This was followed by stringency washes in 0.5 x SSC, 0.1% SDS at the hybridization temperature of 52˚C.

5.5. DETECTION

The membranes were equilibrated in washing buffer (150 mM NaCl, 100 mM maleic acid; pH 7.5), with 0.3% Tween 20, and then blocked in a 2% block solution. Anti-DIG alkaline phosphatase conjugate was added at a 10 000-fold dilution in block solution and incubated for 30 minutes. Membranes were then thoroughly washed in washing buffer, before equilibrating in 100 mM TrisCl, pH9.5, 100 mM NaCl. CSPD® substrate, diluted 100-fold in washing buffer, was added at 200 µl per membrane for chemiluminescent detection. After the addition of the substrate, the membranes were sealed in plastic and incubated at room temperature for 5 minutes, before activating the substrate at 37˚C for 15 minutes. Filters were exposed to X-ray film for about 1 hr before developing.

5.6. SEQUENCING OF CA-REPEAT CONTAINING CLONES

Positive clones, those colonies that hybridized with the CA repeat probe, were cultured overnight in ampicillin-containing LB medium. Double-stranded plasmid DNA was then prepared as template for sequencing using the alkaline lysis method (Sambrook et al, 1989). DNA was suspended in 20 µl TE buffer, containing 0.5 µg/µl RNAase (Pharmacia).

The DNA nucleotide sequence of selected positive clones was determined with ABI Prism® Dye Terminator Cycle sequencing (Perkin Elmer), according to kit instructions, and using 500 ng DNA template per reaction. Extension products were ethanol precipitated using Protocol 1 (Perkin Elmer, 1995). Sequencing reactions were run on 4% denaturing acrylamide gels on an ABI377 PRISM® DNA automatic sequencer.

Glycerol stocks were prepared for each of the positive clones. Overnight LB/Ampicillin cultures were resuspended in fresh LB media and an equal volume of sterile glycerol added. Cultures were snap frozen in liquid nitrogen and stored at -80˚C.

5.7. PCR PRIMER DESIGN

PCR (Polymerase Chain Reaction) primer pairs were designed for the conserved flanking regions of identified microsatellite repeats. Regions of repetitive nucleotide sequence were avoided, and primers were designed so that products were between 150 and 250 base pairs. Other design considerations included avoidance of sequences which may cause primer-dimer formation and internal looping, and care in matching annealing temperatures of the primer pair. Oligonucleotides were synthesized by Bresatec and the forward primer was labelled with a fluorescent tag.

5.8. RESULTS

Of approximately 9000 recombinant colonies screened, almost 200 responded positive to the CA microsatellite repeat probe. A return of 2.2% positive clones compares favourably with other estimates of the abundance of CA repeats in the Atlantic salmon genome of between 1.2 and 1.7% (McConnell et al, 1995b).

Of the 24 positive clones that were sequenced, 18 contained CA microsatellite repeats (Table 5.1). Poor or incomplete sequence data was obtained for five of the remaining clones, and the sixth was a false-positive (clone 1-18: reliable sequence data was obtained for the entire clone DNA but no CA microsatellite repeat was found). The microsatellite repeat sequence for each of the 18 loci is given in Table 5.1, and the complete sequence data for each of the 18 clones is included in Appendix 15.4.1.

Suitable PCR primers could only be designed for eight of the 18 clones containing CA repeats (Table 5.2); two alternative primer pairs were designed for two of these clones.

The other 10 clones were unsuitable for primer design due to the presence of repetitive sequences in the regions flanking the microsatellite, to close proximity of the microsatellite to the end of the clone sequence (in 4 cases), or to unreliable sequence data (Table 5.1). In addition, we restricted the PCR product size to between about 150 and 250 bp.

Each of the 10 designed primer sets (8 microsatellite loci) was tested for success and optimization of PCR amplification. The outcomes of these tests were:

- a 0.2μM concentration of each primer provided optimal amplification;
- the peak profile of a number of loci was improved by touchdown PCR amplification;
- a large number of CA repeats within the microsatellite correlated with the production of highly polymorphic loci as well as with increased formation of stutter peaks - due to higher rates of mutation caused by strand slippage which is dependent upon the number of repeat units (Weber, 1990);
- locus *cmrSs1.2* (primer pair A) produced co-amplification of artifact bands;
- loci *cmrSs1.2* (primer pair B), *cmrSs1.18, cmrSs3.4* and *cmrSs1.35* all produced numerous peaks that were difficult to score - further optimization of the PCR conditions or redesign of the primers may rectify this problem, but due to time constraints they were not taken any further during this project;
- locus *cmrSs1.10B* (primer pair B) failed to amplify;
- locus *cmrSs1.10A* (primer pair A) produced good clean peaks, and from here on is termed locus *cmrSs1.10* with the motif as presented in Table 6.1;
- locus *cmrSs1.22* gave optimal performance when amplified at a higher annealing temperature (60˚C, instead of the calculated 54˚C), this eliminated any artifact band and produced a clean peak profile that was easy to score;
- locus *cmrSs1-30* produced co-amplification of artifact bands, and stutter peaks made scoring difficult;
- only the three loci [*cmrSs1.10* (GenBank accession number AF020846); *cmrSs1.14* (AF020847) and *cmrSs1.22* (AF020848)] that amplified most reliably were selected for use in the genetic variation study presented in Chapter 6.

Table 5.1. Microsatellite motif and suitability for PCR primer design for 24 positive clones. Full sequence data are presented in Appendix 15.4.1, and the PCR primers designed for suitable loci (*) are given in Tables 5.2 and 6.1.

Clone	Microsatellite motif	Comments
*cmrSs1.2**	$(GT)_{34}$	OK; two primer pairs designed
cmrSs1.7	None detected	Poor sequence data, but TA repeat found
*cmrSs1.10**	$(GT)_3$; $(GT)_3$; $(CA)_2$; $(AT)_2$; $(CT)_2$; $(CA)_{14}C(CA)_3$; $(GT)_{30}$	OK; two primer pairs designed
*cmrSs1.14**	$(CA)_{22}$	OK
cmrSs1.15	$(GT)_{73}(GA)_9$	No overlap of sequence data
cmrSs1.16	$(CA)_3$; $(AGG)_3$; $(ACC)_2(CA)_{18}$	OK
*cmrSs1.18**	$(GT)_{43}(GA)_4$	OK
*cmrSs1.22**	$(CA)_3$; $(CA)_{21}$; $(CA)_2$; $(GA)_2$; A_9	OK
cmrSs1.28	$(CA)_{27}$	Microsatellite runs into vector
*cmrSs1.30**	$(GT)_{49}$	OK
*cmrSs1.35**	$(CA)_{28}$	OK
cmrSs1.39	$CTGT(CT)_2GT(CT)_2(GT)_6(CT)_2$;$(CT)_4$; $(GT)_{36}CT(GT)_6$	Repeat nucleotides flanking microsatellite
cmrSs3.1	None detected	Poor sequence data
cmrSs3.3	None detected	Poor sequence data
*cmrSs3.4**	$(CA)_{26}$	Only 20 base pairs before microsatellite
cmrSs3.5	$(GT)_4$; $(GT)_3$; $(GT)_2$; $(GT)_3$; $(GT)_{12}$; $(GT)_4$	Reverse sequence only
cmrSs3.7	$(GT)_{58}$	Microsatellite runs into vector
cmrSs3.8	$(GT)_{36}$	Repeat nucleotides flanking microsatellite
cmrSs3.9	$(CA)A(CA)_5TA(CA)_6$; $(GA)_4$; $(TG)_3(AG)_3$	Limited choices for reverse primer
cmrSs3.10	None detected	Poor sequence data
cmrSs3.11	$(GT)_{59}$	Poor sequence data for primer design
cmrSs3.12	$(GT)_{54}$	No consensus for reverse primer design
cmrSs3.13	$(GT)_{26}$	Poor sequence data for primer design
cmrSs3.14	None detected	Poor sequence data

(note: microsatellite clone nomenclature *cmrSs* - CSIRO Marine Research *Salmo salar*)

Table 5.2. PCR primer sequence (5' – 3') for six microsatellite loci. Alternative primers were designed for 2 loci (*cmrSs1.2* and *cmrSs1.10*). Details of the alternative primers for *cmrSs1.10* and the other two loci (*cmrSs1.14* and *cmrSs1.22*) used in the genetic variation study are presented in Table 6.1

Locus	Primer sequence
cmrSs1.2A	GCT TTC TCT GCT GTC TCC TGT AGT
cmrSs1.2B	GCT GCA CGC TTA TTG TCT GAA AGG GGT CGT GCT TTC TCT GCT AAG GCG GAT TCA TTT TCT GTA GGA
cmrSs1.10B	TGG CGG GGA GGC GGA GTC GAC TCT AGA GGA TCA AAC AAA AT
cmrSs1.18	ATA GCA CAG CCA GGT CAA ACA TCA TTC ATG CTG AAC TCC GCC AAT G
cmrSs1.30	GGT TTG CTA TGG TGT GGC TGT TAT T TGC CTT TAG GGG ATT TAG TTC TCA A
cmrSs1.35	GGA CTT AAC AGA TGG GAG TTT ATA C TTA AGG CCC AGC AAG GAT AGA A
cmrSs3.4	GAT CAA CTT CTC TAC CAC CT AGC CAA CTT AAC ACA ACT G

6. GENETIC VARIATION

6.1. SAMPLES

Four samples of Atlantic salmon were screened for microsatellite genetic variation:

- **River Philip** – collected in November 1992, these 72 fish were a random sample from about 1500 hatchery progeny derived from 16 female and 14 male wild fish returning to the River Philip (Nova Scotia, Canada) in 1991. The same individuals were examined for allozyme and mtDNA variation in 1993 (FRDC project 92/152 and reported in Ward et al. 1994). This sample represents the progenitor population for the Australian Atlantic salmon populations.

- **Tasmania 1** – collected in January 1993 from the SALTAS hatchery, these 72 fish were a random sample of the progeny of the 1992 spawning that consisted of 177 female and 95 male fish predominantly of the 1989 year-class with a small proportion of 1988 year-class repeat spawners. The same individuals were examined for allozyme and mtDNA variation in 1993 (FRDC project 92/152 and reported in Ward et al. 1994).

- **Tasmania 2** – collected in January 1997 from the SALTAS hatchery, these 100 fish were a random sample from the progeny of the 1992 spawning (TL ca.95–115 mm). This spawning consisted of predominantly 1993 year-class fish (226 females and 124 males) with a small proportion of 1992 year-class repeat spawners (39 females and 51 males).

- **Gaden** – collected in April 1997 from the NSW Fisheries hatchery at Gaden, NSW, these 100 fish were a random sample of the progeny of the 1992 spawning (TL ca. 30-45mm); a second small collection (20 fish TL ca. 80-100 mm) was made in July 1997 for an allozyme locus check. The spawning consisted of a mixture of three year classes (three year olds - 1993 year-class, four year olds - 1992 year-class, and five year olds - 1991 year-class) providing a potential nine crosses. A sex ratio of ca. 2:1 females to males was used, with ca. 60 three year old, 25 four year old and 12 five year old females; the ova from 8–10 females were generally mixed. Such a spawning protocol has been used since the early 1980's. The Gaden sample was taken to compare the Tasmania population with its more immediate progenitor population.

The River Philip and Tasmania 1 samples consisted of purified DNA that had been archived at -20°C since 1993. The other two samples consisted of whole fish, frozen soon after collection, from which muscle and liver tissues were dissected and stored at -80°C until required for either allozyme variation analysis or DNA extraction.

6.2. MICROSATELLITE ANALYSIS

Total genomic DNA was isolated from white muscle by a modified CTAB (hexadecyltrimethylammonium bromide) protocol (Grewe et al., 1993).

Eight microsatellite loci were used to examine the genetic variation between samples. The isolation and identification of three loci, namely *cmrSs1.10*, *cmrSs1.14* and *cmrSs1.22* is described above (Chapter 5). The isolation and identification of the three loci *20.19* (GenBank accession number U37490), *D30* (U37493) and *F43* (U37494) is reported in Sanchez et al. (1996), and loci *L5.27* (U37491) and *F49* (U37496/7) were isolated using the same method by Clabby (1996). The microsatellite motifs and PCR primer sequences for each locus are presented in Table 6.1.

PCR amplifications were made with a Perkin Elmer 9600 thermocycler. Individual amplifications were made as 25 µl PCR reactions containing 10 mM TrisHCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.5 µM forward primer (fluorescent labelled), 0.5 µM reverse primer, 200 µM dNTPs, 0.5 units Amplitaq Gold Taq polymerase (Perkin Elmer) and approximately 10 ng genomic DNA template. After an initial denaturation at 95°C for 12 min, samples were subjected to 35 cycles of denaturation at 96°C for 30 s, annealing at locus specific temperature for 30 s (Table 6.1), and extension at 72°C for 30 s. Touchdown cycling parameters were adopted for loci *L20.19*, *F43* and *cmrSs1.14* . In each case this involved 5 cycles at 65°C annealing temperature, followed by 10 cycles at 60°C and finally 15 cycles at 55°C. The final extension step was extended to 2 hr to reduce the formation of split peaks. Amplification products were run on 2% low melting point agarose gels to check the success of the PCR reaction.

The amplified products were diluted 10-fold, mixed with formamide loading dye containing Tamra size standards (ABI), denatured by heating to 95°C for 2 minutes, and loaded on a 4% denaturing acrylamide gel. The samples were run on an ABI377 Prism® DNA autosequencer with accompanying analysis software (Genotyper DNA Fragment Analysis Software version 1.1).

6.3. ALLOZYME ELECTROPHORESIS

Small pieces of muscle or liver tissue were extracted from semi-thawed fish and placed in 1.5 ml microcentrifuge tubes, homogenised manually with 1 or 2 drops of distilled water, and spun at 4°C in a microcentrifuge at 11 000 *g* for 2 minutes. The supernatant was used for electrophoresis.

Seven allozyme loci known to be polymorphic (Ward et al. 1994) were examined:
Adenosine deaminase (*Ada*),
Aspartate aminotransferase-3 (*Aat-3*),
Isocitrate dehydrogenase-4 (*Idh-4*),
Malate dehydrogenase-3,4 (*Mdh-3,4*),
Malic enzyme-2 (*Me-2*),
Phosphoglucomutase-1 regulatory gene (*Pgm-1-t*), and
Sorbitol dehydrogenase-1 (*Sdh-1,2*)

Liver extract was used for five loci, *Ada*, *Aat-3*, *Idh-4*, *Pgm-1-t* and *Sdh-1,2*, and muscle extract for *Mdh-3,4* and *Me-2*. Helena Titan III cellulose acetate plates run at 200 V with a Tris-glycine buffer system (0.020 M tris and 0.192 M glycine, pH 8.5; see Hebert and Beaton, 1993) were used for all enzymes except *Aat-3* which was examined on Helena Titan III cellulose acetate plates run at 150 V with a Tris-citrate buffer system (0.075 M tris and 0.025 M citric acid, pH 7.0). Staining techniques were largely as in Hebert and Beaton (1993).

Alleles within each locus were numbered according to the anodal mobility of their product, except for *Pgm-1-t* where it was not possible to distinguish unequivocally the heterozygotes from one of the homozygotes. For this locus, the allele with no expression was designated 'a' and the other allele 'b' (Ward et al. 1994). The frequency of allele 'a' was determined from the square root of the frequency of the phenotype with no expression (the 'aa' homozygote).

6.4. STATISTICAL ANALYSES

6.4.1 GENETIC VARIATION

Homogeneity of allele frequencies among samples at each locus was tested by the Roff and Bentzen (1989) χ^2 test, which obviates the need to pool rare alleles. A similar Monte-Carlo approach was used to test the conformance of observed genotype frequencies in each sample to those expected under Hardy-Weinberg equilibrium (Zaykin and Pudovkin, 1993). When multiple tests were carried out, the standard Bonferroni procedure was applied to significance levels (Lessios 1992). The predetermined significance level of 0.05 was adjusted according to the number of tests performed.

6.4.2. EFFECTIVE POPULATION SIZE AND GENETIC DRIFT

An estimate of the per-generation (average) effective population size for the Tasmanian introduced stock, and a test of genetic drift based on this estimate, were made by applying the methods described by Hedgecock and Sly (1990). Population size estimates from the microsatellite and allozyme data were compared.

With these analyses a number of assumptions were made:

1. that the allele frequencies in the 1993 River Philip sample were representative of the allele frequencies in that population in the mid-1960s, from which the progenitors of the Australian populations came;

2. that the Tasmanian population had been isolated from its progenitor River Philip population for ten generations when sampled in 1993 (Tasmania 1);

3. that both the Tasmanian and Gaden populations had been isolated for 11 generations from the progenitor River Philip population when sampled in 1997 (Tasmania 2 and Gaden);

4. that the allele frequencies in the Gaden sample were representative of the frequencies present in that population in the mid-1980's, from which the progenitors of the Tasmanian population came;

5. that the Tasmanian population had been isolated from its progenitor Gaden population for two generations when sampled in 1993 (Tasmania 1) and four generations when sampled in 1997 (Tasmania 2); and

6. that the Tasmania 2 sample had only one generation of isolation from the Tasmania 1 sample (broodstock for Tasmania 2 included 1993 and 1992 year-class fish).

Temporal variance of allele frequencies, F_K, for each locus were calculated as described in Hedgecock and Sly (1990) as

$$F_K = \frac{2}{K-1}\sum_{i=1}^{K}\frac{(x_{i,1}-x_{i,0})^2}{x_{i,1}+x_{i,0}},\ K \geq 3 \qquad \text{and } F_K = \frac{1}{K}\sum_{i=1}^{K}\frac{(x_{i,1}-x_{i,0})^2}{\bar{x}_i - x_{i,1}.x_{i,0}},\ K = 2,$$

where K is the number of alleles compared at a particular locus, and $x_{i,1}$ and $x_{i,0}$ are the allele i frequencies in the derived (e.g. Tasmanian collections) and progenitor (e.g. River Philip) populations, respectively. A multilocus estimate of the temporal variance of allelic frequencies, $\overline{F}_K$, was calculated as the average of F_K, weighted by the number of independent alleles per locus, $k = K$-1.

The estimates of the average effective population size (N_K) over t generations were calculated as

$$N_K = \frac{t}{2}(\overline{F}_K - \frac{1}{\sum k_l}\sum k_l(\frac{1}{2S_{1,l}} + \frac{1}{2S_{0,l}}))^{-1}, \quad (1)$$

where l is the number of loci (eight microsatellite or seven allozyme), k the number of independent alleles at locus l ($k = K$-1), and $S_{1,l}$ and $S_{0,l}$ are the number of individuals scored at locus l in the derived and progenitor populations, respectively. The variance of N_K was estimated as

$$\text{Est Variance } N_K = \frac{8(N_K)^4}{t^2(\sum k_l)^2}\cdot\sum k_l(\frac{t}{2N_K} + \frac{1}{2S_{1,l}} + \frac{1}{2S_{0,l}})^2,$$

where variables are as previously defined. The ninety five percent confidence limits (CL) of the estimated effective population size (N_K) were calculated according to the formula in Waples (1989):

$$(1\text{-}\alpha)\text{ CL for } \overline{F}_K = \left[\frac{k\overline{F}_K}{\chi^2_{\alpha/2[k]}}, \frac{k\overline{F}_K}{\chi^2_{1-\alpha/2[k]}}\right]$$

where $\alpha = 0.05$, and $\chi^2_{\alpha/2[k]}$ is the chi-square value at 0.025 probability with k degrees of freedom and $\chi^2_{1-\alpha/2[k]}$ is the chisquare value at 0.975 probability with k degrees of freedom. These bounds for $\overline{F}_K$ were then applied to the above formula (1) to calculate the CL for N_K.

The number of alleles (n_a) expected to remain in the derived population, given the average estimated effective population size (N_K), after t generations of isolation, was calculated as per Hedgecock and Sly (1990), as

$$n_a = \sum(1 - P_0)$$

where summation is over each allele present in the progenitor collection; alleles found only in the derived collection were excluded from the analysis and assumed to have been present in the progenitor population but not detected in our collection. P_o is the probability that a selectively neutral allele, at an initial frequency of p in the progenitor population, would be lost in the derived population after t generations and can be approximated by

$$\left[p(1-p)/C\right]^{1/4}\exp^{-2C/w}$$

where $C = (1/4)[\arccos(1\text{-}2p)]^2$ and $w = t/(2N_K)$. An adjusted chi-square goodness-of-fit test with one degree of freedom was used to test whether the observed number of alleles lost and remaining differed from the expectations (Hedgecock et al. 1992). The chi-square correction for continuity in two-cell case with small sample sizes was made by subtracting 0.5 from the absolute values of the deviation between the observed and expected values (Sokal and Rohlf 1981).

6.5 RESULTS

6.5.1. RIVER PHILIP AND TASMANIA 1 SAMPLES

One of the objectives of the first year of this study (project 95/80) was to determine, using microsatellite loci whether there had been any significant loss of genetic variation in the Tasmanian hatchery stock in comparison with its ancestral River Philip population. These new genetic markers were deployed on the same individual fish as previously analysed in 1993 for allozyme and mitochondrial DNA variation (Ward et al., 1994). Thus, the relative abilities of each genetic technique to detect changes brought about by hatchery practices could be assessed.

6.5.1.1. Microsatellite variation

Variation at eight microsatellite loci was examined in the two Atlantic salmon samples - River Philip and Tasmania 1 (Table 6.2).

Four loci showed the same alleles in both samples. These were the di-allelic loci *5.27* and *F49*, and the tetra-allelic loci *20.19* and *F43.* At three loci, the Tasmania 1 sample showed a loss of one or more alleles, although all lost alleles were rare in the River Philip sample (frequency ≤ 0.055). These loci were *cmrSs1.10* (one of a total of three alleles lost), *cmrSs1.14* (one of five alleles lost), and *D30* (four of six alleles lost). The eighth locus, *cmrSs1.22,* had a total of 14 alleles in these two samples, of which four (frequencies 0.040 to 0.113) were specific to the River Philip sample and two (frequency of both 0.016) to the Tasmania 1 sample.

In all, 38 alleles were observed in the River Philip sample and 30 alleles in the Tasmania 1 sample, giving means of 4.75 and 3.75 alleles per locus respectively (Table 6.3). While suggestive of a loss of allelic diversity in the Tasmanian population, especially since this sample came from a larger parental cohort than the River Philip sample, this difference is not statistically significant.

Observed heterozygosities ranged from 0.016 for locus *cmrSs1.10* in the Tasmania 1 sample to 0.919 for locus *cmrSs1.22* in the River Philip sample (Table 6.4). The mean observed heterozygosity per locus, *Ho*, was significantly less (Wilcoxon test, P = 0.017) in the Tasmania 1 sample (0.377), than in the River Philip sample (0.515). The difference in Hardy-Weinberg expected heterozygosity, *He*, while in the same direction (Tasmania 1, 0.434; River Philip, 0.509), was not quite significant (Wilcoxon, P=0.069). Loss of heterozygosity in the Tasmania 1 fish was 26.7 % for *Ho* and 14.7% for *He*.

Genotype frequencies were in agreement with Hardy-Weinberg equilibrium except for three deviations that remained significant after Bonferroni correction (16 tests, Table 6.4). One was at the highly variable *cmrSs1.22* locus in the River Philip sample (P = 0.001) associated with a small heterozygote excess (Selander's D = 0.070). The other two ($P < 0.001$) were for the *F43* locus, which had a substantial heterozygote deficiency in both River Philip (D = -0.438) and Tasmania 1 (D = -0.430) fish.

A heterozygote deficiency at *F43* was not observed in European Atlantic salmon samples analysed by Clabby (1996) and Sanchez et al. (1996). It is possible that the Canadian population from which our two samples originated has a non-amplifying or null allele. The presence of such an allele at a frequency of about 0.100 would account for the heterozygote deficiency. No null homozygotes were observed, but at an allele frequency of 0.100, only 1 in 100 fish would be expected to have such a genotype. A null allele at a microsatellite locus was reported in populations of Chinook salmon at frequencies of 0.385 to 0.678 (Scribner et al., 1996), and non-amplifying alleles have been described in other species (e.g. Callen et al., 1993; Pemberton et al., 1995).

Significant allele frequency differences between the River Philip and Tasmania 1 samples were found at four loci: *cmrSs1.14* ($P < 0.001$), *cmrSs1.22* ($P < 0.001$), *cmrSs1.10* ($P = 0.005$) and *D30* ($P = 0.005$). For these loci, the Tasmania 1 sample showed a decrease in both observed (*Ho*) and Hardy-Weinberg expected (*He*) heterozygosity (Table 6.4). Allele-frequencies at the remaining four loci were not significantly different after Bonferroni correction for multiple tests: *20.19* ($P = 0.150$), *F43* ($P = 0.021$), *F49* ($P = 0.030$) and *5.27* ($P = 0.033$).

The microsatellite allele frequency variance estimates between the assumed progenitor River Philip sample and the Tasmania 1 sample gave a multilocus estimate, $\overline{F}_K$, of 0.0926 (Table 6.9.I and section 6.5.4). This is similar to the value of 0.0618 calculated using the seven allozyme loci data from Ward et al. (1994) on the same fish (Table 6.9.II). The resulting per-generation effective population size estimates, N_K, were therefore similar at 65.2 ± 19.7 (s.d.) for the microsatellites, and 106.1 ± 74.4 for the allozymes (Table 6.10); combining the two data sets gave a value of 70.1 ± 19.4. These values are the estimated mean effective population size over the 10 generations of isolation. The estimates are lower than expected given the number of individuals currently spawned (177 females and 95 males Ward et al. 1994), but it should be remembered that they are the harmonic mean over the 10 generations of isolation (see section 6.5.4.4 for full discussion).

With an estimated per-generation effective population size of 65.2, and assuming the loss of alleles is due only to genetic drift, we calculated an expected loss of 7.6 microsatellite alleles in the Tasmanian 1 sample after ten generations of isolation. We recorded a loss of ten alleles from the total 38 observed in the River Philip sample; such a loss was not significantly greater than expected under random genetic drift ($\chi^2 = 0.594$, $P > 0.250$). Based on the same assumption and an effective population size of 106.1, we calculated an expected loss of 0.8 allozyme alleles, and Ward et al. (1994) showed no loss of alleles in the seven di-allelic loci examined.

6.5.1.2. Comparison of three genetic techniques

In the two samples of Atlantic salmon (River Philip and Tasmania 1), only one of 18 restriction enzymes applied to the mtDNA revealed variation (Ward et al., 1994). Only two haplotypes were found, at significantly different frequencies in the two samples ($P = 0.021$), and the haplotype diversity was higher in the Tasmania sample (0.31 cf. 0.11). Mitochondrial DNA (mtDNA) is haploid and maternally only inherited and is generally expected to be more sensitive to genetic drift than nuclear DNA as it has an effective population size only about 1/4 that of nuclear DNA (e.g. Birky et al., 1989). However, in Atlantic salmon, mtDNA variation is more restricted (e.g. Bermingham et al., 1991; Birt et al., 1991) than in many other fishes, and the power of any derived conclusions is therefore limited.

Only a minority of the allozyme loci that have been screened in Atlantic salmon are polymorphic, and these typically have only two or three alleles per locus (Cross and King, 1983; Ward et al. 1994; Tessier et al., 1995). Of the seven di-allelic polymorphic allozyme loci examined in the River Philip and Tasmania 1 samples (Ward et al., 1994), two showed allele frequency variation (*Mdh-3,4* $P = 0.008$ and *Pgm-1-t* $P = 0.033$), but neither was significant after Bonferroni adjustment for multiple tests. There was no evidence of allele loss in the Tasmanian fish, and the mean expected heterozygosity per locus of the Tasmanian fish (0.207) was no less than that of the River Philip fish (0.182).

Thus neither the mtDNA nor the allozyme examination flagged any loss of diversity in the Tasmanian hatchery stock (Ward et al. 1994).

Our microsatellite data, however, revealed a somewhat different picture. Four of the eight loci showed significant allele frequency differences ($P \leq 0.005$), and there was some evidence of allele loss in the Tasmanian fish (~21%), although this was not statistically significant. Observed heterozygosity in the Tasmanian fish (0.377) was about 27% less (and significantly so) than the Canadian fish (0.515), and the decline in expected heterozygosity (0.434 cf. 0.509, ~15% loss) verged on statistical significance.

Thus while both mtDNA and allozyme data failed to indicate any reduction in genetic variation in the Tasmanian hatchery stock, such a loss was indicated by the microsatellite data. Assuming neutrality of these various classes of variants, all should have responded similarly, at least in direction, to any bottleneck effects. MtDNA is only a single genetic marker, and is relatively invariant in Atlantic salmon; both features reduce its effectiveness for monitoring small genetic changes in hatchery stocks. In a somewhat analogous fashion, microsatellite data have revealed more than either allozymes or mtDNA data about the population structure of the cod *Gadus morhua* (Wright and Bentzen, 1994; Ruzzante et al., 1996).

It is likely that the microsatellite loci, with higher numbers of alleles per locus and higher heterozygosities, are more sensitive to changes in effective population size than the allozyme loci. Similar estimates of per-generation effective population size were obtained from microsatellite (65.2 ± 19.7) and allozyme (106.1 ± 74.4) data, though the microsatellite data gave a much lower error estimate (coefficient of variation 13% cf 70%). The higher number of microsatellite alleles per locus also provided a better examination of the loss of alleles due to genetic drift over the period of isolation.

6.5.2. 1997 TASMANIA 2 AND GADEN SAMPLES

The analysis of the 1993 Tasmania 1 sample suggested a loss of microsatellite genetic variation, compared with the River Philip sample, and a lower than expected per-generation effective population size estimate for the Tasmanian population.

Hatchery records for the early years of Atlantic salmon in New South Wales are poor, but anecdotal evidence suggests that breeding numbers were initially low. However, in the late 1970s and early 1980s several hundred broodstock were used each year in that hatchery. A similar practice has been followed at SALTAS, which uses around 400 female and 250 male broodstock of at least two year-classes each year. Such large numbers of broodstock – assuming that most do contribute to the gene pool – are unlikely to lead to a detectable erosion of genetic variation in the sample sizes we used. Therefore, if the above microsatellite results are a true reflection of the 'genetic health' of the Tasmanian population, this may reflect small population sizes in the early years (late 1960's) of the Australian domestication of Atlantic salmon in New South Wales, or more recently either in New South Wales prior to the importation to Tasmania in the late 1980's or in fact in the first few generations in Tasmania.

Questions that arise from this initial analysis of the population therefore include:

1. Is the 1992 River Philip sample a true representation of the progenitor population for the Australian populations? (i.e. how would that sample compare with a late 1960's sample from the River Philip?)
2. Is the 1993 Tasmania 1 sample a true representation of the Tasmanian population? (i.e. is a sample of 60–70 fish representative of the population, and/or is this 1993 sample, progeny from the 1989 year-class, representative of all year-classes?)
3. If the result is valid, did the apparent loss of variation occur early in the population's history or is it a continual event occurring with each generation of isolation?

At the time of collection, the sample from the River Philip was the most suitable given the techniques available at the time, but it may be possible with new technology to improve on this aspect (see section 10. Further Developments). Sample sizes were in

hindsight probably slightly low, but constrained to a certain extent by resources. With an average common allele frequency of about 0.7, sample sizes of 60–80 will detect a difference in allele frequency of 0.2 with only a 5% chance of finding a biologically spurious but statistically significant difference, and a 20% chance of missing significant cases (Richardson et al. 1986). To reduce the chance of missing a significant case to around 10% the sample size needs to be increased to around 100, but to increase the detecting power (i.e. reduce the difference in frequency detected) sample sizes would have to be increased to over 300 per sample to reduce the difference detected from 0.2 to 0.1; such numbers become unrealistic with limited resources. The major problem, however, with this sample was that the fish were the progeny of a small number of broodstock captured by the hatchery (section 6.1).

To address the questions of year-class differences and historical timing of loss of genetic variation, both microsatellite and allozyme variation was examined in a second Tasmanian sample (Tasmania 2) and a sample from the New South Wales hatchery at Gaden, both collected in 1997. The Tasmania 2 sample is also the first step in a suggested regular monitoring of the Tasmanian population.

6.5.2.1 Microsatellite variation

The same alleles were observed in the Tasmania 2 and Gaden samples as were previously observed, in the River Philip and Tasmania 1 samples, at three of the eight microsatellite loci (Table 6.2). These were the di-allelic loci *5.27* and *F49*, and the tetra-allelic locus *20.19*. The Tasmania 2 sample shared the same four alleles at locus *F43* as were previously observed, but only three alleles were observed in the Gaden sample.

At locus *D30*, the two more common alleles originally observed were again present in the Tasmania 2 and Gaden samples. Two rare alleles (frequency ≤ 0.055) previously recorded only in the River Philip sample were observed, one only in the Gaden sample and the other in both samples; the latter at higher frequencies than observed in the River Philip sample. However, two other rare alleles, recorded in the River Philip sample were not observed in any of the 259 Australian fish.

The same common allele (frequency > 0.900) was found in all four samples at locus *cmrSs1.10,* with three rare alleles - two found in the River Philip and Tasmania fish, and one new allele recorded only in the Gaden fish (Table 6.2).

Of the five *cmrSs1.14* alleles observed in the River Philip sample, four were common to all samples and the fifth was not observed in any Australian fish, but a new rare allele (frequency 0.010) was observed only in the Tasmania 2 sample.

At the highly polymorphic *cmrSs1.22* locus eight new alleles were observed in the two 1997 samples. Of the total 22 alleles observed at this locus (Table 6.2), only two were specific to the River Philip sample, while ten were specific to either one or all of the three Australian samples, and of these five were specific to the Gaden sample. One allele, *cmrSs1.22*210,* was common (frequency 0.129 to 0.258) in all samples. While it was the most common allele in the River Philip sample it was not so in the Australian fish, where allele **208* was the most common (0.305 – 0.508 cf 0.081 in River Philip). As to be expected most of the remaining 20 alleles were at low frequencies, with some exceptions - **192, *212* and **214* were at a frequency > 0.100 in the River Philip sample but at a low frequency or absent from the Australian samples, and alleles **190, *206* and **228* were at a low frequency or absent from three samples but present at a frequency > 0.100 in Tasmania 2, Gaden and Tasmania 1 respectively.

There still appears to be a loss of alleles (Tables 6.2 and 6.3) in Australian fish compared with the River Philip sample (except of locus *cmrSs1.22*), albeit not as marked as previously suggested. The earlier indication of a 21% loss of alleles in the Tasmanian population (3.75 alleles per locus Tasmania 1) compared with the River

Philip sample (mean of 4.75), was not as dramatic in the second analysis (8% loss for Tasmania 2 with mean of 4.37, and 5% loss for combined Tasmania sample with mean of 4.50), and the Gaden sample in fact showed an increase (+8%, mean 5.13). These latter figures are, however, heavily influenced by the number of rare alleles present at locus *cmrSs1.22*.

Of the total 38 alleles observed in the River Philip sample, five were not observed in any Australian fish (n = 250–263). On the other hand, of the 46 alleles observed in all the Australian fish, 12 were not observed in the River Philip sample, but 10 of these were at the highly polymorphic locus *cmrSs1.22*. In the Gaden sample alone 41 alleles were observed, of which nine were not observed in the Tasmanian fish and seven of these were at locus *cmrSs1.22* (Table 6.2).

Ignoring the highly polymorphic locus *cmrSs1.22*, the mean number of alleles for seven microsatellite loci in the River Philip sample decreases to 3.71 and there was a 23% loss recorded for the Tasmania 1 (mean 2.86), 11% loss for the Tasmania 2 (3.29) and 19% loss for the Gaden (3.00) samples. As previously, not one of these losses in allelic diversity is statistically significant.

The Tasmanian population shows a loss of alleles (but not significant) compared with the Gaden sample. For all eight loci, there was a 27% loss for the Tasmania 1 sample and 15% loss for the Tasmania 2 sample; the combined Tasmania data suggests a 12% loss in alleles.

Observed heterozygosities at each of the eight loci in the Tasmania 2 and Gaden samples ranged from 0.000 for locus *cmrSs1.10* in the Gaden sample to 0.822 for locus *cmrSs1.22* also in the Gaden sample; a similar range to that previously recorded for the Tasmania 1 and River Philip samples (Table 6.4).

The mean observed heterozygosities per locus (*Ho*) were significantly less in the two Tasmanian (Wilcoxon tests both $P = 0.017$) and Gaden ($P = 0.042$) samples than in the River Philip sample (Table 6.4); although the differences do not remain significant after Bonferroni adjustment. The differences in Hardy-Weinberg expected heterozygosities (*He*), while in the same direction, were not significant (Table 6.4).

Losses of heterozygosity in the three Australian samples compared to the River Philip sample (progenitor population) were:

	Ho	*He*
Tasmania 1 vs River Philip	26.8 %	14.7%
Tasmania 2 vs River Philip	15.9%	5.3%
Gaden vs River Philip	15.3%	3.5%,

and the losses of heterozygosity in the two Tasmanian samples compared to the Gaden sample (progenitor population) were:

	Ho	*He*
Tasmania 1 vs Gaden	13.5%	11.6%
Tasmania 2 vs Gaden	0.7%	1.8%.

The Tasmania 2 sample had an increased ($Ho = 14.8\%$, $He = 11.1\%$) mean heterozygosity compared with the Tasmania 1 sample.

Our observed reduction in mean heterozygosity over the eight microsatellite loci, while verging on significance, is less than the statistically significant reduction reported by Clifford et al. (1998) in a farm strain of Atlantic salmon (0.281) compared to wild samples (0.532) over three minisatellite loci.

For the two 1997 samples (Tasmania 2 and Gaden), there were four significant deviations from Hardy-Weinberg equilibrium (Table 6.4). As with the initial analyses (River Philip and Tasmania 1), the two most significant ($P < 0.001$) deviations were for the *F43* locus, with substantial heterozygote deficiencies (Selander's D: Tasmania 2 = -0.489 and Gaden = -0.521, compared with River Philip = -0.438 and Tasmania 1 = -0.430). As discussed above (section 6.5.1.1), this may be due to the presence of a null or non-amplifying allele. The other significant deviations were at locus *D30* in the Gaden sample ($P = 0.001$), associated with a small heterozygote deficiency (D = -0.045), and at locus *cmrSs1.10* also in the Gaden sample where no heterozygotes were observed.

Significant allele frequency differences ($P < 0.001$) existed between the four samples at all but two of the eight loci (after Bonferroni adjustment):

	χ^2	*P*
cmrSs1.10	57.283	<0.001
cmrSs1.14	35.859	<0.001
cmrSs1.22	393.388	<0.001
20.19	49.199	<0.001
5.27	6.027	0.110
D30	72.624	<0.001
F43	68.638	<0.001
F49	11.166	0.013

Pair-wise χ^2 tests revealed significant differences between all pairs of samples at between three to five loci (Table 6.5). Such differences were not unexpected as the three populations (River Philip, Tasmania and Gaden) have been isolated for a number of generations, however the difference between the two Tasmanian samples was not expected to be as marked as it was, with significant differences at five loci. The Tasmania 2 sample consisted of progeny from a 1996 spawning of predominantly 1993 year-class fish, but with some 1992 year-class fish. The Tasmania 1 sample was from the 1992 year-class (see 6.6 for further discussion).

6.5.2.2. Allozyme variation

Seven di-allelic enzyme loci were examined in the Tasmania 2 and Gaden samples (Table 6.6), and the variation compared with the previous results obtained for the other two samples (Ward et al. 1994). Poor resolution at the *Ada* locus for the smaller Gaden fish resulted in this locus not being included in the analyses involving that sample. Two (rare) alleles previously observed in the River Philip and Tasmania 1 samples were not observed in the Tasmania 2 or Gaden samples, these were at *Ada* and *Me-2*. Genotype frequencies in the Tasmania 2 and Gaden samples were in agreement with Hardy-Weinberg equilibrium (Table 6.7).

The observed heterozygosities ranged from zero to 0.510 and the mean per locus observed (*Ho*) and expected (*He*) heterozygosities for the four samples were similar at around 20% (Table 6.7). As with the microsatellite results, the Gaden sample was suggestive of a higher heterozygosity than the other samples, but the difference was not significant.

Significant heterogeneity (after Bonferroni adjustment) between the four samples was observed at four loci:

	χ^2	*P*	
Ada	2.795	0.191	(three samples only)
Aat-3	8.542	0.039	
Idh-4	4.818	0.175	
Mdh-3,4	21.596	<0.001	
Me-2	15.180	0.006	
Pgm-1-t	23.160	<0.001	
Sdh-1,2	20.690	<0.001	

Unlike the microsatellite loci, pairwise comparisons at these loci showed some structure to the observed differences (Table 6.8). The River Philip sample was significantly different to the other three samples at *Mdh-3,4*, from Tasmania 2 and Gaden at *Me-2* , and Gaden also at *Pgm-1-t*. The Gaden sample also differed significantly from the Tasmania 2 sample at both *Sdh-1,2* and *Pgm-1-t*. However, the two Tasmanian samples, despite the apparent loss of two alleles in the Tasmania 2 sample, did not differ from each other.

6.5.3. GENETIC DRIFT

6.5.3.1. Tasmania and Gaden vs Progenitor – River Philip

The ranges of allele-frequency variance estimates, F_K, for the eight microsatellite loci were similar for the three comparisons - from 0.0287 to 0.1428 in the Tasmania 1 comparison, from 0.0114 to 0.1094 in the Tasmania 2 comparison and from 0.0044 to 0.1675 in the Gaden comparison (Table 6.9.I). The multilocus estimates, $\overline{F_K}$, were therefore similar, and there was no significant difference between the three sets of data. The resulting per-generation effective population size estimates, N_K, were similar at 65.2 ± 19.7 (s.d.) for Tasmania 1 over 10 generations, and 87.9 ± 25.4 for Tasmania 2 and 92.1 ± 24.9 for Gaden over 11 generations (Table 6.10).

For the seven allozyme loci, similar variances and multilocus estimates were recorded for the two Tasmanian sample comparisons, and although values were higher for the Gaden comparison (Table 6.9.II) they were not significantly different [the Tasmania 1 and Gaden comparison (Wilcoxon $P = 0.028$) was not significant with Bonferroni adjustment]. The corresponding effective population size estimates were similar to the microsatellite results for the two Tasmania samples (106.1 ± 74.4 and 70.8 ± 43.7) but lower at 33.2 ± 20.6 for the Gaden sample (Table 6.10). The largest contributing F_K value to the Gaden allozyme result was *Pgm-1-t* (twice as large as for either Tasmania sample).

Mean effective breeding sizes of less than 100 are lower than would be expected based on the current hatchery practice of stripping several hundred broodstock. The current number of breeding individuals (stripped) in Tasmania falls outside the upper 95% confidence limits of the estimated mean effective population sizes; those for Gaden are less than the upper limit (Table 6.10). However these low values should be treated with caution and do not suggest that current hatchery management is poor. Firstly, these are mean (harmonic mean) per generation figures estimated over the ten or eleven generations of isolation, and therefore one or two low effective spawnings (or bottlenecks), perhaps in the early years of domestication in Australia, would have significant and permanent effects on reducing this average value. Secondly, it is not known how representative the 1992 River Philip sample is of the actual progenitor population. Only further analyses of either the original progenitor population or time series data on the Tasmanian population will confirm the suggestion of a bottleneck event and validate the current hatchery practices (see section 6.6 for further discussion).

The largest allele-frequency variance for the comparison between the River Philip sample and the two Tasmanian samples was recorded at the highly polymorphic locus *cmrSs1.22* (Table 6.9.I). If this locus is removed from the analyses, the multilocus variance estimates reduce from 0.0926 to 0.0583 for Tasmania 1 and from 0.0756 to 0.0502 for Tasmania 2, with resulting increased estimates of the average effective population size - 117.9 ± 31.2 for Tasmania 1 and 147.5 ± 35.9 for Tasmania 2. The variance at this locus between River Philip and Gaden was not as high and its omission from the analyses had little effect on the overall result ($\overline{F}_K$ reduced to 0.0724 and N_K increased to 92.8 ± 17.5).

Locus *cmrSs1.22* may well provide the best indication that a bottleneck event occurred during the early years of the introduction to Australia. We can logically assume that the rare alleles observed only in the Gaden and/or Tasmanian samples were originally present at very low frequencies in the River Philip progenitor population, and were not observed in the 1993 sample due to the small sample size and possible small parental cohort. Then our observed shift in the common allele (that with the highest frequency) from **210* to **208,* and the apparent increase in frequency of the rare alleles, could well be a consequence of a bottleneck (or small effective population size) at one time. A bottleneck event could as easily cause a loss of rare alleles as an increase in rare alleles, even more so if there was some selection pressure also included (i.e. shift in environmental conditions and 6-month shift in spawning period).

With the above (eight locus) estimated per-generation effective population sizes, and assuming loss of alleles from the progenitor population (River Philip) is due to genetic drift, we calculated an expected loss of 7.5 microsatellite alleles in the Tasmania 1 sample after ten generations of isolation - we observed a loss of 10 (Table 6.11). Likewise, we calculated a loss of 6.4 and 6.2 alleles respectively for Tasmania 2 and Gaden after 11 generations of isolation, and observed a loss of 8 in each. Ignoring the *cmrSs1.22* locus, we still observed a greater loss of alleles than expected - Tasmania 1 calculated loss of 2.4 and observed 6, Tasmania 2 calculated 2.3 and observed 4, and Gaden calculated 3.5 and observed 6.

For the allozyme analysis, Ward et al. (1994) observed no loss of alleles in the seven di-allelic loci examined for the River Philip and Tasmania 1 samples, we calculated an expected loss of 0.8 alleles. In the Tasmania 2 sample we observed a loss of 2 alleles and calculated an expected loss of 1.2, and observed the loss of one allele in the Gaden analysis (only six loci) and calculated an expected loss of 1.5.

Microsatellite loci with higher numbers of alleles and heterozygosity, as discussed before, appear to be more sensitive to changes in genetic variation in hatchery populations, and this seems to be supported with the calculation of expected maintenance (or loss) of alleles. However, the observed allele losses in the Australian fish are not significantly greater than would be expected under random genetic drift (Table 6.11).

6.5.3.2. Tasmania vs Progenitor – Gaden

Similar estimates of allele-frequency variance for the eight microsatellite loci were observed for the two comparisons with Tasmanian samples and there was no significant difference between the two sets of data (Table 6.9.I). As expected after only two (Tasmania 1) and four (Tasmania 2) generations of isolation, the resulting multilocus estimates (0.0505 and 0.0407, respectively) were lower than those obtained for the previous comparisons with the River Philip sample. However, these relatively lower variances still resulted in lower than expected (compared to number of parents stripped) mean effective population size estimates, from the microsatellite data, of 26.9 ± 12.0 (s.d.) for Tasmania 1 and 65.8 ± 20.5 for Tasmania 2 (Table 6.10). [Note: from

equation (1) the effective population size (N_K) is inversely related to the allele-frequency variance (F_K), and directly related to the number of generations of isolation (t)]

The variance at *cmrSs1.22* was not high in these two comparisons and had little effect on the overall result, in fact omitting the locus from the analyses would increase the multilocus variance estimate and decrease the effective population size estimates. The variance at this locus between these samples compared to that between them and the River Philip sample strengthens the possibility that allele frequency shifts at this locus supports the likelihood of a bottleneck early in the domestication history of Atlantic salmon in Australia. Alternatively, it may reflect the low number of broodstock that contributed to the River Philip 1993 sample.

Although the allele-frequency variances at the six allozyme loci were similar and there was no significant difference between the two data sets, the multilocus estimates reflect the increase in number of generations of isolation - Tasmania 2 after four generations of isolation (0.0821) was nearly twice that for Tasmania 1 (0.0422) after only two generations of isolation (Table 6.9). The corresponding mean effective population size estimates were therefore similar - Tasmania 1, 34.5 ± 29.0 and Tasmania 2, 28.1 ± 17.1. The largest contributing F_K value in the Tasmania 2 estimate was at *Sdh-1,2* (three times as large as for Tasmania 1).

We calculated an expected loss of 6.2 microsatellite alleles in the Tasmania 1 sample after 2 generations of isolation, and observed a loss of 13 (Table 6.11), and expected a loss of 5.1 alleles for Tasmania 2 after 4 generations and observed 10. Both these observed losses of alleles were significantly greater than was expected by random genetic drift. For the allozyme analysis no loss of alleles was observed, and we expected no loss after 2 generations for the Tasmania 1 sample and a loss of only 0.3 alleles after 4 generations (Tasmania 2).

The relatively high allele-frequency variances between the Gaden and Tasmania samples, higher than expected loss of alleles and subsequent low effective population sizes after just two or four generations is of concern and needs examining. While we can suggest that the genetic variation observed in the 1997 Gaden sample may not be a true representation of the variation that was present in the mid 1980's at the time of importation to Tasmania, this seems unlikely given the hatchery practice of stripping about 100 females and 50 males; albeit lower numbers than in the Tasmanian industry.

6.5.3.3. Comparison of Tasmania 1 and Tasmania 2

Is the 1993 Tasmania 1 sample a true representation of the Tasmanian population? The results we have obtained would suggest this is not so.

There were significant differences in allele frequencies between the two Tasmania samples at five microsatellite loci - *cmrSs1.14* , *cmrSs1.22*, *20.19, D30* and *F43* (Table 6.5). The level of variation between the two samples is also seen in the relatively high average allele-frequency variances between the two samples despite only one generation of isolation (microsatellites 0.0627 and allozymes 0.0268; Table 6.9), and this was reflected in a very low estimated effective population size (microsatellites 10.1 ± 3.4, allozymes 35.6 ± 36.3; Table 6.10), and a significant loss of allozyme alleles (Table 6.11).

Our results, based on two samples, would not suggest a detrimental loss of genetic variation between the two Tasmania samples, as the more recent Tasmania 2 sample had a higher (but not significant) level of microsatellite heterozygosity and a similar level of allozyme heterozygosity to the first sample (Tables 6.3 and 6.7).

Yet just how related are these two samples; are they two samples of one population or from two separate lines?

There were three successive introductions of ova to Tasmania (1984, 1985 and 1986), and each would have provided mature fish for spawning after 3 years, so commencing three spawning lines (first spawning 1987, 1988 and 1989, respectively). The annual hatchery practice has been to strip predominantly 3 year-old broodstock but with a proportion of repeat spawning 4 year-olds included (i.e. crossing spawning lines). The Tasmania 2 sample, therefore, was a random sample from the 1996 spawning which consisted of 226 females and 124 males from the 1993 year-class, and 39 females and 51 males from the 1992 year-class fish; the latter being the year-class from which the Tasmania 1 sample was randomly chosen. Hence the Tasmania 2 sample was partially derived from the broodstock represented by the Tasmania 1 sample. Yet with the stripping and mixing of such numbers of broodstock, the effective population size would be expected to be well in excess of 200.

The low estimated effective population size from the Tasmania comparison may reflect non-random sampling of the progeny from either spawning (i.e. only progeny from a small number of crosses contributed to the sample) and/or existence of (three) separate spawning lines. The crossing of spawning lines (3 and 4 year-olds) in recent generations would not have had sufficient time to impact on the genetic variation in each line, i.e. bring them closer together.

To fully interpret our results will require both long term monitoring of the three spawning lines, and a closer examination of the effectiveness of the spawning protocol, in particular the contribution of the mixed year classes. The common hatchery procedure of pooling the gametes from several males and females has been shown previously with salmonids to result in lower than expected effective breeding numbers, as all broodstock, but especially males, do not contribute equally in pooled crosses (Gharrett and Shirley 1985, Withler 1988, Simon et al 1986, Gile and Ferguson 1989).

Suggested research would include firstly, annual assessment of the genetic variation in random samples of parr, and in the first instance this would include multiple samples to assess the effectiveness of the random selection protocol and numbers of individuals required to efficiently assess the population. Secondly, application of microsatellite or AFLP (amplified fragment length polymorphism, see section 7.2.1) markers to identify broodstock individuals and then progeny test samples of parr produced from the pooled gametes of those individuals, thereby assessing the actual contribution of each parent to the gene pool, and so an assessment of the actual effective population size.

6.6 DISCUSSION

Our results are suggestive of a possible decline in genetic variation in Australian Atlantic salmon compared to the Canadian progenitor population, and of a low mean effective population size over the years of isolation. The latter is lower than anticipated based on the number of individuals stripped each year (at least for the past 4 or 5 generations) at the hatcheries to produce the next generation. However the heterozygosity differences are not statistically significant, and there are a number of assumptions with the analysis of effective population size (Waples 1989, Hedgecock et al. 1992) that may be invalid here and so put a different interpretation on the results. In particular, the validity of the 1993 River Philip sample to represent the genetic variation present in the original importation of ova to Australia may be questioned.

The relevant assumptions and their impact on the interpretation of our results are:

1. Constant effective population sizes. The estimation of the mean effective population size (N_K) is in essence a summary of the population's breeding history over the successive generations (t) of isolation from the progenitor population. The mean value

is the harmonic mean of the effective breeding numbers (N) at each of those generations (Simon et al. 1986):

$$N_K = [1/t\ (1/N_1 + 1/N_2 + ... + 1/N_t)]^{-1}$$

With low numbers of generations, the mean value is closer to the smallest breeding number than the larger. For example: if the breeding numbers were 50 for two generations and then 500 for eight successive generations, the mean effective breeding number over the ten generations would be approximately 180.

If we assume that the effective breeding numbers for the first seven generations of the Australian population at the Gaden hatchery were 150 followed by three generations at 250 in Tasmania, then over ten generations the effective mean population size would be approximately 170. An additional generation with breeding numbers of 500 would raise that mean value to 180 over 11 generations. However, if as the anecdotal evidence suggests there were very low breeding numbers in the first few generations, the mean value would decline, e.g. breeding numbers of 50 for two generations, followed by five generations at 150 and three at 250, would return a mean of around 120. Such values are within the 95% confidence limits of the estimates we have obtained for the comparisons with the River Philip sample, consistent with the suggestion that perhaps there was a bottleneck early in the domestication of the Australian population.

However, our low effective breeding estimates from the comparison of the Tasmania samples with the Gaden sample can not be explained by this means.

2. Progenitor population representation. We have analysed samples taken at the same time from the progenitor and derived populations (i.e. River Philip and Tasmania 1 in 1993, and Gaden and Tasmania 2 in 1997), and have assumed that the progenitor sample is a truly random and unbiased representation of that population at the time of exporting the fertilized ova to start the derived population. In the case with the River Philip sample this seems potentially invalid due to low broodstock numbers used to derive the progeny from which the sample was taken (Ward et al 1994).

Hedgecock et al (1992) discuss the point that the variance in allele frequencies between samples taken at the same time from the progenitor and derived populations actually represents an accumulation of changes along two lines of descent - one for the progenitor population and the other the derived hatchery population. They further suggest that when the number of generations (t) is not large, the expected variance becomes $E(F) \approx t/(2N_p) + t/(2N_d) + 1/(2S_p) + 1/(2S_d)$, where N_p, N_d, S_p and S_d are the effective population sizes and sample sizes of the progenitor and derived populations respectively (compare to equation (1) in section 6.2). This reduces to $E(F) \approx t/(N_e) + 1/(2S_p) + 1/(2S_d)$, where N_e is the harmonic mean of the effective population sizes over both lines of descent. The estimator of N_e, is then $N_k = t/[F - 1/(2S_p) - 1/(2S_d)]$, when this is compared to equation (1) (section 6.2), it lacks the 2 present in the denominator. Therefore, Hedgecock et al. argue that the N_k estimates made should be multiplied by two and interpreted as the harmonic mean of the effective population size of the pair of populations over the number of generations of isolation. Such an interpretation for our data would return estimates between 120 and 200 for the comparisons with the River Philip sample.

Yet again this does not assist in explaining the results of the Gaden and Tasmania comparisons.

3. Neutral alleles. Hedgecock et al (1992) suggested that the assessment of the goodness-of-fit of the number of alleles lost and remaining was a suitable means to evaluate the neutrality of alleles, assuming that random genetic drift was the only process removing alleles during the isolation from the progenitor population. The six assessments (microsatellites and allozymes for three samples) of change from the River Philip sample were all non-significant suggesting selectively neutral alleles.

However significant results were obtained with the microsatellite alleles (but not allozymes) when comparing the Gaden and Tasmania samples; the majority of lost alleles occurred at one locus. Such a loss would bias the allele-frequency variance upwards and so lower the estimated effective population size.

4. No mutation. Given the low number of generations involved in the derivation of the Australian Atlantic salmon population (maximum 11), mutation would not be expected to play a major role in changing allele frequencies, and can be safely ignored.

5. Random selection of samples. It is assumed that the four samples collected were each a random sample of parr from the entire population, and thus that the allele frequencies are an unbiased estimate of those in the population. Although the samples were 'random selections' from the available progeny, they may not be a true representation of the population due to the sample size or the actual selection method based on the mixing of parr in the hatchery. This potential sampling error needs investigation to determine the best sampling protocol and the number of samples required to fully represent the population. If the samples were not fully representative of the entire population they would most likely cause an increased estimate of allele-frequency variance between the two samples and result in lower estimates of effective breeding numbers.

6. No migration. It can be safely assumed that there was no immigration into the Gaden and Tasmanian samples from other populations after the respective initial importations. The same cannot be assumed for the River Philip sample, as that was derived from a small number of wild returning individuals. Such an event would again increase the allele-frequency variance and lower the estimated effective population size.

Discrete lines of Atlantic salmon may have arisen in the Australian population as a result of the early bottleneck (low breeding numbers) in New South Wales soon after importation. The subsequent number of generations with mixed age-group (line crossing) spawnings, has not been large and would not have resulted in a relatively uniform stable population. The Gaden and Tasmania 2 samples collected in the same year would most likely have been from the same line, and so have a relatively lower allele-frequency variance and higher N_K value, than the comparison of Tasmania 1 with Gaden which would be from different lines.

If three discrete lines were introduced to Tasmania (either from three lines or an initial genetically stable stock), then allele frequencies in the progeny of the 1996 spawning when two lines were mixed, would be expected to move back towards the original values, reduce genetic drift, reduce the allele-frequency variance and so increase the effective breeding size estimate. We have observed a possible increase in effective population size between the two Tasmanian samples - at least from the microsatellite data. The presence of discrete lines could explain the observed differences between the two Tasmania samples.

While it seems possible to explain the results of the comparisons of the Australian samples with the River Philip sample, and the conclusion that a bottleneck occurred early in the introduction to Australia, it is not possible to satisfactorily explain the results of the comparison of the Tasmania samples with the Gaden sample. Despite the bottleneck and subsequent low effective breeding numbers, there appears to have been a sufficient breeding population present to maintain a relatively high degree of genetic variation in the population. However this comparison is based on the 1993 River Philip sample which may well have a reduced level of variation compared to that present in the mid-1960's at the time of importation to Australia. This question needs further investigation to firstly establish a better estimate of the effective breeding size and secondly to confirm that breeding numbers are sufficient to maintain a 'genetically healthy' population.

6.7. CONCLUSIONS

- Microsatellite loci, with higher numbers of alleles per locus and higher heterozygosities, are more sensitive to changes in effective population size than allozyme loci, providing a lower error associated with estimates of per-generation effective population size, and a better examination of the loss of alleles due to genetic drift over the period of isolation.
- Allozyme loci, while not as polymorphic as microsatellite loci, and therefore less informative, are still useful markers for assessing genetic variation in hatchery populations of Atlantic salmon.
- MtDNA is relatively invariant in Atlantic salmon making it rather ineffective for monitoring genetic changes in hatchery stocks.
- Observed microsatellite locus heterozygosities ranging from 0.016 to 0.919 and number of alleles from 2 to 22 are typical for microsatellite loci in salmonids (e.g. McConnell et al. 1995a, Slettan et al. 1996).
- There has been a decrease in microsatellite mean heterozygosity in the Tasmanian population compared to the River Philip and Gaden samples.
- The loss of microsatellite and allozyme alleles in the Australian population was not significantly greater than that expected by random genetic drift, and suggests neutrality of alleles. However, significant losses were recorded between the Tasmania samples and the Gaden sample, the majority of loss occurring at one microsatellite locus.
- Estimates of average per-generation effective population size for the Australian population were less than 100 based on the allele-frequency variance of eight microsatellite loci, but increased to over 100 if the highly polymorphic locus *cmrSs1.22* was omitted from the analyses. Such low estimates when compared to actual broodstock numbers and assuming equal probability of those contributing to the offspring, is suggestive of a bottleneck event at some stage in the isolation of the Australian population.
- If it is accepted that a bottleneck occurred early in the Australian domestication, it can be assumed that current broodstock numbers are providing relatively large effective population sizes. Whether these are sufficient to maintain a high genetic diversity in the population requires continued monitoring of the local population and further investigation of the progenitor population.
- Locus *cmrSs1.22* , with a large number of rare alleles and a shift in frequency of the most common allele, may provide the best evidence to suggest that the putative bottleneck event occurred early in the history of the Australian Atlantic salmon population. However, this locus may also reflect the low number of parents contributing to the River Philip sample.
- Locus *F43* appears to have a non-amplifying or null allele present at a frequency of ≤ 0.100 in populations of Atlantic salmon originating from the River Philip in Canada. This 'allele' requires further examination and verification.
- The two Tasmanian samples have different levels of heterozygosity and number of alleles, and show significant variation in microsatellite allele frequencies. This needs further examination to determine whether this is due to a sampling artifact or is a consequence of historical discrete lines in the population.
- The apparent low effective population sizes may reflect a low contribution rate to the gene pool by some individuals with the mixed spawning protocol adopted by the hatcheries.

Table 6.1. Microsatellite motif, PCR primer sequences (5'– 3'), number of alleles observed, expected size of alleles and locus-specific annealing temperatures (˚C) for eight Atlantic salmon microsatellite loci.

Locus	Motif	Primer sequences	Alleles		Annealing temperature
			No.	Size[3]	
cmrSs1.10	$(gt)_3..(gt)_3..(ca)_2..(at)_2..(ct)_2..(ca)_{14}c(ca)_3$	GAG AGT GTG CGC TTG TGC T CTG TCC GCC TCC CCG CCA TTA	3	218	60
cmrSs1.14	$(ca)_{22}$	GTC AGG TCG TTA CGG ATG GAT TGG CTT CTG TCT GAC GCG GCT CTG C	5	160	65-60-55[4]
cmrSs1.22	$(ca)_3..(ca)_{21}..(ca)_2..(ga)_2..a_9$	TCC TCG GCC AGC TGG TTC TTT A TGT GTA CGC ATG GAT AGT CTC	14	204	60
20.19[1]	$(ac)_{13}g(ac)_4$	TCA ACC TGG TCT GCT TCG AC CTA GTT TCC CCA GCA CAG CC	4	96	65-60-55[4]
5.27[2]	$(tc)_6..(tc)_4..(tc)_2..(tc)_2..(tc)_4$	GTT ACC TTG CTC CTA G CCA GTG TGC CAC CCC	2	128	50
D30[1]	$(tc)_5..(tc)_7..(tc)_5..(tc)_8..(tc)_9..(tc)_{10}..(tc)_{10}$	AGC AGT AAA GAG AGA GAC TG TGT TGA CTT CCT TCC CCA AG	6	240	53
F43[1]	$(ac)_{18}$	AGC GGC ATA ACG TGC TGT GT GAG TCA CTC AAA GTG AGG CC	4	121	65-60-55[4]
F49[2]	$(tc)_3(t)_4(tc)_6tg(tc)_9tgta(tc)_4g(ct)_4cac(t)_2$	ACA GCT GAC CTC TCC CAT CT AGA GGT TTA GGC CAG CAC TG	2	175	58

[1]Sanchez et al. (1996). [2]Clabby (1996). The (..) among motif descriptions represents a gap of several non-repetitive nucleotides. [3]Expected PCR product size (bp) based on cloned Atlantic salmon allele of specified repeat number. [4]Touchdown cycling temperatures, see text for details.

Table 6.2. Allele frequencies for eight polymorphic microsatellite loci in four samples of Atlantic salmon. *n* = number of individuals scored. – = allele not detected.

Locus	Allele (bp)	R. Philip	Tasmania 1	Tasmania 2	Gaden
cmrSs1.10					
	216	–	–	–	0.051
	218	0.911	0.992	0.935	0.949
	220	0.024	–	0.050	–
	222	0.065	0.008	0.015	–
	n	*62*	*62*	*100*	*97*
cmrSs1.14					
	136	0.185	0.213	0.170	0.250
	138	–	–	0.010	–
	148	0.015	–	–	–
	158	0.131	0.115	0.145	0.133
	160	0.492	0.648	0.520	0.520
	162	0.177	0.025	0.155	0.097
	n	*65*	*61*	*100*	*98*
cmrSs1.22					
	182	–	–	–	0.042
	184	–	–	0.035	0.053
	186	–	0.016	–	0.032
	188	–	–	0.035	0.037
	190	0.016	0.016	0.130	0.016
	192	0.113	–	–	0.016
	198	0.040	–	–	–
	202	–	–	0.005	0.037
	204	0.016	0.040	0.060	0.032
	206	0.048	0.065	0.045	0.121
	208	0.081	0.508	0.310	0.305
	210	0.258	0.129	0.210	0.158
	212	0.105	0.024	0.015	0.021
	214	0.169	0.048	0.045	0.021
	216	0.089	–	–	0.005
	218	–	–	–	0.005
	222	–	–	–	0.010
	224	–	0.016	0.045	0.042
	226	–	–	–	0.005
	228	0.008	0.137	0.065	0.016
	230	–	–	–	0.026
	238	0.056	–	–	–
	n	*62*	*62*	*100*	*95*
20.19					
	78	0.047	0.095	0.025	0.153
	84	0.359	0.294	0.180	0.240
	96	0.055	0.111	0.155	0.158
	98	0.539	0.500	0.640	0.449
	n	*64*	*63*	*100*	*100*

Table 6.2 cont.

5.27					
	124	0.844	0.726	0.795	0.756
	128	0.156	0.274	0.205	0.244
	n	*64*	*62*	*100*	*88*
D30					
	230	0.023	–	–	–
	234	0.273	0.402	0.281	0.204
	236	0.055	–	–	0.031
	238	0.016	–	–	–
	240	0.625	0.598	0.597	0.684
	242	0.008	–	0.122	0.082
	n	*64*	*61*	*98*	*98*
F43					
	109	0.082	0.068	0.071	0.046
	115	0.467	0.652	0.505	0.713
	121	0.262	0.136	0.372	0.241
	125	0.189	0.144	0.051	–
	n	*61*	*66*	*98*	*87*
F49					
	174	0.817	0.914	0.860	0.781
	176	0.183	0.086	0.140	0.219
	n	*63*	*64*	*100*	*96*

Table 6.3. Comparison of allele numbers over eight microsatellite loci. Total number of alleles, range per locus and mean number per locus. n = range of number of individuals analysed per locus.

Sample	n	Total	Range	Mean per locus
River Philip	62–65	38	2—12	4.75
Tasmania 1	61–66	30	2—10	3.75
Tasmania 2	98–100	35	2—12	4.37
Tasmania combined	159–164	36	2—13	4.50
Gaden	88-100	41	2—20	5.13

Table 6.4. Microsatellite observed (*Ho*) and Hardy-Weinberg expected (*He*) heterozygosity estimates, and probabilities of genotype conformance to Hardy-Weinberg equilibrium (*P*). Sample means and standard errors are also presented.

Gene		River Philip	Tasmania 1	Tasmania 2	Gaden
cmrSs1.10					
	Ho	0.177	0.016	0.110	0.000
	He	0.166	0.016	0.123	0.098
	P	*1.000*	*1.000*	*0.012*	*<0.001*
cmrSs1.14					
	Ho	0.692	0.541	0.700	0.653
	He	0.680	0.526	0.656	0.640
	P	*0.034*	*0.579*	*0.187*	*0.237*
cmrSs1.22					
	Ho	0.919	0.726	0.820	0.822
	He	0.866	0.703	0.826	0.854
	P	*0.001*	*0.043*	*0.425*	*0.991*
20.19					
	Ho	0.766	0.524	0.530	0.663
	He	0.580	0.647	0.533	0.693
	P	*0.022*	*0.137*	*0.776*	*0.497*
5.27					
	Ho	0.313	0.355	0.290	0.352
	He	0.266	0.401	0.326	0.370
	P	*0.214*	*0.505*	*0.383*	*0.776*
D30					
	Ho	0.578	0.410	0.449	0.480
	He	0.535	0.485	0.550	0.502
	P	*0.634*	*0.297*	*0.108*	*0.001*
F43					
	Ho	0.377	0.303	0.306	0.207
	He	0.676	0.536	0.599	0.432
	P	*<0.001*	*<0.001*	*<0.001*	*<0.001*
F49					
	Ho	0.302	0.141	0.260	0.313
	He	0.301	0.158	0.241	0.342
	P	*1.000*	*0.393*	*0.678*	*0.563*
Mean ± SE	*Ho*	0.515 ± 0.093	0.377 ± 0.081	0.433 ± 0.085	0.436 ± 0.096
	He	0.509 + 0.086	0.434 + 0.084	0.482 ± 0.082	0.491 ± 0.083

Table 6.5. Probability results of pairwise χ^2 tests for allele frequency homogeneity at eight microsatellite loci between the four collections of Atlantic salmon. RP = River Philip, T1 = Tasmania 1, T2 = Tasmania 2 and G = Gaden. Bold values are significant after Bonferroni adjustment.

cmrSs1.10	T1	T2	G
RP	**0.005**	0.021	**<0.001**
T1		0.040	0.395
T2			**<0.001**

cmrSs1.14	T1	T2	G
RP	**0.001**	0.438	0.065
T1		**<0.001**	0.035
T2			0.092

cmrSs1.22	T1	T2	G
RP	**<0.001**	**<0.001**	**<0.001**
T1		**<0.001**	**<0.001**
T2			**<0.001**

20.19	T1	T2	G
RP	0.127	**0.001**	**<0.001**
T1		**<0.001**	0.184
T2			**<0.001**

5.27	T1	T2	G
RP	0.032	0.299	0.069
T1		0.158	0.578
T2			0.409

D30	T1	T2	G
RP	**0.003**	**<0.001**	**<0.001**
T1		**<0.001**	**<0.001**
T2			0.009

F43	T1	T2	G
RP	0.020	**0.001**	**<0.001**
T1		**<0.001**	**<0.001**
T2			**<0.001**

F49	T1	T2	G
RP	0.019	0.371	0.482
T1		0.172	**0.002**
T2			0.039

Table 6.6. Allozyme allele frequencies for four samples of Atlantic salmon. River Philip and Tasmania 1 data from Ward et al 1994. Gene and allele nomenclature as per Ward et al. 1994 (see text). *n* = number of individuals scored. - = allele not observed. na = locus not analysed (see text). Allele frequencies for *Pgm-1-t* were calculated assuming Hardy-Weinberg equilibrium (see text).

Gene	Allele	River Philip	Tasmania 1	Tasmania 2	Gaden
Ada					
	100	0.986	0.984	1.000	na
	90	0.014	0.016	-	
	n	71	63	92	
Aat-3					
	100	0.153	0.077	0.111	0.064
	50	0.847	0.923	0.889	0.936
	n	72	65	98	102
Idh-4					
	116	0.160	0.117	0.147	0.076
	100	0.840	0.883	0.853	0.924
	n	72	64	99	59
Mdh-3,4					
	120	0.181	0.323	0.405	0.378
	100	0.819	0.677	0.595	0.622
	n	72	65	100	102
Me-2					
	125	0.958	0.977	1.000	1.000
	100	0.042	0.023	-	-
	n	72	65	99	102
Pgm-1-t					
	b	0.882	0.686	0.682	0.495
	a	0.118	0.314	0.318	0.505
	n	72	71	99	98
Sdh-1,2					
	100	0.075	0.080	0.026	0.154
	-50	0.924	0.919	0.974	0.846
	n	69	65	95	94

Table 6.7. Allozyme observed (*Ho*) and Hardy-Weinberg expected (*He*) heterozygosities, and probabilities of genotype conformance to Hardy-Weinberg equilibrium (*P*). River Philip and Tasmania 1 data from Ward et al 1994. na = locus not analysed (see text). nd = not determined (see text). Means (and standard error) are given for each sample.

Gene		River Philip	Tasmania 1	Tasmania 2	Gaden
Ada					
	Ho	0.028	0.032	0.000	na
	He	0.028	0.031	0.000	
	P	1.000	1.000		
Aat-3					
	Ho	0.278	0.154	0.202	0.128
	He	0.259	0.142	0.198	0.119
	P	0.687	1.000	1.000	1.000
Idh-4					
	Ho	0.292	0.234	0.273	0.153
	He	0.268	0.207	0.250	0.141
	P	0.663	0.562	0.476	1.000
Mdh-3,4					
	Ho	0.194	0.308	0.510	0.441
	He	0.296	0.437	0.482	0.470
	P	0.013	0.025	0.683	0.685
Me-2					
	Ho	0.083	0.046	0.000	0.000
	He	0.080	0.045	0.000	0.000
	P	1.000	1.000		
Pgm-1-t					
	Ho	nd	nd	nd	nd
	He	0.208	0.443	0.434	0.500
Sdh-1,2					
	Ho	nd	nd	0.053	0.309
	He	0.139	0.147	0.051	0.261
	P			1.000	0.144
Mean±SE	*Ho*	0.175 ± 0.052	0.155 ± 0.053	0.173 ± 0.081	0.206 ± 0.077
	He	0.183 ± 0.039	0.207 ± 0.064	0.202 ± 0.075	0.249 ± 0.082

Table 6.8. Probability values from pair-wise chi-square tests for four allozyme loci showing significant heterogeneity among the four samples. Values in bold are significant after Bonferroni adjustment for multiple tests

	River Philip	Tasmania 1	Tasmania 2
Mdh-3,4			
Tasmania 1	**0.006**		
Tasmania 2	**<0.001**	0.157	
Gaden	**<0.001**	0.343	0.631
Me-2			
Tasmania 1	0.470		
Tasmania 2	**0.003**	0.058	
Gaden	**0.007**	0.045	-
Sdh-1,2			
Tasmania 1	1.000		
Tasmania 2	0.062	0.055	
Gaden	0.027	0.065	**<0.001**
Pgm-1-t			
Tasmania 1	0.031		
Tasmania 2	0.033	1.000	
Gaden	**<0.001**	0.030	**0.004**

Table 6.9. Estimates of allele-frequency variance, F_K, between each pair of the four collections of Atlantic salmon for (I) eight microsatellite loci and (II) seven allozyme loci. Top listed collection is assumed to represent the progenitor for the bottom collection. The weighted average multilocus estimate of F_K was calculated as the average over loci, weighted by the number of independent alleles per locus

6.9.I. Microsatellite loci

	River Philip			Gaden		Tasmania 1
Locus	Tasmania 1	Tasmania 2	Gaden	Tasmania 1	Tasmania 2	Tasmania 2
cmrSs1.10	0.0720	0.0407	0.0939	0.0600	0.0774	0.0538
cmrSs1.14	0.0769	0.0114	0.0244	0.0405	0.0196	0.0631
cmrSs1.22	0.1428	0.1094	0.0734	0.0348	0.0285	0.0436
20.19	0.0287	0.0816	0.0919	0.0200	0.0894	0.0618
5.27	0.0808	0.0162	0.0478	0.0047	0.0087	0.0260
D30	0.0509	0.0779	0.0477	0.1223	0.0380	0.1434
F43	0.0519	0.0671	0.1675	0.1201	0.0799	0.1152
F49	0.0792	0.0136	0.0044	0.0714	0.0229	0.0289
Weighted average F_K	0.0926	0.0756	0.0729	0.0505	0.0407	0.0627

6.9.II. Allozyme loci

	River Philip			Gaden		Tasmania 1
Locus	Tasmania 1	Tasmania 2	Gaden	Tasmania 1	Tasmania 2	Tasmania 2
Mdh-3,4	0.1042	0.2284	0.1839	0.0132	0.0031	0.0288
Aat-3	0.0560	0.0153	0.0802	0.0026	0.0275	0.0135
Me-2	0.0114	0.0840	0.0840	0.0460	0.0000	0.0460
Idh-4	0.0154	0.0013	0.0667	0.0192	0.0502	0.0078
Sdh-1,2	0.0003	0.0502	0.0598	0.0515	0.1905	0.0589
Ada	0.0003	0.0280	na	na	na	0.0320
Pgm-1-t	0.2447	0.2216	0.5945	0.1207	0.1394	0.0008
Weighted average	0.0618	0.0898	0.1782	0.0422	0.0821	0.0268

Table 6.10. The estimated mean effective population sizes N_K , standard deviation (SD) and 95% confidence limits (CL) of the stocks based on the weighted average multilocus estimates of variances in allele frequencies (F_K see Table 6.4) between the two samples over the assumed number of generations of isolation (G).

	G		N_K	Est. SD	CL
River Philip/Tasmania 1	10	Microsatellites	65.2	19.7	34.2, 126.8
		Allozymes	106.2	74.4	20.7, 405.1
River Philip/Tasmania 2	11	Microsatellites	87.9	25.4	47.6, 149.8
		Allozymes	70.8	43.7	15.3, 202.7
River Philip/Gaden	11	Microsatellites	92.1	24.9	52.2, 167.6
		Allozymes	33.2	20.6	6.5, 89.6
Gaden/Tasmania 1	2	Microsatellites	40.4	13.1	13.8, 50.2
		Allozymes	51.7	43.5	5.2, 231.4
Gaden/Tasmania 2	4	Microsatellites	65.8	20.5	34.6, 119.3
		Allozymes	28.1	17.1	5.2, 85.9
Tasmania 1/Tasmania 2	1	Microsatellites	10.1	3.4	4.9, 18.9
		Allozymes	35.6	36.3	5.1, -479.5

Table 6.11. Number of alleles expected to remain, *na*, in a population after the number of generations of isolation (G), based on the mean effective population size estimates, N_K, presented in Table 6.10. The actual number of alleles present in the progenitor (P) and derived (D) stocks are presented as well as the number of alleles observed to be lost or gained; goodness-of-fit of loss of alleles against the expected loss due to random genetic drift was tested by chi-square with one degree of freedom.

			Number of alleles					
	G		*na*	P	D	Lost	Gained	χ^2
River Philip/Tasmania 1	10	Microsatellites	30.5	38	30	10	2	0.594
		Allozymes	13.2	14	14			0.119
River Philip/Tasmania 2	11	Microsatellites	31.6	38	35	8	5	0.227
		Allozymes	12.8	14	12	2		0.082
River Philip/Gaden	11	Microsatellites	31.8	38	41	8	11	0.326
		Allozymes	10.5	12	11	1		0.000
Gaden/Tasmania 1	2	Microsatellites	34.8	41	30	13	2	6.541*
		Allozymes	11.0	11	12		1	0.000
Gaden/Tasmania 2	4	Microsatellites	35.9	41	35	10	4	4.335*
		Allozymes	9.7	10	10			0.137
Tasmania 1/Tasmania 2	1	Microsatellites	26.1	30	35	1	6	1.698
		Allozymes	13.9	14	12	2		19.741*

(* significant, $\chi^2 = 3.841$, $P = 0.05$).

7. Y-CHROMOSOME MARKER

7.1 MICROSATELLITE MARKERS

7.1.1 MALE/FEMALE SPECIFICITY OF EIGHT LOCI

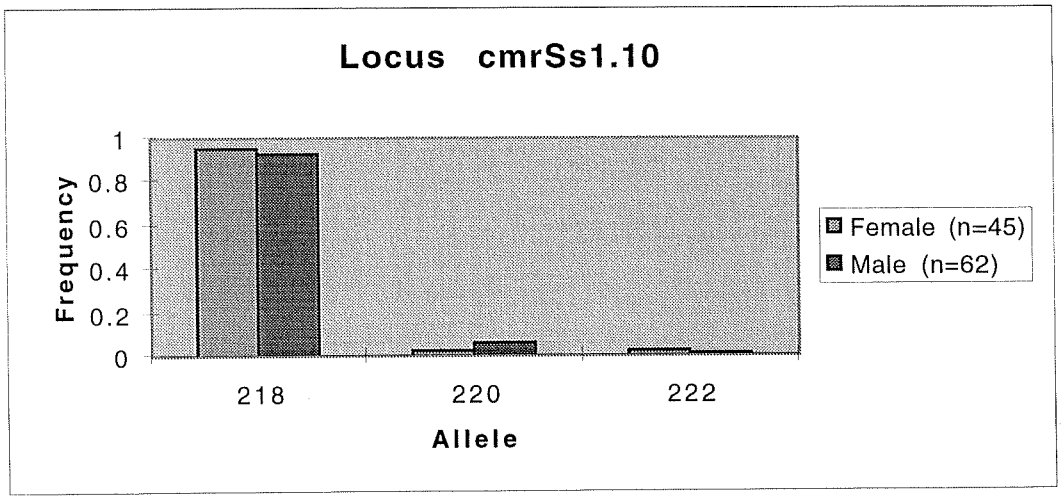
A total of 107 sexed individuals (45 females and 62 males) were screened for the eight microsatellite loci used in the population study (Section 6.2) to determine whether any alleles were sex-specific. These comprised 96 individuals from the Tasmania 2 sample (Section 6.1) and eleven (6 males and 5 females) mature individuals; all were provided by SALTAS. The microsatellite analyses was performed as described in section 6.2.

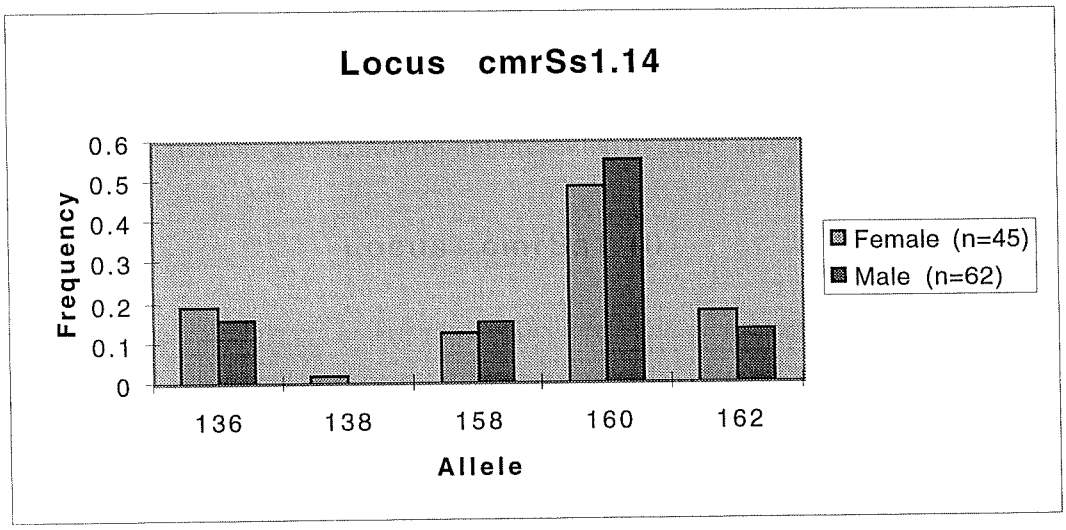
For each of the eight loci, there was no significant difference between the two sexes observed in either allele frequencies (Fig 7.1), genotype frequencies or percentage of homozygotes (Table 7.1). There was therefore no indication that any of these eight loci occurred on or near the male-sex determining region of the genome (Y-chromosome), i.e. were sex specific.

Table 7.1 Percentage of each sex that were homozygous at each of eight microsatellite loci. Number of females 45 and males 62, except for locus *F43* (44 females and 60 males). Probability values from chi-square tests of allele and genotype frequencies.

Locus	Percentage homozygotes		Chi-square *P* values	
	Female	Male	Genotypes	Alleles
cmrSs1.10	93.3	87.1	*0.189*	*0.237*
cmrSs1.14	28.9	33.9	*0.582*	*0.308*
cmrSs1.22	13.3	19.4	*0.978*	*0.714*
20.19	51.1	46.7	*0.532*	*0.908*
5.27	64.4	62.7	*0.504*	*0.329*
D30	48.9	53.2	*0.566*	*0.647*
F43	65.9	65.0	*0.793*	*0.597*
F49	84.4	62.7	*0.058*	*0.162*
Mean	56.3	55.3		

Figure 7.1. Comparison of the frequency distribution of each allele by sex for the eight microsatellite loci. Alleles are presented as allele size in basepairs.





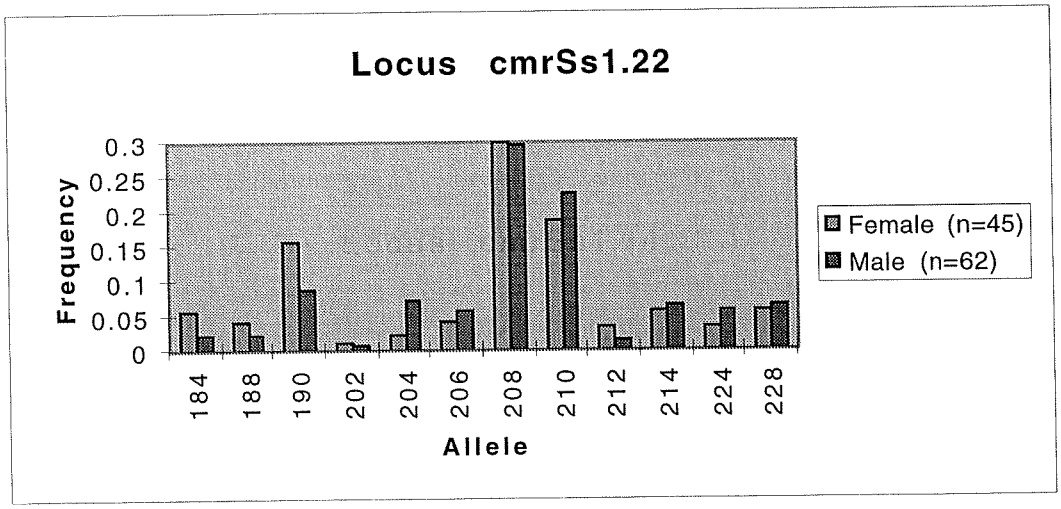
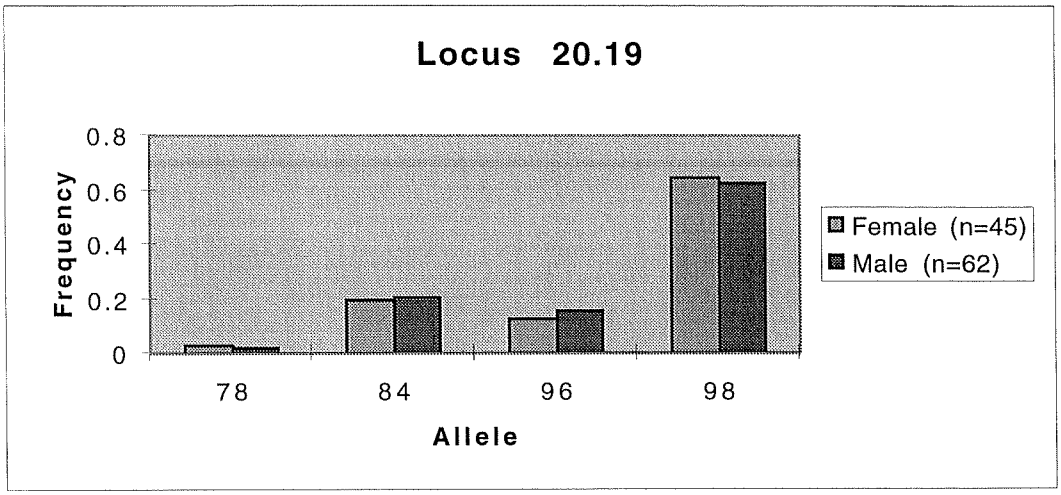
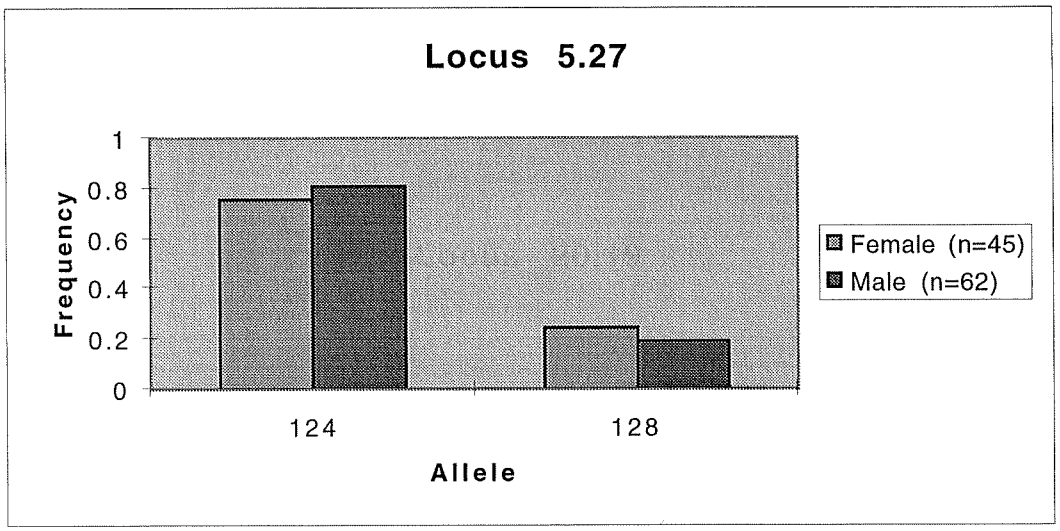


Figure 7.1 cont.





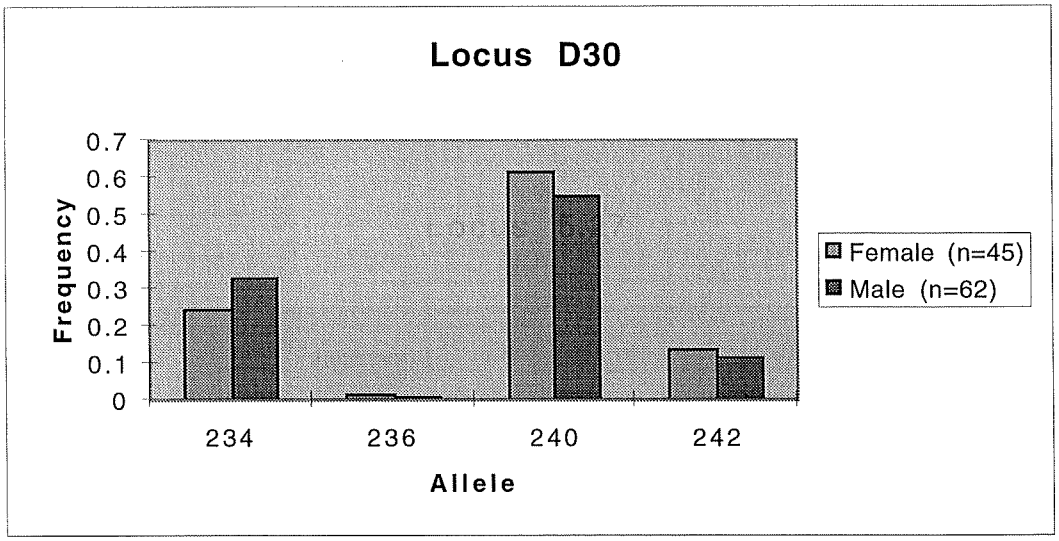
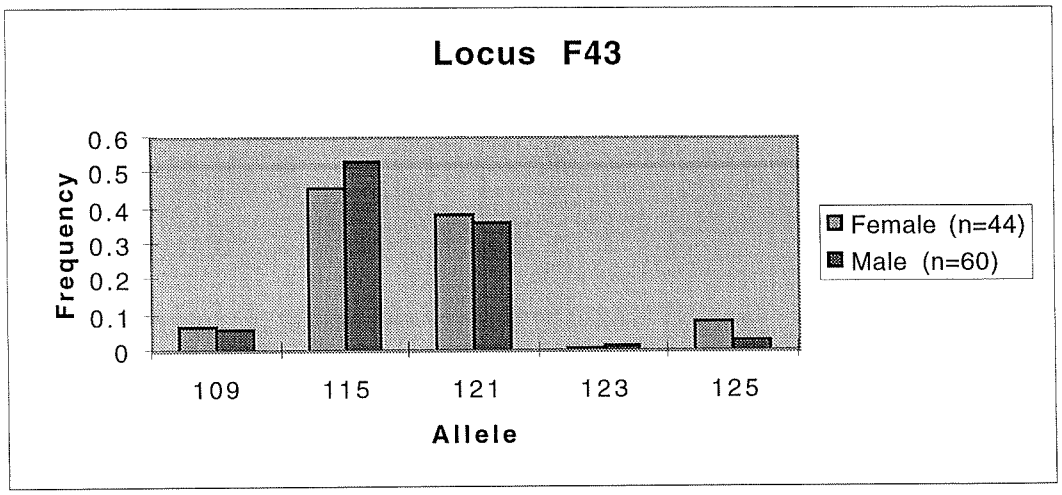
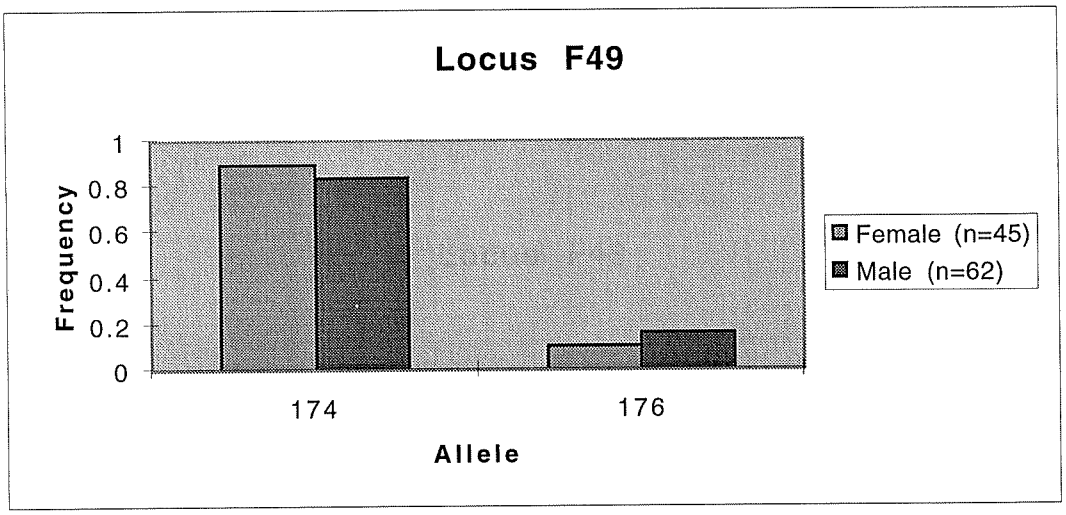


Figure 7.1 cont.





7.1.2 CLONE SCREENING WITH TOTAL GENOMIC MALE/FEMALE PROBES

7.1.2.1 Method

Male and female genomic probes were prepared for hybridization with CA-repeat microsatellite containing clones derived from a male fish (Section 5.1). Genomic DNA was extracted from muscle tissue (50 mg) of ten individual SALTAS fish (5 males and 5 females), suspended in 50 µl TE buffer and treated with RNAase A (Pharmacia) at 10 µg/ml DNA solution. The DNA of the five individuals of each sex was pooled and 50 µl digested with 50 units of each of the restriction enzymes *Alu*I and *Sau*3AI. A 1µg aliquot of each pooled genomic DNA digest was random-prime labelled with the Boehringer DIG High prime kit, according to kit instructions. However, before labeling, the denatured male and female probes were quenched in liquid nitrogen to keep DNA in single-stranded form. Labeling was allowed to proceed overnight at 37˚C.

Plasmid DNA from the 200 microsatellite-containing clones (prepared according to section 5.1.6) from the male genomic DNA library described in section 5.1.1, was denatured at 95˚C for 10 minutes, before chilling on ice. Aliquots (1µl) of plasmid DNA were spotted out in duplicate on each of two positively charged nylon membranes (Hybond N+, Amersham). DNA was fixed by exposing to UV light for 10 seconds on each side of the membrane.

Membranes were prehybridised for 2 hrs at 65˚C in standard hybridization buffer (section 5.1.4). Both male and female probe solutions were made up to a final concentration of 7.5 ng/ml, by adding 7.5µl of probe to 50 ml hybridization solution. Hybridization was performed overnight at 65˚C. Membranes were washed in 2 x SSC, 0.1% SDS at room temperature with gentle rocking. This was followed by stringency washes in 0.5 x SSC, 0.1% SDS at the hybridization temperature of 65°C. Detection was performed as described in section 5.1.5.

7.1.2.2 Results

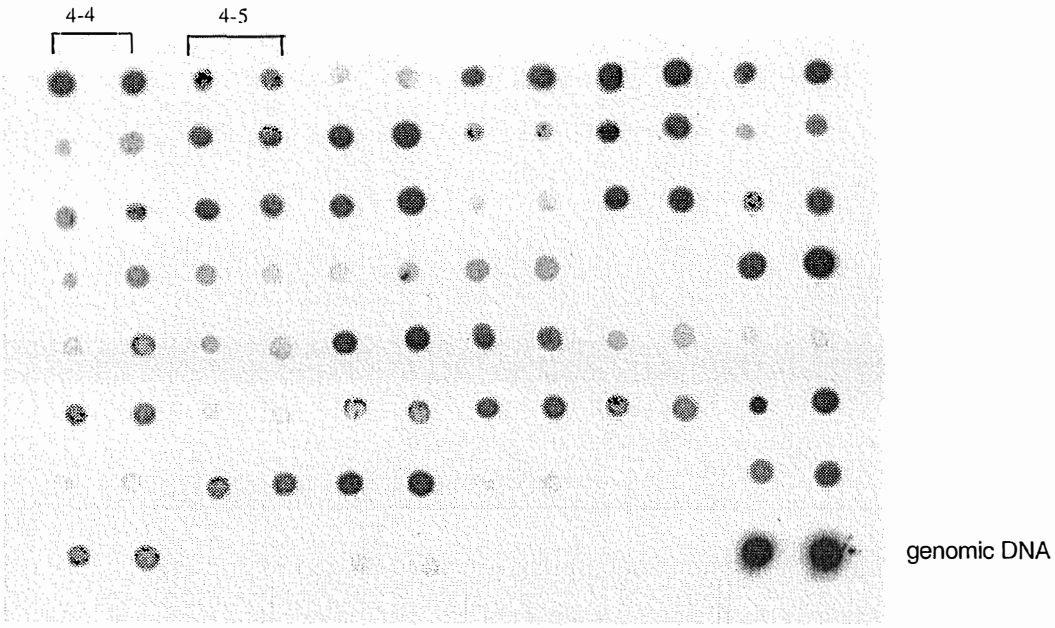
All CA microsatellite-containing clones reacted with both male and female probes. Although the signal intensity varied between individual clones, each clone reacted to an equal extent with each of the male and female probes. A typical result is illustrated in Figure 7.2. The microsatellite-containing clones that were tested did not appear to contain any sex specific sequence.

Sequences that differed in copy number with sex or that are unique to males could potentially have been detected by this method. However, microsatellite clones are expected to be problematic to probe in this manner since a large number of similar CA repeats present in the total genomic probes would react non-specifically with these clones.

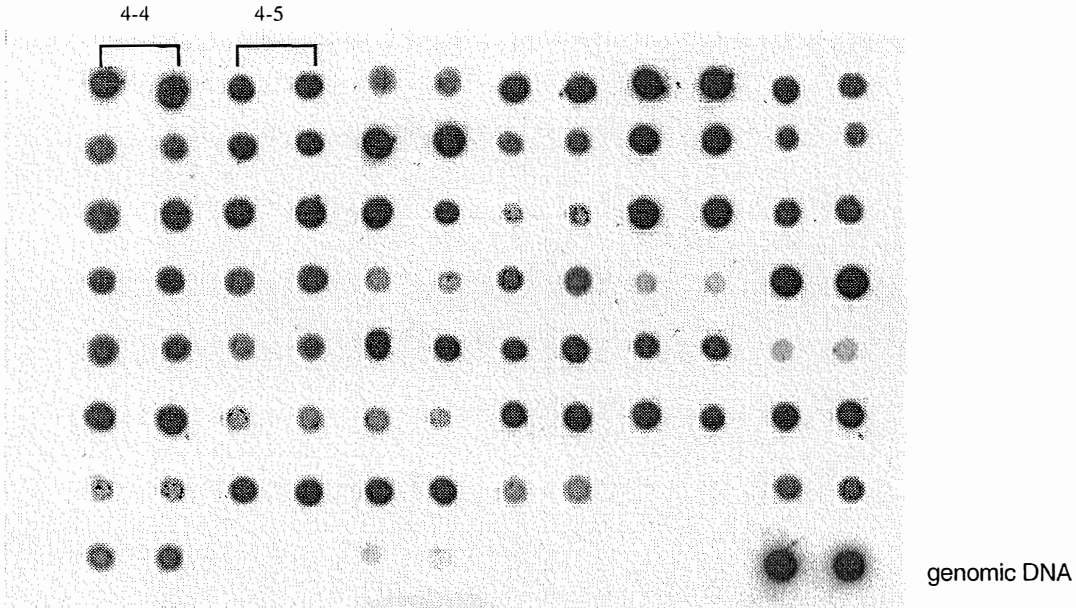
Our original plan was to screen the entire genomic plasmid-based library (Section 5.1) with the two sex probes. However, after the results of the above experiment with 200 of the potential 9000 clones, it was concluded that success in locating a unique Y-chromosome (male) DNA fragment was unlikely with this approach. The stringency of this approach was questioned as the clones contained other genomic material as well as the microsatellite. In excess of 4 million clones would have to be screened to give a good chance of finding the particular gene fragment of interest (Sambrook et al. 1989, p 9.3). A review of recent molecular techniques suggested that the identification of unique DNA fragments required the application of a different approach, most likely a form of subtractive hybridization.

Figure 7.2 Dot blot screening of microsatellite-containing clones with total genomic male and female probes

Dot blot of clones probed with male genomic DNA probe



Dot blot of clones probed with female genomic DNA probe



Shown above are dot blots performed on randomly selected clones. Each clone (44 shown here) was blotted out in duplicate on each of two filters

7.2 REPRESENTATIONAL DIFFERENCE ANALYSIS (RDA)

7.2.1 RDA AS PREFERRED METHOD

In recent years a number of molecular techniques have been designed for identification of genetic differences or specific genes, and those considered for this project included:

- RAPD (random amplification of polymorphic DNA). A polymerase chain reaction (PCR) based genetic marker technique that identifies DNA polymorphisms between individuals. It is similar to the microsatellite approach (Section 7.1.1), and with a similar low potential for success. Unlike the microsatellite primers, however, the primers for RAPD loci will amplify a random sequence of the genome and repeatability with this technique is not 100%. Other laboratories have applied the RAPD technique to identifying a sex marker for Atlantic salmon but with no success (R. Devlin pers comm.).
- AFLP™ (Amplified fragment length polymorphism). A novel and powerful DNA fingerprinting technique based on the selective amplification of restriction fragments from a total DNA digest of genomic DNA. The technique involves restriction of the genomic DNA and selective amplification of a subset of all the fragments. This method is described as a cost-effective way to generate marker-assisted genetic maps for discriminating between closely related lines that differ for a particular trait of interest.
- Subtractive hybridization. A group of techniques designed for isolating rare differentially expressed genes or non-coding DNA differences. The basic principle of this methodology is that two sets of DNA are prepared, one with the specific gene or sequence of interest, and one without. The two samples of DNA are then digested with a restriction enzyme that cuts the DNA into a large number of small fragments. The hybridization process then removes the DNA fragments that are common to the two samples – in principle leaving only unique fragments containing the gene or sequence of interest. The various techniques have proved useful for the isolation of differentially expressed genes in studies of a variety of organisms, and have been applied recently in other sex identification studies in fish (e.g. Nakayama et al. 1997).
- PCR-Select™ cDNA. A subtraction technique for sequences of DNA common to two cDNA samples that suppresses undesirable polymerase chain reaction (PCR) amplification, rather than by physically separating single-stranded and double-stranded DNAs. However, this method works with differentially expressed messenger RNA (mRNA). The mRNA associated with the Y-chromosome genes in Atlantic salmon is assumed to be present only during a short interval of time (within the first days or weeks after hatching) when the practice of sex-reversal in hatcheries is carried out. To undertake this method would have required initial identification of this window of opportunity to obtain the mRNA, time for this was not available during the project.
- Differential display. A PCR based alternative method to subtractive hybridization that requires the combination of various primers to efficiently amplify all cDNA derived from mRNA. In principle similar to the RAPD technique, but more robust and potentially more powerful.
- Representational Difference Analysis (RDA). Our preferred option that combines a modified subtractive hybridization technique between two DNA samples, with a DNA enrichment by second order kinetics of self-reassociation. This method uses genomic DNA and is not reliant on obtaining mRNA for the particular gene(s) under investigation.

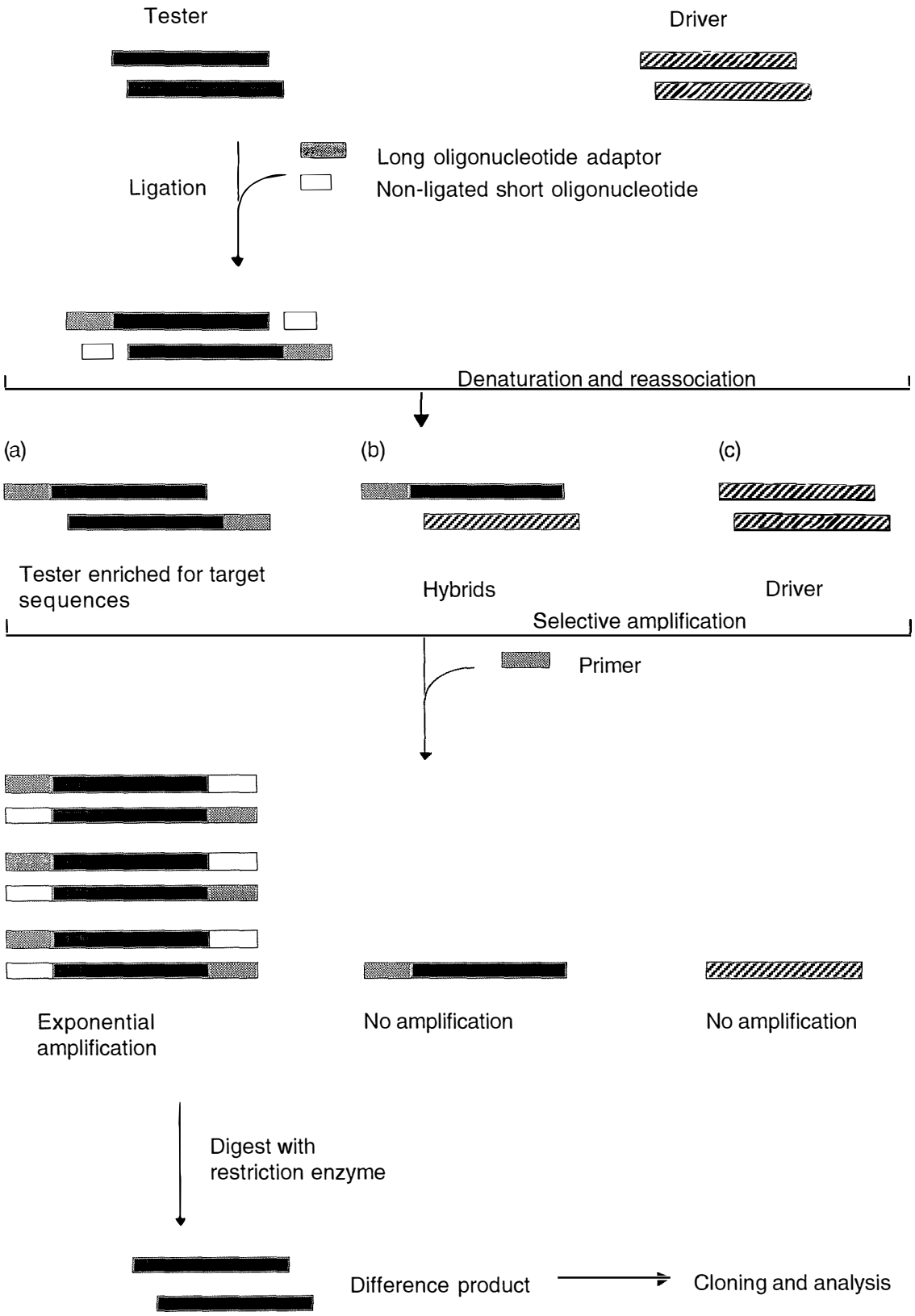
7.2.2 BASIC PRINCIPLES OF RDA

RDA is a modified DNA subtractive hybridization method designed to clone differences between two complex genomes. The subtractive methodology is such that "tester" DNA (in our case a male, Y-chromosome carrying individual) is hybridized to an excess of "driver" (female) DNA to remove similarities or common hybridizing sequences and enrich for the "target" or rare sequences which are unique to the "tester". In order for efficient hybridization and enrichment to occur, the tester (male) DNA and driver (female) DNA are cut into small fragments with a restriction enzyme. Specific DNA primers are attached to one end of each strand of the tester DNA fragments and used to specifically amplify these pieces of DNA by PCR. DNA is denatured to single-stranded and allowed to hybridise in the presence of excess driver (female) DNA. A variety of hybrid molecules form between the tester and driver where there is a common sequence. These are then "subtracted out" as they do not have the specific primers on both ends, and can only amplify linearly; driver DNA cannot amplify and is also subtracted out. Target fragments unique to the tester will then self-reanneal to form double stranded DNA with adaptors on both ends. In theory, only self-reannealed tester DNA with adaptors at each end will amplify exponentially, thereby enriching. This procedure is illustrated in Figure 7.3.

The original RDA protocol (Lisitsyn et al. 1993) was designed for application to the human genome, and involves many complex steps. A detailed protocol was provided to us by the authors (Lisitsyn and Wigler), however, modifications were required due to the different genome and the different constraints of laboratory equipment (e.g. PCR thermocycler).

We have undertaken three RDA experiments with Atlantic salmon to locate a male or Y-chromosome marker. Details of the modified RDA protocol, used in our third, and most complete experiment are given in Appendix 15.1. Variations to this protocol used in the two earlier experiments are noted below.

Figure 7.3 Schematic outline of RDA procedure (adapted from Lisitsyn 1995).



The tester DNA sample is ligated to oligonucleotide adaptors. Tester and driver DNA samples are then mixed, denatured and allowed to reassociate. Three types of reannealed duplexes form: (a) target strands self-reanneal with adaptor oligonucleotides adaptors on both 5' ends; (b) driver-tester hybrids have adaptor on one end only; and (c) self-reannealed driver. The target fragments are selectively amplified after filling in the oligonucleotide cohesive ends with Taq DNA polymerase. The procedure is repeated after ligation of a new adaptor to the amplified DNA fragments.

7.3 METHOD DEVELOPMENT AND TROUBLESHOOTING

7.3.1 EXPERIMENT 1 –*Bgl*II REPRESENTATION

7.3.1.1. Protocol

The modified protocol provided in Appendix 15.1 was used with the following exceptions.

- DNA was extracted from muscle tissue by the modified CTAB method (Grewe et al, 1993). As recommended in the original RDA protocol, a *Bgl*II (a 6-base restriction enzyme with recognition site AGATCT) representation was made. After a 3 hr incubation, digests were checked by 1% agarose gel electrophoresis. Digests did not look complete so a further 10 units of restriction enzyme were added to each tube and the digest continued for a further 2 hrs.
- The PCR reactions contained 50µl diluted ligate, 4µl R*Bgl*II 24-mer (Set 1 adapters, see Appendix 15.1.2) as primer and 7.5 units AmpliTaq DNA polymerase. A positive control was not included due to a limited amount of primers and poor amplifications of human gDNA (the recommended control).
- Driver (70µg) and tester (5µg) DNA were digested with 10 units/µl *Bgl*II. They were performed in 800µl and 200µl volumes for driver and tester DNAs respectively and incubated for 2 hrs at 37°C. Driver DNA was resuspended in 58.3µl TE (1.2µg/µl) and tester DNA in 62.5µl TE (0.08 µg/µl). A 50 x excess of driver was used for hybridization instead of the recommended 100 x excess; 20µg of driver amplicon (see 15.1.1) digest was hybridized with 0.4µg tester DNA ligate.
- For selective pre-amplification reactions, standard PCR reactions were prepared using 40µl of diluted hybridized DNA, 7.5 units of AmpliTaq DNA polymerase and 5µl J*Bgl*II 24-mer (Set 2, Appendix 15.1.2) as primer.
- Following the mung bean nuclease treatment (removes adaptors from single-stranded DNA, leaving only double-stranded molecules, Appendix 15.1.9.3) and inactivation, selective amplification reactions were set up containing 40µl nuclease-treated difference product, 5µl J*Bgl*II 24-mer and 7.5 units of AmpliTaq DNA polymerase. PCR amplification was for 20 cycles, followed by an additional three after an extra 7.5 units of AmpliTaq polymerase was added. In optimising selective amplification conditions, the PCR product was sampled after every second cycle from cycle 14 to cycle 24.

7.3.1.2. Results

After three rounds of selective hybridization, the *Bgl*II representation failed to reveal any unique difference products. The DP1, DP2 and DP3 amplicons (first, second and third hybridization difference products formed by PCR amplification, Appendix 15.1.1) all looked very similar or identical to the representation amplicon with heavy repetitive element bands accompanied by a faint smear (Figure 7.4). The DP1 amplicon should consist of heavy smears in the amplifiable range (~150 to 1500 bp) with perhaps a couple of heavier repetitive element bands visible (Lisitsyn pers. comm.). Only smaller products (up to 500 bp in size) appeared to be amplifying. These small fragments were thought to indicate that most of the difference product was denatured to single-stranded DNA. With large and/or repetitive DNA targets, the number of PCR cycles required to generate difference product should be reduced as overamplification can deplete the Taq polymerase enzyme, causing the PCR reaction to slow, resulting in single-stranded DNA (Lisitsyn, pers. comm.). However, changing the number of cycles of the selective amplification step to between 14 and 24 cycles had no effect upon the PCR amplification size profile, only on the amount of product.

This our first experiment failed to produce any difference products between the tester and driver for further analysis. The only bands visible on the gel were those also present in the initial representation amplicon (15.1.1), and believed to be highly repetitive elements such as satellite repeats. The lack of difference may have been due to poor digestion, denaturing or other technical errors with the protocol.

7.3.2 EXPERIMENT 2 - *Bgl*II, *Hind*III, *Bam*HI REPRESENTATIONS

7.3.2.1. Protocol

For our second RDA experiment it was decided to repeat the *Bgl*II representation, and to include a *Hind*III (6-base restriction enzyme recognition site AAGCTT) and *Bam*HI (6-base restriction enzyme recognition site GGATCC) representation. DNA was extracted and digested in the same manner as in Experiment 1. In this experiment the standard PCR mixture contained 2µl RBglII 24-mer, 8µl dNTPs, 10µl diluted ligate and 7.5 units Taq.

7.3.2.2. Results

The *Bgl*II representation amplicons resulting from this second experiment looked more encouraging than those from the first (Fig. 7.4), producing larger amplification products. In the first RDA experiment more ligate had to be used to get efficient amplification; 50µl diluted ligate with 4µl R*Bgl*II 24-mer and 16µl dNTPs compared with 10µl diluted ligate with 2µl R*Bgl*II 24-mer and 8µl dNTPs. In this second experiment there was more efficient amplification of the representation amplicon allowing direct scaling down of the reagents compared with the original protocol. This suggested that the restriction digests may have been better this time or that the ligation of adaptors may have been more efficient. This improved result is not completely understood since DNA was prepared and digested, and adaptors ligated, in the same way as in Experiment 1.

The *Bam*HI and *Hind*III representation amplicons did not amplify very well under the conditions used, although some tubes amplified better than others. Neither of these representations appeared superior to the *Bgl*II representation.

The improvement in the *Bgl*II representation amplicons was not considered significant enough to yield improved hybridization results so the experiment was not taken further than the PCR amplification of the representation amplicon, and advice was sort on the problems being experienced.

7.3.3. DNA EXTRACTION AND RESTRICTION DIGEST IMPROVEMENT

The results we obtained with Experiments 1 and 2 were discussed with researchers experienced with the RDA technique. Two Melbourne laboratories were visited, and discussions held with Dr Eric Moses, Royal Women's Hospital, Carlton and Drs Jill Maddox and Peter Roche, School of Veterinary Science, University of Melbourne, Parkville (Report of visit at Appendix 15.3).

It was agreed that our problems were most likely due to only partial cutting of the genomic DNA, as indicated by the high quantities of large fragments or intact genomic DNA. If DNA is not restricted (cut by the restriction enzyme) to completion, there will be little material in the preferred amplifiable range (i.e. ~150bp to 1500bp), and this was supported by our poor amplification success of the representation amplicons.

Our next step was to attempt to produce better quality, cleaner DNA.

7.3.3.1. Alternative DNA extraction techniques

The presence of inhibitors in the DNA extracts was one possible cause of our restriction digest problems. To determine whether the extraction method influenced the digestion profile, several different extraction methods were attempted. These protocols were recommended by various researchers working with Atlantic salmon DNA.

1. DNA from muscle tissue using a modified CTAB protocol (Grewe et al, 1993) – standard protocol in our laboratory and used for the genetic variation study (Section 6) and RDA experiments 1 and 2.
2. DNA from muscle tissue using the extraction method of Sambrook et al (1989) (with additional extractions, see 15.1.3) – compared with the CTAB protocol, it included an extended proteinase K treatment, more solvent extractions and RNAase treatment early, rather than at the end of the extraction.
3. Qiagen QIAamp® Tissue kit was tested since this procedure does not use phenol, commonly believed to inhibit enzyme restrictions. DNA was extracted from 25 mg muscle tissue according to the recommended protocol (Qiagen, 1996).
4. SDS lysis genomic DNA extraction from blood, recommended by Dr Richard Powell, (University College, Galway, Ireland). Dr Powell recommended using blood, as he believed he had previously experienced some digestion problems when DNA was extracted from Atlantic salmon muscle.
5. Modified version of Taggart et al. (1992) protocol for DNA extraction from muscle tissue (provided by Dr John Taggart, Stirling University, Scotland) – essentially similar to that of Sambrook et al (1989).
6. Salt precipitation of genomic DNA from blood (provided by Dr Lars-Erik Holm, Danish Institute of Agricultural Science, Denmark). This protocol was promising since it was a salt precipitation method, therefore not requiring any organic extractions.

The latter three protocols for DNA extraction are presented in Appendix 15.2.

7.3.3.2. Additional Restriction Digest Precautions

Additional precautions were taken to ensure good quality restriction enzyme digestions, including:

- heating the DNA to 60°C for 10 minutes to ensure DNA is fully suspended
- pre-equilibration of all reagents except the enzyme at 37°C for 4 hrs
- overnight digest incubation
- addition of a further aliquot of enzyme and incubation for an additional 2 hrs
- inclusion of Bovine serum albumin (BSA) in some reactions
- changing digest volume reactions to optimise digest volume
- restriction enzymes tested from a different company (Stratagene), as the quality of our enzymes may have been compromised due to shipping conditions
- a variety of restriction enzymes were tested to determine if the problem was enzyme specific. Enzymes tested include: *Bgl*II, *Bam*HI, *Hind*III, *Dpn*II, *Eco*RI

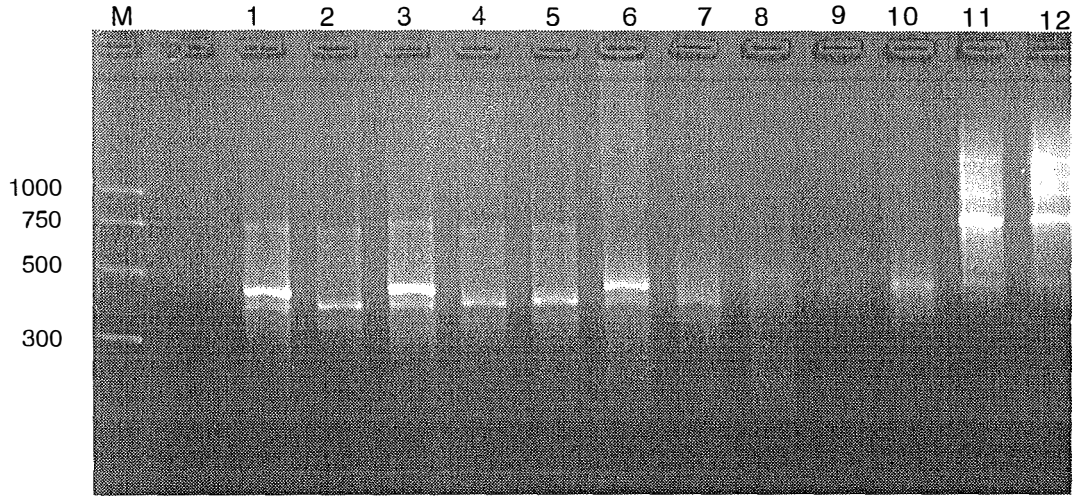
7.3.3.3. Results

The DNA extraction methods of Powell (pers. comm.) and Sambrook et al. (1989) were found to be the most reliable for production of large quantities of unsheared DNA. The method from Lars-Erik Holm (pers. comm.) did not reliably produce large quantities of DNA, and the method from Taggart (pers. com.) had problems with precipitation of EDTA. The Qiagen method resulted in poor quality (sheared) and quantity of DNA.

Restriction digest efficiency with each DNA extraction method was tested. All genomic DNAs gave fairly similar digests to each other (Fig. 7.5). Despite taking the additional precautions with the restriction digests, all were similar to that described in our first experiment. There still appeared to be significant amounts of intact or very large DNA fragments. Digest profiles from the various 6-base restriction enzymes were all similar. However, digestion with a 4-base restriction enzyme (*Dpn*II, recognition sequence GATC) produced a smaller size range of DNA fragments and alleviated, though not completely, the problem of incomplete digests.

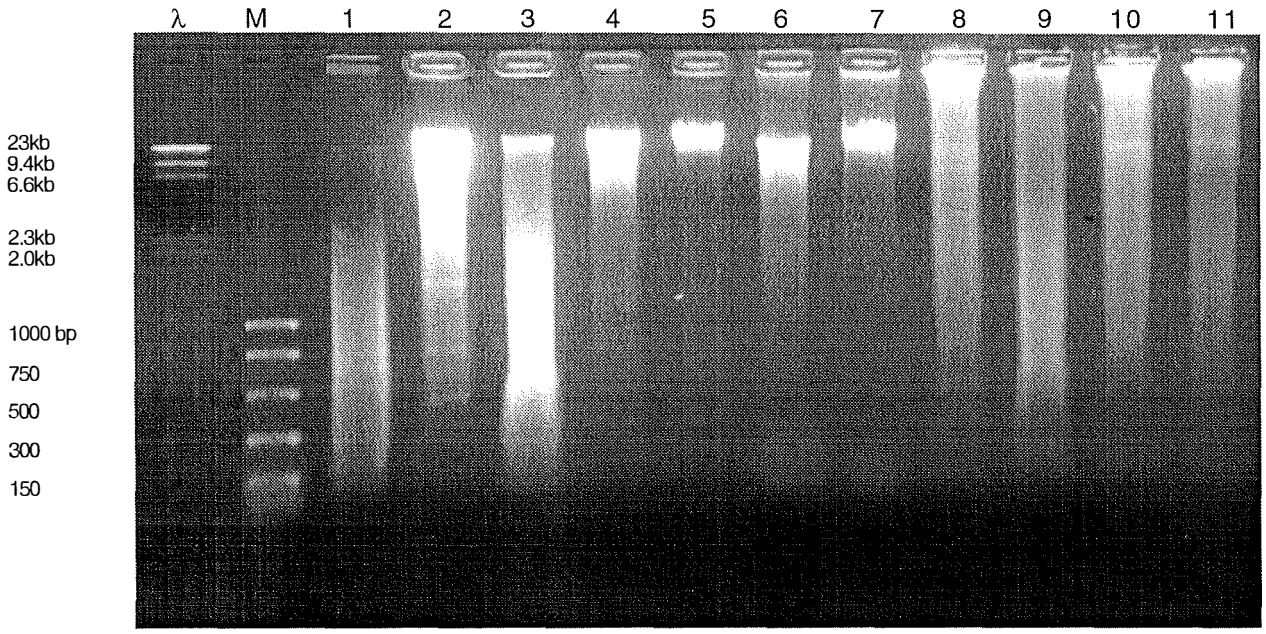
It was concluded that complete digestion failure for Atlantic salmon DNA was most likely due to the presence of satellite DNA, or to the nature of the nucleotide sequence within the Atlantic salmon genome; similar restriction patterns result from prawn DNA (pers. comm. Steve Moore, CSIRO Tropical Agriculture, Brisbane). Having established that this unusual restriction profile is characteristic of Atlantic salmon genomic DNA (Taggart pers. comm.), certain alterations to the RDA protocol were made prior to commencing Experiment 3 .

Figure 7.4 Comparison of BglII representations of experiments 1 and 2



Approximately 200 ng of each product was loaded on a 2% TBE agarose gel. M, PCR markers (Promega); lanes 1-10 Experiment 1 and lanes 11,12 Experiment 2. Lane 1, RBglII driver amplicon, uncut; lane 2, RBglII driver amplicon, BglII cut; lane 3, RBglII tester amplicon, uncut; lane 4, RBglII tester amplicon, cut; lane 5, RBglII tester amplicon, cut and purified; lane 6, DP1 uncut; lane 7, DP1, BglII cut; lane 8, DP2 uncut; lane 9, DP2 BglII cut; lane 10, DP3; lane 11, Experiment 2 RBglII driver amplicon, uncut; lane 12, Experiment 2 RBglII tester amplicon, uncut

Figure 7.5 Comparison of restriction digests performed on genomic DNAs resulting from various extraction methods



λ HindIII marker; M, PCR markers (Promega); lane 1, human genomic DNA Sau3A digest; lanes 2-5, gDNA extracted by method of Powell (blood); lanes 6,7, gDNA extracted by method of Taggert (muscle); lanes 8-11, gDNA extracted by method of Sambrook (muscle). Lane 2, Powell BglII digest; lane 3, Powell DpnII digest; lane 4, Powell BamHI digest; lane 5, Powell uncut DNA; lane 6, Taggert BamHI digest; lane 7, Taggert uncut DNA; lane 8, Sambrook BglII digest; lane 9, Sambrook DpnII digest; lane 10, Sambrook BamHI digest; lane 11, Sambrook uncut DNA

Despite the various precautions taken with extractions and digestions, there was still significant amounts of very large or intact DNA remaining after digestion.

7.4. RDA EXPERIMENT 3

The full protocol for Experiment 3 is presented in Appendix 15.1.

7.4.1 RESULTS – DIFFERENCE PRODUCTS

Although the genome size of Atlantic salmon is about double the size of humans (haploid genome = 6 x 10^9 bp, Hamada et al. 1982), our earlier results suggested that much of the Atlantic salmon genomic DNA was not included in the restriction enzyme representations due to a failure of some of the DNA to cut, thereby reducing the overall complexity. Given this reduction in complexity, and in view of the failure of 6-base restriction enzyme representations to exhibit difference products, it was decided to try a four-base restriction enzyme (*Dpn*II). Hubank and Schatz (1994) selected a *Dpn*II representation for use with mammalian cDNA, where the effective DNA complexity is only about 1-2% of the total genome and the RDA method could be applied without the need to further reduce the genome complexity. As some Atlantic salmon DNA appears to fail to cut (or remains in large fragments), the complexity of the genome might already be effectively reduced by ten-fold or more - the level generally required in the mammalian genome for efficient hybridization and enrichment of the target fragment (Lisitsyn and Wigler, pers. comm.). The *Dpn*II digests produced a bulk of fragments in the size range ~400 bp to 2000 bp, ideal for production of representation amplicons. A restriction digest with *Dpn*II was also compatible with using the protocol recommended *Bgl*II adaptors.

The *Dpn*II digest (section 15.1.4) consisted of a heavy smear over the amplification range and gave far better representation amplicons compared with a *Bgl*II digest. The *Bgl*II representation amplicons consisted of a couple of dominant heavy bands with relatively poor amplification elsewhere (Fig. 7.6).

Efficient amplification was achieved using only 2µl of amplicon PCR product as template and performing only 10 cycles instead of the usual 20 cycles (section 15.1.6). It was observed that amplification from ligates a few days post-ligation reaction did not amplify efficiently. PCR reactions were pooled and purified, resulting in 18µg *Dpn*II tester, 18µg of *Dpn*II control tester, and 240µg *Dpn*II driver DNA to be split evenly between the positive control driver (driver DNA spiked with lambda DNA, 15.1.3) and the experimental driver.

For the *Dpn*II representation, the first round difference product, DP1, consisted of a smear with a few heavy bands in the 250 to 400 bp region. The experimental amplicon DP1 produced some unique bands compared with the control amplicon DP1, but there were also some of a common size. Additional bands became obvious after the second (DP2) and third (DP3) rounds of selective hybridization (15.1.9), although these all appeared to be common to the control. (Fig 7.7)

Since no obvious lambda *Dpn*II specific bands corresponding to the spiked control DNA (15.1.3) were observed in the control experiment and several bands were in common between the control and experimental difference products, it was decided to probe for lambda-specific bands. This would determine if the positive control had been successful, and whether there had been any cross-contamination of the experiment with lambda DNA. The results of a southern hybridization experiment (see 15.1.10.1) gave good sensitivity levels for the lambda DNA positive control, suggesting that the control RDA experiment had been unsuccessful. That is the "difference products" achieved for the positive control were not derived from lambda DNA, and there was no evidence of cross-contamination of the experiment with lambda DNA. However, failure of the positive control experiment suggests that several of the bands in the experimental difference products might be artifacts that are not real differences between the tester (male) and driver (female) DNA; they could be repetitive elements of the driver. It was

considered worthwhile to clone the observed difference products (15.1.10.2) and test the resulting clones for sex specificity by hybridization (15.1.10.3).

7.4.2. CLONING OF EXPERIMENTAL DIFFERENCE PRODUCTS

Sixteen recombinant colonies resulted from the cloning (15.1.10.2) of first hybridization difference product, DP1 (clones 3-1 to 3-16); 14 recombinant colonies resulted from the cloning of DP2 (clones 4-1 to 4-14) and 21 colonies from DP3 (clones 5-1 to 5-21). Double restriction enzyme digests, *Sac*II/*Nde*I (6-base restriction enzymes with recognition sites CCGCCG and CATATG respectively), showed clones containing inserts ranging in size from about 230 bp up to 420 bp (Fig 7.8 and Table 7.2).

It was anticipated that some of these 51 clones may have been multiple copies of individual clones. Each was therefore sequenced to determine their similarity to each other, and to reported gene sequences deposited in the GenBank database. The latter may assist in the selection of clones, for hybridization experiments, with potentially greater chance of showing a sex difference.

7.4.3. SEQUENCING AND BLAST SEARCHES

The sequencing data from the 51 clones (Appendix 15.4) was checked for similarity to sequences deposited in GenBank using a Blast (Basic Local Alignment Search Tool) search (Table 7.2). Some sequences confirm the identity of the difference products DNA as originating from a salmonid genome, e.g. a 55 bp region of clones 3-5 and 3-6 matched with the end of a published sequence of Chinook salmon (*Oncorhynchus tschawytscha*) prolactin II gene (Xiong et al, 1992), with 96% homology. The sequence coincided with the 3' flanking sequence of the prolactin gene and therefore appears to be intron DNA (non-coding). Clones 3-5 and 3-6 each contained 2 bands when cut with *SacII/NdeI*, and were picked from next to each other on the plate. The start of the DNA sequences for the two clones varied, and it is possible that these are two closely related copies of the same sequence, or allelic variation, but unlikely as they were next to each other on the plate.

The size range of the DP1 clones obtained by sequencing correlated with the agarose gel findings that the *Sac*II/*Nde*I digests resulted in fragments ranging in size from about 220 bp to 420 bp. The first difference product (DP1) appeared to contain six discrete bands which we anticipated would be repeated within the 16 recombinant colonies obtained. However, the sequence data (Appendix 15.4) showed that there were 11 different clones amongst the 12 DP1 clones successfully sequenced. In addition, the sequencing results showed that clones of similar size were not identical as anticipated, but were actually different to each other. For example, clones 3-7, 3-10 and 3-11 are 258, 268 and 272 bp respectively and each has a unique sequence. It therefore appears that each major DP1 band on an agarose gel, might actually be composed of several bands. This tends to suggest there may be too much complexity within the representation to allow efficient subtractive hybridization.

Of the 12 successful DP1 clone sequences , two showed good homology with the flanking sequence of the Chinook salmon prolactin II gene, however, of the nine successful DP2 clone sequences, two had good homology with rainbow trout 28SrRNA; no homology with rainbow trout was observed in DP1 nor with Chinook salmon in DP2. However, of the 19 DP3 product clones, five had some homology with the Chinook prolactin flanking sequence and six had some homology with the rainbow trout 28SrRNA.

Despite having similar band profiles (on agarose gel) to the DP1 and DP2 products, the DP3 product clones showed less variation in clone sequence. There were eight different clones from the 12 successful sequences of DP1 clones, six from nine DP2 clones and

eight from 19 DP3 clones. This result, plus the multiple copies of the two salmonid sequences, may indicate that enrichment has occurred through the successive rounds of subtractive hybridization.

7.4.4. HYBRIDIZATION TESTING OF CLONES FOR SEX DIFFERENCES

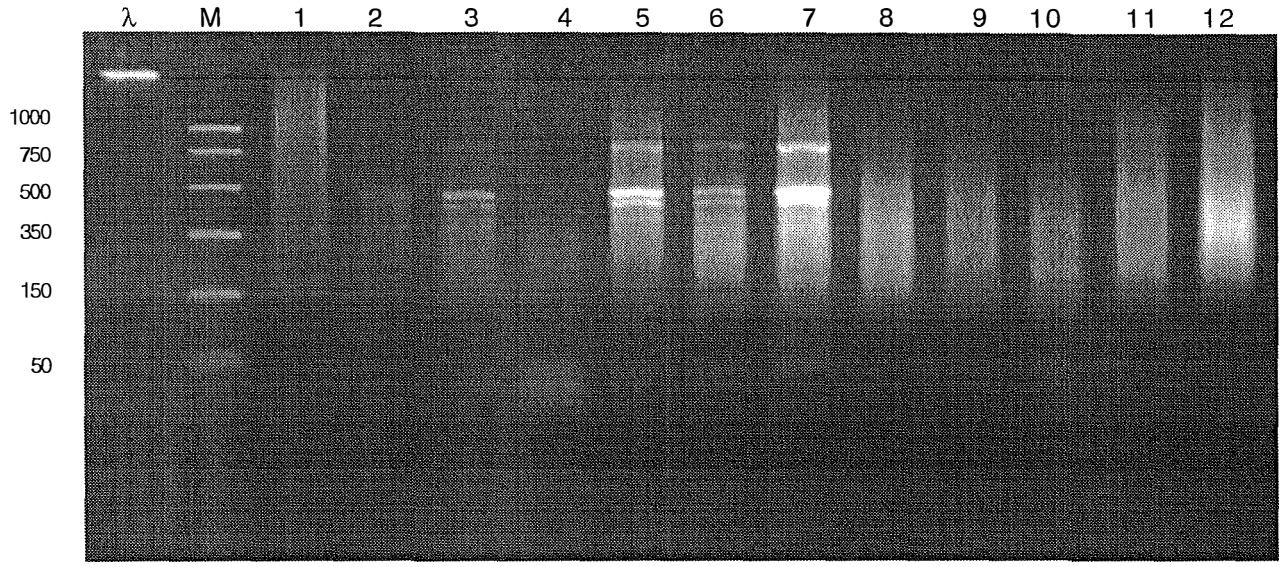
Resolution of difficulties experienced during the initial RDA experiments unfortunately resulted in insufficient project time to test all the difference product clones from the third RDA experiment for sex differences. Initial assessments were made with dot blot hybridization (15.1.10.3) on three clones.

Clone 3-1: This DNA probe from this clone appeared to hybridize with all templates, including the negative control (see 15.1.10.3), but reacted most strongly with the male genomic DNA and with the tester (male) amplicon. It reacted with the female genomic DNA, and relatively weakly with each of the three difference products. The clone 3-1 sequence contains a microsatellite repeat (Appendix 15.4) which could be quite common within the Atlantic salmon genome, explaining the fairly good hybridization with the male and female genomic DNAs. The strength of the reaction with male material, however, suggests that this clone was worth checking further (with Southern hybridization) for possible sex differences.

Clone 3-5: The sequence data for this clone matched well with that of a flanking sequence of the prolactin gene in Chinook salmon (Table 7.2). The clone DNA probe hybridized fairly strongly with DP1, DP2 and DP3, and the tester (male) amplicon, with weaker hybridization with the driver (female) amplicon. The probe also weakly hybridized with male genomic DNA (gDNA), and relatively strongly with two individual female gDNAs, but no hybridization with the pooled female DNA. These inconsistent results may be due to unequal loading of the template, although, DNAs were quantitated so this seems unlikely.

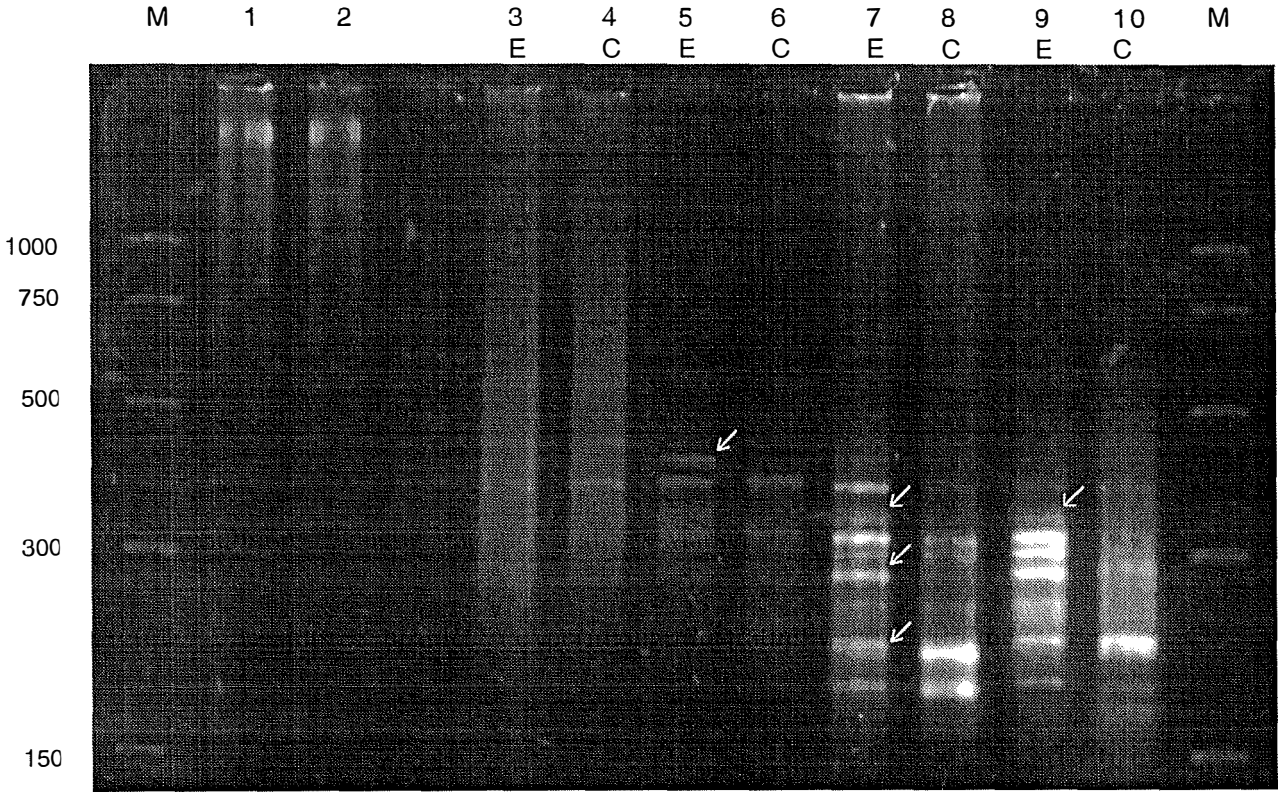
Clone 4-4: On a first hybridization, this probe appeared fairly non-specific. It hybridized with male genomic DNA, with the tester (male) amplicon and weakly with DP2, from which it was derived. However, it also reacted weakly with female genomic DNA. A repeated dot blot showed better results but there was no hybridization with either of the difference products. There was weak reactivity with the human genomic DNA (negative control, 15.1.10.3), but none with the tester or driver amplicons. It was concluded that this clone was a false positive.

Figure 7.6 Comparison of *Bgl*II and *Dpn*II representation amplicons



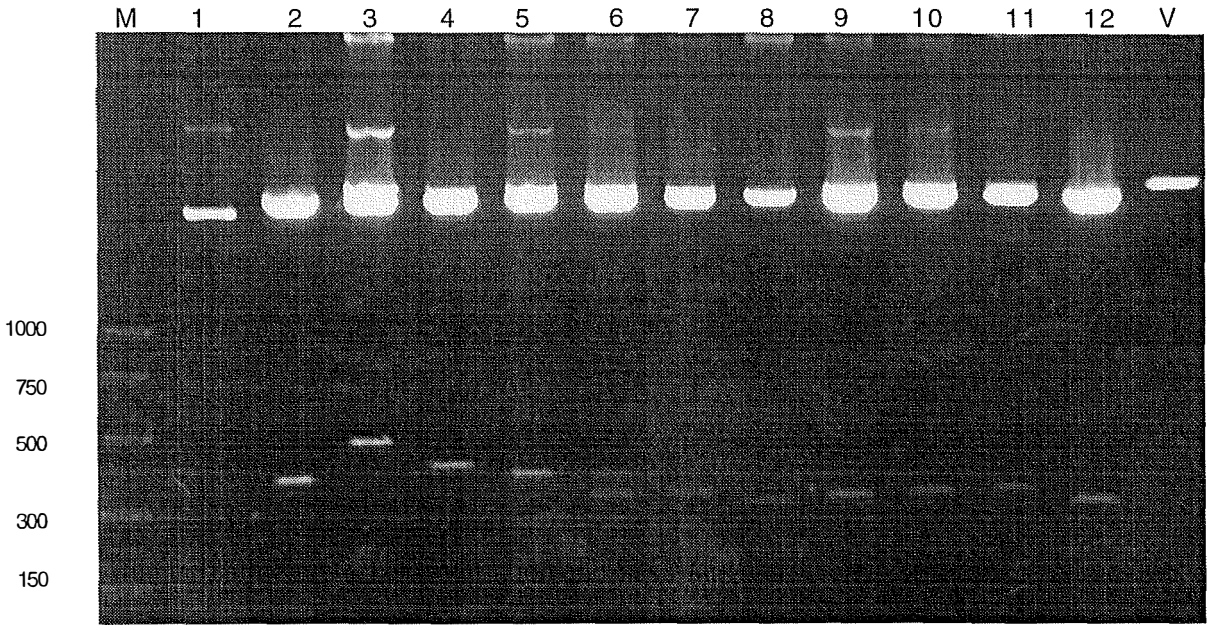
λ HindIII marker; M, PCR markers (Promega); lane 1, human genomic DNA Sau3A digest; lanes 2-7 BglII representations under various PCR conditions; lanes 8-12 DpnII representations under various PCR conditions; lane 2, 10 ng template, 1 µL Taq; lane 3, 20 ng template, 1µl Taq; lane4, 50 ng template, 1µl Taq; lane 5, 10 ng template, 2 µl Taq; lane 6, 20 ng template, 2µl Taq; lane 7, 50 ng template, 2µl Taq; lane 8, 10 ng template, 1µl Taq; lane 9, 20 ng template, 1µl Taq; lane 10, 50 ng template, 1µl Taq; lane 11, 10 ng template, 2µl Taq; lane 12, 20 ng template, 2µl Taq

Figure 7.7 Results of third RDA experiment, illustrating various difference products



M, PCR markers (Promega); lane 1, genomic driver DNA, DpnII digested; lane 2, genomic tester DNA, DpnII digested; lane 3, experimental tester amplicon; lane 4, control tester amplicon; lane 5, experimental DP1; lane 6, control DP1; lane 7, experimental DP2; lane 8, control DP2; lane 9, experimental DP3; lane 10, control DP3. Arrows indicate difference products which appear specific to Atlantic salmon

Figure 7.8 Restriction enzyme analysis of various RDA clones



M, PCR marker (Promega); lane 1, clone 3-1uncut; lane 2, clone 3-1 SacII/NdeI digested; lane 3, clone 3-2 SacII/NdeI digest; lane 4, clone 3-3 SacII/NdeI; lane 5, clone 3-4 SacII/NdeI; lane 6, clone 3-5 SacII/NdeI; lane 7, clone 4-1 SacII/NdeI; lane 8, clone 4-2 SacII/NdeI; lane 9, clone 4-3 SacII/NdeI; lane 10, clone 4-4 SacII/NdeI; lane 11, clone 5-2 SacII/NdeI; lane 12, clone 5-3 SacII/NdeI; V, pGEM®-T cloning vector SacII/NdeI

7.5. CONCLUSIONS

The RDA subtractive hybridization protocol has been modified, and RDA difference products have been obtained in our third experiment. We are confident that it is a suitable technique for the investigation of a sex specific DNA fragment in Atlantic salmon. The disappointing aspect of the work has been the amount of unforeseen resources required to modify the protocol and overcome initial difficulties working with the Atlantic salmon genome. These difficulties have been overcome, but due to time restrictions the resulting clones (DNA fragments) from the difference products have not been fully tested to identify any sex differences, and await further resources to complete the experiment.

There was a degree of inconsistency with the three dot blot hybridizations that were completed, and it is recommended that the Southern blot hybridization technique, although more time consuming, may provide better resolution. In addition, it is recommended that more specific DNA probes from the clones be prepared by excising the insert DNA (clone specific) from the plasmid vector prior to labeling, or PCR amplification of the insert fragment. Such probes, of approximately 250 bp (cf. 3000 bp for all plasmid), would be expected to be more specific than those derived from labeling the whole plasmid DNA.

Table 7.2: Blast data of sequenced RDA clones (sequence data Appendix 15.4.2).

Clone	Size (bp)	Blast results
3.1	264	No significant matches
3.2	379	Human lung (72 bp, 65% homology); *Mus musculus* immunoglobulin S alpha region (37 bp, 83%)
3.3	very poor sequence	No blast data
3.4	260	No significant matches
3.5	256	Chinook salmon prolactin II gene (55 bp, 96% homology)
3.6	266	Identical to 3-5 but with some variation in first few bases
3.7	258	Soares ovary tumor (41 bp, 80% homology)
3.8	very poor sequence	No blast data
3.9	poor sequence	Arabidopsis IGF (54 bp, 70% homology)
3.10	286	Arabidopsis IGF (53 bp, 69% homology)
3.11	272	Soares fetal heart *Homo sapien* cDNA (57 bp, 68% homology)
3.12	very poor sequence	No matches
3.13	328	Adenovirus type 12/rat right junction (82 bp, 97%)
3.14	387	*Rattus norvegicus* cell cycle checkpoint protein kinase (28 bp, 100%)
3.15	216	No significant matches
3.16	244	Human CpG DNA (59 bp, 89%)
4.1	poor sequence	No matches
4.2	225	Rainbow trout 28SrRNA (181 bp, 91% homology)
4.3	186	Human procadherin mRNA (185 bp, 84%)
4.4	poor sequence	No significant matches
4.5	224	Rainbow trout 28SrRNA (181 bp, 90% homology) cf clone 4-2
4.6	poor sequence	No significant matches
4.7	268, incomplete	*Fugu rubripes* GSS sequence (130 bp, 93%)
4.8	118	*Gallus gallus* tendon fibroblast cDNA (32 bp, 93%)
4.9	215	*Schistosoma japonicum* cDNA clone (70 bp, 82%); *Gallus gallus* tendon fibroblast cDNA (36 bp, 88%)
4.10	387	*Schistosoma japonicum* cDNA clone (158 bp, 82%)
4.11	326	Adenovirus type 12/ rat right junction (82 bp, 87%); (63 bp; 96%)
4.12	164	*Gallus gallus* tendon fibroblast cDNA (35 bp, 91%)
4.13	very poor sequence	No blast data

Table 7.2 cont.

Clone	Size (bp)	Blast results
4.14	no insert	No blast data
5.1	very poor sequence	No blast data
5.2	131	Chinook salmon prolactin II gene (57 bp, 94%)
5.3	167	Rainbow trout 28SrRNA (167 bp, 97%)
5.4	377	Human mRNA - HIV associated non-Hodgkin's lymphoma (27 bp, 96%)
5.5	162	*Rattus norvegicus* trg mRNA (81 bp, 70%)
5.6	245	Human CpG (59 bp, 93%)
5.7	135	Rainbow trout 28SrRNA (135 bp, 95%)
5.8	225	Rainbow trout 28SrRNA (181 bp, 91%)
5.9	197	Atlantic salmon, SINE repeat region (124 bp, 91%)
5.10	169	Chinook salmon prolactin II gene (57 bp, 94%)
5.11	177	No significant matches
5.12	266	Chinook salmon prolactin II gene (57 bp, 96%)
5.13	266	Flour beetle mariner transposase gene (46 bp, 76%)
5.14	259	Chinook salmon prolactin II gene (57 bp, 98%)
5.15	224	Rainbow trout 28SrRNA (193 bp, 84%)
5.16	168	Rainbow trout 28SrRNA (162 bp, 95%)
5.17	157	No matches
5.18	167	Rainbow trout 28SrRNA (167 bp, 97%)
5.19	267	Chinook salmon prolactin II gene (57 bp, 96%)
5.20	165	*Drosophila melanogaster* (52 bp, 71%)
5.21	very poor sequence	No blast data

8. NON-DESTRUCTIVE SAMPLING

Genetic markers will be an essential tool in future selective breeding programs. Such markers will be used for identifying broodstock to provide particular qualitative or quantitative traits, e.g. sex reversed individuals (i.e. XX 'males') for producing all-female smolt, faster growth, meat quality and disease resistance. These individuals will need to be identified as early in the life cycle as possible and will become valuable broodstock, therefore a non-lethal and non-destructive method for obtaining DNA is required. The availability of PCR (polymerase chain reaction) technology enables the amplification of DNA from small initial amounts of tissue. We chose to investigate the use of fin clippings as a source of DNA as they are often used to mark individual fish, and the procedure has little apparent effect on the individual.

DNA was extracted from ca. 50 mg of adipose fin tissue from two individuals using the modified CTAB method. The DNA was of good quality and yield, providing approximately five times the amount of DNA compared to an equivalent sample of muscle tissue. Each of the eight microsatellite loci used in the genetic variation study (section 6.2) were successfully amplified with the DNA from adipose fin tissue, and the products run on the ABI sequencer.

The results suggest that non-lethal sampling through the use of adipose fin clippings will be reliable and provide suitable quality DNA for future molecular genetic research.

Analysis of genetic variation at microsatellite loci using DNA extracted from old scale samples of Atlantic salmon has recently been reported (Nielsen et al. 1997). That study successfully extracted and analysed DNA from scales collected over 60 years ago. Thus non-lethal sampling of broodstock is possible through either fin clippings and/or scale samples.

9. BENEFITS

The principal benefactor of the molecular genetics research undertaken in Projects 95/80 and 96/347 has been SALTAS, although their customers (the Tasmanian salmon growers) are the ultimate benefactors. The salmon farming industry benefits from the knowledge that the extensive SALTAS breeding program is maintaining a high degree of genetic variation within their cohorts. In addition, an Atlantic salmon genomic library has been developed that contains several hundred potential microsatellite markers that could be further developed for use in future genome mapping and selective breeding projects.

Unfortunately, a sex specific probe has not been identified. However, there have been significant gains in expertise in the use of subtractive hybridization techniques that will have benefits for future molecular genetics projects in the aquaculture industry.

10. INTELLECTUAL PROPERTY

No commercial intellectual property has occurred out of the work to date.

11. FURTHER DEVELOPMENT

The following areas of research in relation to Atlantic salmon in Australia are recommended following the interpretation of the results we have obtained and international efforts in salmonid genetic research.

Genetic variation
The genetic variation in the Australian population (both Tasmania and New South Wales) should be monitored regularly, using both microsatellites and allozymes, to determine whether the observed loss of variation has stabilized, is continuing or in fact is reversed with the continued spawning of mixed age groups.

The sampling protocol to obtain the annual samples should be investigated to determine the sample size and 'random selection' protocol required to effectively represent the population.

Additional microsatellite markers be investigated to make available ten or more informative makers.

Investigate the actual contribution of all parents to the progeny of spawning involving the mixing of ova and sperm from multiple individuals; this should be through the use of microsatellite and AFLP markers for progeny testing.

Obtain archived scales from River Philip fish caught in the late 1960's/early 1970's and examine microsatellite DNA variation (see Nielsen et al. 1997 for feasibility of extracting DNA from historical scales). This would provide an assessment of whether the 1993 sample accurately reflects the genetic composition of the mid 1960's fish used to establish the Australian population. Initial contact with the Canadian authorities suggests scales from the early 1970's are available.

Microsatellite markers

Our sequencing results indicate a very high rate of success with our library screening and we predict that nearly all of the 200 positive clones isolated from the library would contain CA microsatellite repeats. It is suggested that a larger number of these di-nucleotide repeat microsatellite loci be fully developed, as well as screening and development of some tri- and tetra-nucleotide repeat microsatellites.

These markers are required for use in genome mapping, Quantitative Trait Loci (QTL) maps and progeny testing, all of which will be useful in future selective breeding programs. Other microsatellite loci may also prove more informative in assessing changes in genetic variation than those currently available.

The CSIRO Division of Marine Research has been invited (subject to local funding) to be associated with a three-year (1997-2000) international project, SALMAP, funded by the European Commission to contribute towards the development of a genetic marker map for salmonids. Such a map will be a decisive tool in the search for loci/chromosome regions that have major effects on economically important traits (QTLs). An important part of SALMAP is the production of highly informative DNA markers with an emphasis on microsatellites. Such collaboration by Australia would strengthen the international mapping effort and bring major benefits to the local industry in regards to future selective breeding programs. In particular via access to far more potential markers and genetic information useful to the local industry than would be possible if we worked in isolation. The hundreds of potential microsatellite markers that could be characterized from our existing DNA library could be applied to SALMAP reference families, and in future to local family lines.

It is further suggested that the use of AFLP markers be investigated in Atlantic salmon, for use both in genome mapping, and in the investigation of Y-chromosome markers.

Y-chromosome marker
It is suggested that further RDA subtractive hybridization experiments be undertaken, and that the remaining clones from the difference products obtained in our final experiment be fully investigated for possible sex specific fragments. The ground work

has been done, and a modified protocol established for the technique with Atlantic salmon; difference products have been found but not fully investigated.

To reduce the DNA complexity between the tester (male) and driver (female) DNAs in the experiments, the use of material from individual family lines is strongly recommended. None are currently available in Tasmania, but participants with SALMAP have access to family line material from Scotland.

12. STAFF

Dr Nicholas G. Elliott	25%
Ms. Anne Reilly	100%
Dr Peter M. Grewe	5%
Dr Robert D. Ward	5%

13. ACKNOWLEDGEMENTS

The authors thank Harry King (SALTAS) for the supply of samples, assistance with sexing parr and his continued support and interest in this work; Sam Crocker (NSW Fisheries, Gaden Trout Hatchery) for the supply of samples and useful comments; Peter Lee and co-workers at Aquatas for supply of blood samples; Richard Powell (University College, Galway) and Catherine Clabby (University of Hull, England) for use of microsatellite loci; Lars-Eric Holms (Danish Institute of Agricultural Science), Bjorn Hoyheim (Norwegian College of Veterinary Medicine), John Devlin (Fisheries and Oceans, Canada), Richard Powell and John Taggart (Stirling University, Scotland) for sharing ideas, knowledge and protocols; Kathy Haskard (CSIRO) for statistical advice; Nikolai Lisitsyn (University of Pennsylvania) and Michael Wigler (Cold Spring Harbor Laboratory) for the RDA protocol details and advice; and Eric Moses (Royal Womens Hospital, Melbourne), Jill Maddox and Peter Roche (University of Melbourne) for their time and advice on the RDA protocol.

14. REFERENCES

Baldocchi, R.A., Tartaglia, K.E., Bryda, E.C., and Flaherty, L. 1996. Recovery of probes linked to the *jcpk* locus on mouse chromosome 10 by the use of an improved representational difference analysis technique. *Genomics* 33: 193-198.

Bermingham, E., Forbes,S.H., Friedland, K. and Pla, C., 1991. Discrimination between Atlantic salmon (*Salmo salar*) of North American and European origin using restriction analyses of mitochondrial DNA. *Can. J. Fish. Aquat. Sci.* 48: 884–893.

Birky, C.W., Fuerst, P. and Maruyama, T., 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, 121: 613–627.

Birt, T.P., Green, J.M. and Davidson, W.S., 1991. Mitochondrial DNA variation reveals genetically distinct sympatric populations of anadromous and nonadromous Atlantic salmon, *Salmo salar. Can. J. Fish. Aquat. Sci.*, 48: 577–582.

Boehringer Mannheim, 1995 The DIG System User's Guide for Filter Hybridization. Boehringer Mannheim Gmbtl Biochemica, Germany.

Callen, D.F. Thompson, A.D., Shen, Y., Phillips, H., Richards, R.I., Mulley, J.C. and Sutherland, G.R., 1993. Incidence and origin of 'null' alleles in the (AC)*n* microsatellite markers. *Amer. J. Hum. Genet.,* 52: 922–927.

Clabby, C., 1996. Cloning and molecular analysis of tandem repetitive DNA elements from the Atlantic salmon, *Salmo salar* L. Unpublished PhD thesis, Department of Microbiology, University College Galway, Ireland.

Clifford S.L., McGinnity, P. and Ferguson, A., 1998. Genetic changes in an Atlantic salmon population resulting from escaped juvenile farm salmon. *J. Fish Biol.* 52: 118–127.

Cross, T.F. and King, J., 1983. Genetic effects of hatchery rearing in Atlantic salmon. *Aquaculture*, 33: 33–40.

Devlin, R.H., McNeil, B.K., Groves, T.D. and Donaldson, E.M. (1991). Isolation of a Y-chromosome DNA probe capable of determining genetic sex in Chinook salmon(*Oncorhynchus tshawytscha*). *Can. J. Fish. Aquat. Sci.*, 48, 1606–1612.

Gharrett, A.J., and Shirley, S.M., 1985. A genetic examination of spawning methodology in a salmon hatchery. *Aquaculture* 47: 245–256.

Gile, S.R. and Ferguson, M.M., 1990. Crossing methodology and genotypic diversity in a hatchery strain of rainbow trout (*Oncorhynchus mykiss*). *Can. J. Fish. Aquat. Sci.,* 47: 719–724.

Grewe, P.M., Krueger, C.C., Aquadro, C.F., Birmingham, E., Kincaid, H.L. and May, B., 1993. Mitochondrial DNA variation among lake trout (*Salvelinus namaycush*) strains stocked into Lake Ontario. *Can. J. Fish. Aquat. Sci.*, 50: 2397–2403.

Hanada, H., Petrino, M.G. and Kakunga, T. (1982). A novel repeated element with Z-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* 79: 6465–6469.

Hebert, P.D.N. and Beaton, M.J. (1993). Methodologies for Allozyme Analysis using Cellulose Acetate Electrophoresis. 2nd Ed. Helena Labs, Beaumont, TX.

Hedgecock, D., Chow, V. and Waples, R.S., 1992. Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. *Aquaculture* 108: 215–232.

Hedgecock, D. and Sly, F., 1990. Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas. Aquaculture*, 88: 21–38.

Hubank, M., and Schatz D.G. 1994. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Research* 22 (5): 5640-5648.

Lessios, H.A., 1992. Testing electrophoretic data for agreement with Hardy-Weinberg expectations. *Mar. Biol.,* 112: 517–523.

Lisitsyn, N.A. 1995. Representational difference analysis: finding the difference between genomes. *TIG* 11: 303–307.

Lisitsyn, N., Lisitsyn, N., and Wigler, M. 1993. Cloning the Differences Between Two Complex Genomes. *Science* 259: 946-951.

Lisitsyn, N. and Wigler, M. 'Applications of Representational Difference Analysis to Detection of Genetic Lesions in Cancer', Cold Spring Harbor Laboratory, New York, pp 1-17.

McConnell, S., Hamilton, L., Morris, D., Cook, D., Paquet, D., Bentzen, P. and Wright, J., 1995a. Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon. *Aquaculture*, 137: 19-30.

McConnell, S.K., O'Reilly, P., Hamilton, L., Wright, J.M., and Bentzen, P. 1995b. Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): genetic differentiation of North American and European populations. *Can. J. Fish. Aquat. Sci.* 52: 1863-1872.

Nakayama, I., Grossi, T., Foresti, F. and Chourrout, D. 1997. Atypical sex chromosomes revealed by subtracted probes in *Leporinus elongatus.* 6th International Symposium on Genetics in Aquaculture. University of Stirling, Scotland, June 1997.

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from small number of individuals. Genetics, 89: 583–590.

Nielsen, E.E., Hansen, M.M. and Loeschcke, V., 1997. Analysis of microsatellite DNA from old scale samples of Atlantic salmon *Salmo salar*: a comparison of genetic composition over 60 years. *Mol. Ecol.* 6: 487-492.

O'Connell, M. and Wright, J.M., 1997. Microsatellite DNA in fishes. *Rev. Fish Biol.* 7: 331-363.

Pemberton, J.M., Slate, J., Bancroft, D.R. and Barrett, J.A., 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.*, 4: 249-252.

Perkin Elmer 1995. ABI Prism™ Dye Terminator Cycle Sequencing Ready Reaction Kit Protocol. P/N 402078 Revision A.

Qiagen September 1996 'QIAamp® Blood Kit and QIAamp Tissue Kit Handbook', pp. 18-21.

Reed, K.M., Bohlander, S.K. and Phillips, R.B. (1995). Microdissection of the Y chromosome and fluorescence *in situ* hybridization analysis of the sex chromosome of lake trout, *Salvelinus namaycush. Chromosome Research*, 3, 221–226.

Richardson, B.J., Baverstock, P.R. and Adams, M. (1986). *Allozyme Electrophoresis. A Handbook for Animal Systematics and Population Structure.* Academic Press, Sydney.

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

Ruzzante, D.E., Taggart, C.T., Cook, D. and Goddard, S. 1996. Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: microsatellite DNA variation and antifreeze level. *Can. J. Fish. Aquat. Sci.,* 53: 634–645.

Sambrook, J., Fritsch, E.F. and Maniatis, T., 1989. 'Molecular Cloning: A Laboratory Manual' 2nd edn. (Cold Spring Harbor Laboratory: New York).

Sanchez, J.A., Clabby, C., Ramos, D., Blanco, G., Flavin, F., Vazquez, E. and Powell, R., 1996. Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon). *Heredity*, 77: 423–432.

Scribner, K.T., Gust, J.R. and Fields, R.L., 1996. Isolation and characterization of novel salmon microsatellite loci: cross-species amplification and population genetic applications. *Can. J. Fish. Aquat. Sci.*, 53: 833–841.

Slettan, A., Olsaker, I. and Lie, O., 1996. Polymorphic Atlantic salmon, *Salmo salar* L., microsatellites at the *SSOSL438*, *SSOSL439* and *SSOSL444* loci. *Anim. Genet.*, 27: 57–58.

Simon, R.C., McIntyre, J.D. and Hemmingsen, A.R., 1986. Family size and effective population size in hatchery stock of coho salmon (*Oncorhyncus kisutch*). *Can. J. Fish. Aquat. Sci.* 43: 2434–2442.

Sokal, R.R. and Rohlf, F.J., 1981. Biometry 2nd Ed. W.H. Freeman & Co. New York.

Taggart, J.B., Hynes, R.A., Prodohl, P.A. and Ferguson, A., 1992. A simplified protocol for routine total DNA isolation from salmonid fishes. *J. Fish Biol.* 40: 963–965.

Tessier, N., Bernatchez, L., Presa, P. and Angers, B., 1995. Gene diversity analysis of mitochondrial DNA, microsatellites and allozymes in landlocked Atlantic salmon. *J. Fish. Biol.*, 47 (suppl. A): 156–163.

Waples, R.S., 1989. A generalised approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121: 379–391.

Ward, R.D., Grewe, P.M. and Smolenski, A.J., 1994. A comparison of allozyme and mitochondrial DNA in Atlantic salmon from Tasmania and from the ancestral population in Canada. *Aquaculture*, 126: 257–264.

Weber, J.L. 1990. Informativeness of Human (dC-dA)n.(dG-dT)$_n$ Polymorphisms. *Genomics* 7: 524-530.

Withler, R.E., 1988. Genetic consequences of fertilizing chinook slamon (*Oncorhyncus tshawytscha*) eggs with pooled milt. *Aquaculture* 68: 15–25.

Wright, J.M. and Bentzen, P. 1994. Microsatellites: genetic markers for the future. *Rev. Fish Biol. Fish.*, 4: 384–388.

Xiong, F., Chin, R.A., and Hew, C.L. 1992. A gene encoding chinook salmon (*Oncorhynchus tschawytscha*) prolactin: gene structure and potential cis-acting regulatory elements. *Mol. Mar. Biol. and Biotech.* (1992) 1 (2): 155-164.

Zaykin, D.V. and Pudovkin, A.I., 1993. Two programs to estimate significance of χ^2 values using pseudo-probability tests. *J. Hered.*, 84: 152.

15. APPENDICES

15.1 DETAILED MODIFIED RDA PROTOCOL

15.1.1 DEFINITIONS: 'REPRESENTATION' AND 'AMPLICON'

A "representation" is a reproducible subpopulation of DNA fragments from a genome. Representations that are ten-fold less complex than the mammalian genome are generally required for efficient hybridization and enrichment of target DNA fragments (Lisitsyn and Wigler, detailed protocol). Enzyme restriction is an efficient means of producing a representation.

An "amplicon" is a representation formed by PCR amplification of DNA fragments that arise from cleavage with a particular restriction enzyme, followed by ligation to an oligonucleotide PCR adaptor.

15.1.2 OLIGONUCLEOTIDE PCR ADAPTORS

The sequences of the three sets of oligonucleotide adaptors used in our RDA experiments were:

Set 1:	R*Bgl*II24	5'-AGCACTCTCCAGCCTCTCACCGCA-3',
	R*Bgl*II12	5'-GATCTGCGGTGA-3';
Set 2:	J*Bgl*II24	5'-ACCGACGTCGACTATCCATGAACA,
	J*Bgl*II12	5'-GATCTGTTCATG-3';
Set 3:	N*Bgl*II24	5'-AGGCAACTGTGCTATCCGAGGGAA-3';
	N*Bgl*II12	5'-GATCTTCCCTCG-3'

Adaptors were synthesized by Bresatec as high purity, reverse-phase purified unphosphorylated oligonucleotides. Oligonucleotides were suspended in sterile water at 62 pmol/µl.

15.1.3. SAMPLES

To reduce the number of difference products derived from individual DNA polymorphisms, DNAs of five individual female fish were pooled to be used as driver DNA. The male, tester DNA was derived from one individual. Controls were included to demonstrate whether subtraction was effective. Positive control tester amplicons were prepared from ligate that was comprised of 2µg *Dpn*II digested driver DNA that had been spiked with 32pg λ*Dpn*II digested DNA. This was considered a more relevant positive control for our RDA experiments than using the human gDNA control suggested in the protocol (Lisitsyn et al, 1993).

Genomic DNA was extracted according to Sambrook et al. (1989). Briefly, 50 mg muscle tissue samples were ground, suspended in extraction buffer (10mM Tris. Cl (pH 8.0), 0.1 M EDTA, 0.5% SDS) containing pancreatic RNAase (Pharmacia) at 20µg/ml and incubated at 37˚C for 1 hour. Proteinase K was added to a final concentration of 100µg/ml and incubated at 50˚C overnight. Samples were then extracted twice with phenol and then once with chloroform. DNA was precipitated with 0.2 volumes of 10M ammonium acetate and 2 volumes of ethanol at room temperature. Pellets were washed twice with 70% ethanol. DNA pellets were allowed to resuspend in distilled water overnight.

In an attempt to further clean some genomic DNA samples, the proteinase K step was repeated, followed by additional phenol and phenol/chloroform extractions.

15.1.4. RESTRICTION OF GENOMIC DNA FOR A *Dpn*II REPRESENTATION

Two micrograms of each driver (female) DNA and the tester (male) DNA were digested separately with 20 units of the restriction enzyme *Dpn*II in a 200µl volume, containing bovine serum albumin (NEB) at 100µg/ml. Following overnight incubation, an additional 10 units of restriction enzyme was added and the reaction incubated an additional 2 hrs. Digests were checked by 1% agarose gel electrophoresis. Following digestion, tRNA was added as carrier and the mixture was extracted twice with phenol/chloroform. The DNA was then ethanol precipitated, dried and resuspended in water. The digests were quantitated by fluorometer and adjusted to 0.1 mg/ml. One microgram of each driver DNA digest was pooled prior to ligation, according to Baldocchi et al. (1996).

A similar experiment was set up for a *Bgl*II restriction to compare the 2 representation amplicons.

15.1.5. LIGATION OF ADAPTORS

One microgram of each digested driver or tester DNA was ligated to 465 pmol of each of the R*Bgl*II oligonucleotide PCR adaptors (Set 1) in a total reaction volume of 30µl. The reaction was placed in a glycerol-filled heating block that had been preheated to 55˚C. The block was then placed in the fridge until the temperature dropped to 15˚C. The reaction tubes were cooled on ice for 3 minutes before the addition of 400 units T4 DNA ligase (New England Biolabs). Ligations were performed overnight at 15˚C. Each ligate was diluted to 1 ml with TE buffer (10mM Tris.Cl pH8.0; 1mM EDTA) containing tRNA (20µg/ml).

15.1.6. PCR AMPLIFICATION OF REPRESENTATION AMPLICON

PCR reactions were performed as 100µl reactions instead of the 400µl reactions recommended in the protocol since our Perkin Elmer 9600 PCR machine is limited to 100µl reactions.

Four tubes were used for preparation of both tester and control tester amplicons, and 24 for preparation of both driver and control driver amplicons. Each tube contained 20 µl 5 x PCR buffer (335 mM Tris-HCl pH 8.9, 20 mM $MgCl_2$, 50 mM 2-mercaptoethanol, 0.5 mg/ml bovine serum albumin), 16 µl dNTPs (4mM stock), and 52 µl water. 10 µl diluted ligate (section 6.3.5) and 2 µl R*BglII* 24-mer were added to reactions and placed in the thermocycler preheated to 72˚C for 5 minutes to dissociate the 12-mer. To fill in 3' recessed ends, 10 units of AmpliTaq DNA polymerase (Perkin Elmer) were added to each tube, mixed and incubated for 5 minutes at 72˚C. Amplification then proceeded for 20 cycles of 95˚C for 1 min and 72˚C for 3 min, with the last cycle followed by a 72˚C hold for 10 min. To check the quantity and quality of amplicons, 10 µl of PCR product was run on a 2% TBE agarose gel.

15.1.7. REMOVAL OF ADAPTORS FROM AMPLICONS

All tester (male) tubes were combined into one and driver (female) tubes were combined into six tubes. These were extracted twice with equal volumes of phenol/chloroform. The DNA was precipitated by addition of a 1/10 volume of 3 M sodium acetate and 600 µl of isopropanol. The mixture was incubated for 15 min on ice and centrifuged to recover the DNA. The pellets were washed with 70% ethanol and dried. Pellets were resuspended overnight in 40 µl TE buffer and quantitated by fluorometer.

To cleave the adaptors, the amplicons were digested with *Dpn*II enzyme. 10 µg of tester and positive control amplicons and 120 µg of driver and positive control amplicons were digested using 10 units/µl *DpnII*. Digests were performed in 800µl and 200µl volumes for driver and tester DNAs respectively and incubated for 2 hrs at 37˚C.

To aid in precipitation, 10µg tRNA was added to tester amplicon digest. DNAs were cleaned and precipitated as outline above. Driver amplicon was resuspended in 58.3µl TE (1.2µg/µl) and tester amplicon in 62.5µl TE (0.08 µg/µl). The driver amplicon was vortexed twice for 30 sec to suspend the highly concentrated DNA solution. Driver DNA was diluted ten-fold and approximately 0.2, 0.4 and 0.6 µg of driver and tester amplicon digests were loaded on a 2% agarose gel along with the same quantities of undigested amplicon to check digestions.

15.1.8. CHANGE OF ADAPTORS ON TESTER AMPLICON

To clean away PCR adaptors present in the tester amplicon DNA digest, the Qiagen QIAquick PCR purification kit was used. The digested tester amplicon (5 µg DNA in total) was made up to 100 µl in TE buffer and five volumes of Buffer PB was added. The sample was purified according to kit instructions and eluted by addition of 30 µl TE buffer. The column was allowed to stand for 1 minute before spinning for one minute. The DNA concentration was checked on a gel and also quantitated by fluorometer. Adaptor set 2 (*JBgl*II) (section 15.1.2) was then ligated to 1 µg of purified tester DNA amplicon as in section 15.1.5. DNA concentration was adjusted to 10 µg/ml by adding TE buffer containing tRNA (20µg/ml).

15.1.9. HYBRIDIZATION/SELECTION

15.1.9.1 First round hybridization
Driver amplicon digest (40 µg) was hybridized with 0.4 µg tester DNA ligate. DNAs were mixed in a total volume of 120 µl and extracted once with an equal volume of phenol/chloroform. To this was added 30 µl 10M ammonium acetate and the DNA precipitated by addition of ethanol with chilling at -70˚C for 10 min, followed by incubation at 37˚C for 2 min and centrifugation. The pellet was washed twice with 70% ethanol and vacuum dried. The pellet was resuspended in 4 µl of buffer (EE x 3) containing 30 mM EPPS buffer (Sigma) pH 8.0 and 3 mM EDTA, and vortexed for 2 min. The sample was briefly centrifuged and overlayed with 35 µl mineral oil . The DNA was denatured for 4 min at 98˚C in a heating block and 1µl of 5M NaCl carefully added, followed by 1 µl EE x 3 buffer and 1.5 µl NaCl solution. This was incubated at 67˚C for 20 hr in a water bath.

15.1.9.2. Selective PCR pre-amplification
After removing the oil, 8 µl of 5 mg/ml tRNA was added to the hybridized DNA and 390 µl TE buffer added. To fill-in the adaptor ends, eight tubes were prepared containing 20 µl 5 x PCR buffer, 16 µl dNTPs (4mM stock), 38 µl water and 20 µl of diluted hybridized DNA. Reactions were placed in the thermocycler at 72˚C for 5 minutes to dissociate the 12bp oligonucleotide adaptor. Ten units of AmpliTaq DNA polymerase were added and the reactions incubated for 5 min at 72˚C. After the addition of 4 µl *JBglII* 24-mer, self-reannealed tester amplicon molecules were amplified by performing 10 cycles of 95˚C for 1 min and 70˚C for 3 min, followed by a final extension of 72˚C for 10 min. All tubes were combined and 10 µg tRNA added, prior to two phenol/chloroform extractions. DNA was precipitated by isopropanol precipitation, as described above. The pellet was dissolved in 40µl TE buffer.

15.1.9.3. Mung bean nuclease treatment
An equal volume of 2 x mung bean nuclease buffer (NEB) was added to 20 µl difference product before treatment with 20 units of mung bean nuclease (NEB). This mixture was incubated at 30˚C for 30 minutes.

Mung bean nuclease is a single-strand specific DNA endonuclease that will degrade only single-stranded extensions from the ends of DNA molecules, leaving only blunt ends. This treatment will remove adaptors from single-stranded DNA, leaving only double-stranded molecules to amplify.

The mung bean nuclease was inactivated by a 5 min incubation at 98˚C following addition of 160 µl of 50mM Tris-Cl pH 8.9.

15.1.9.4. Selective amplification
Three 100 µl PCR tubes of standard PCR mixture were prepared containing 5 µl J*Bgl*II 24-mer and 20 µl nuclease-treated difference product. After preheating at 72˚C for 5 min, 10 units of AmpliTaq DNA polymerase was added. Amplification was as above, except only 20 cycles were performed. An additional 10 units of AmpliTaq polymerase were added and amplification continued for three more cycles, followed by a 72˚C hold for 10 min. A 10 µl aliquot of this PCR product was run on a 2% TBE agarose gel and quantitated by fluorometer.

15.1.9.5. Optimization of selective amplification of difference product
The optimal number of PCR cycles needed to amplify a high yield of double-stranded difference product 1 (DP1) was determined according to Lisitsyn (pers. comm.) and Baldocchi et al. (1996). After 14 cycles, the reaction was sampled and more AmpliTaq DNA polymerase was added. PCR was continued and the reactions sampled after every subsequent two cycles up to 24 cycles.

15.1.9.6. Change of adaptors for subsequent hybridization/selection
The reactions in the three PCR tubes were pooled and twice phenol/chloroform extracted followed by isopropanol precipitation of the DNA. The DNA pellet was suspended in 80 µl TE buffer and quantitated by fluorometer. Five micrograms of first difference product (DP1) was digested with *Dpn*II to remove the adaptors. Digestion was performed as a 100 µl reaction using a total of 50 units of *Dpn*II enzyme. To aid in precipitation, 10 µg tRNA was added prior to performing two phenol/chloroform extractions and an ethanol precipitation of the DNA. The pellet was dissolved in 100 µl TE buffer, before quantitation by fluorometer. The DNA concentration was adjusted to 20 ng/µl. Primer pair set 3 (N*Bgl*II) was ligated to 100 ng of digested difference product 1 in a 30 µl volume. To the difference product ligate, 50 µl TE containing tRNA (20 µg/ml) was added.

15.1.9.7. Subsequent hybridization/selection steps
For the second hybridization, 50 ng of first round difference product ligated to primer set 3 (N*Bgl*II), as described directly above, was hybridized with 40 µg driver amplicon digest. Subsequent hybridization and selective amplification steps were as described above. The adaptor on the second round difference product was changed to primer pair set 2 (J*Bgl*II, section 15.1.2). The concentration of DNA ligate was then adjusted to 2.5 pg/µl in TE buffer containing tRNA (20 µg/ml).

For the third hybridization, 100 pg of difference product 2 (DP2) ligated to primer pair set 2 (J*Bgl*II) was hybridized with 40µg of driver amplicon DNA digest.

15.1.10. CHARACTERISATION OF DIFFERENCE PRODUCTS

15.1.10.1. Southern Hybridization to detect λ fragments
A southern hybridization was performed to determine whether any of the positive control difference products were indeed lambda DNA specific and whether there was any cross-contamination of experimental difference products with lambda DNA. Probe was prepared by labeling 1 µg lambda DNA, previously digested with *Dpn*II, with the Boehringer DIG-High Primer kit as in section 5.3.1.

One microgram of various experimental and positive control difference products, along with lambda DNA were run on a 2% TAE gel at ~50V; pUC19 DNA was included as a negative control. The southern hybridization was performed according to the DIG user

guide. The gel was denatured for 2 x 15 minutes at room temperature, rinsed in distilled water and then neutralised for 2 x 15 minutes in neutralization solution. DNA was blotted overnight to Hybond N+ membrane in a 20 x SSC solution. DNA was fixed to the membrane by UV irradiation for 1.5 minutes on each side. The membrane was prehybridised according to section 5.1.4. Probe was added at 5 ng/ml by addition of ~2 µl denatured probe to 20ml hybridization solution. Hybridization, washes and detection were as described in 5.1.4 and 5.1.5. A 30 minute film exposure was used.

15.1.10.2. Cloning of experimental difference products: DP1; DP2 and DP3
Each of the difference products were blunt-end cloned into the Promega pGEM®-T vector system. The pGEM-T vector was commercially prepared by cutting pGEM-5Zf(+) with the restriction enzyme *Eco*RV and adding a 3' terminal thymidine to both ends. The 3'-T overhangs improve the efficiency of ligation of a PCR product with 3'A ends (added non-specifically by Taq). Ligation reactions were prepared in a 1:1 molar ratio of insert:vector, assuming an average insert size of 225 bp, 11.25 ng of each of the uncut experimental difference products was ligated to vector. The ligates were transformed into Stratagene JM109 cells according to section 5.1.1.

Plasmid DNA was isolated from each of the resulting recombinant clones by the alkaline lysis method (Sambrook et al, 1989). *Sac*II/*Nde*I double digests were performed on each of the plasmid DNAs to reveal the presence and approximate size of the cloned inserts. The clones were sequenced using ABI Prism® Dye Terminator cycle sequencing or using ABI Prism® BigDye® Terminator Cycle Sequencing, according to kit instructions.

15.1.10.3. Dot blot hybridization
Hybridizations were performed to determine whether any of the RDA difference product clones were indeed sex-specific. Probe DNA was prepared by labeling 500 ng of clone plasmid DNA with Boehringer DIG-High prime kit according to 5.3.1. The dot blot procedure was similar to that outlined in 5.3.1.

The templates tested were individual male gDNA; individual female gDNA; female pooled gDNA; DP1; DP2; DP3; tester (male) amplicon and driver (female) amplicon. An amount of 100ng of each of these DNAs was used as template. It was necessary to remove any remaining linkers from the various difference products and amplicons, and this was achieved with *DpnII* digests. Plasmid DNA was included as positive control (10 ng) and lambda and human gDNA as negative controls (100 ng).

Probe was added at 7.5µl per 50 ml hybridization solution (3.75 ng/ml) for initial hybridization experiments, but reduced subsequently to 2.5µl per 50 ml (i.e. 1.25 ng/ml) for hybridization of clone 3-5.

15.2. ALTERNATIVE DNA EXTRACTION PROTOCOLS

15.2.1. SDS LYSIS GENOMIC DNA EXTRACTION FROM BLOOD (recommended by Richard Powell, University College, Galway, Ireland)

Genomic DNA was extracted from 100µl of fresh blood. Lysis was performed by diluting the blood up to 5 ml with lysis buffer (10mM Tris.Cl pH7.5, 2mM $MgCl_2$, 10mM EDTA, 400 mM NaCl, 1% SDS) containing 0.2 mg/ml proteinase K, and then incubating at 37˚C for 8 hr. The solution was then phenol extracted using a 65˚C incubation for 5 min, with gentle mixing. Phases were separated by centrifuging at 5 000 rpm for 10 min. A chloroform extraction was then performed. The DNA was precipitated with 2 volumes of ethanol, hooked out with a bent pasteur pipette, dried and resuspended in 500µl TE (pH8.0).

15.2.2. MODIFIED VERSION OF TAGGART ET AL, 1992 (John Taggart, Stirling University, Scotland)

This method uses muscle tissue and makes use of a different detergent, sodium lauroylsarcosine.

Materials required:

Soln. 1. 0.2M EDTA pH 8.0; 0.5% sodium lauroylsarcosine. (SIGMA L-5125)
Soln. 2. 20 mg/ml proteinase K (or pronase) (see Sambrook et al. 1989 B16)
Soln. 3. RNAse (DNAse free) 20 mg/ml (see Sambrook et al. 1989 B17)
Soln. 4. Phenol pH 8.0 equilibrated
Soln. 5. Chloroform:isoamyl alcohol (24:1)
Soln. 6. 92% ethanol (room temp)
Soln. 7. 70% ethanol (room temp)
Soln. 8. TE pH 8.0 (10mM Tris, 1mM EDTA) -Sterile

[microfuge tubes with tight fitting caps (e.g. Treff), Sterile pipette tips, and waterbath / incubator @ 55˚C / 37˚C]

DAY 1

1. To each microfuge tube add 375 µl Soln. 1 and 12 µl proteinase K (or 25 µl pronase)
2. Add tissue to each tube. (e.g. 70 mg muscle, 10 mg liver)
3. Mix briefly and incubate overnight at 55˚C (for proteinase K) or 37˚C (for pronase)

DAY 2

4. Add 10 µl RNAse to each tube. Shake very vigorously and incubate for 60 min @ 37˚C.
5. Add 400 µl phenol to each tube. Shake very vigorously for 10 sec - then gentle overend turning for 15-20 minutes.
6. Add 400 µl chloroform : isoamyl alcohol to each tube. Shake very vigorously for 10 sec - then gentle overend turning for 15-20 minutes.
7. Centrifuge the tubes (e.g. 10,000g) for 5 minutes.
8. Carefully remove 300 µl of top aqueous layer, (using a micropipette with wide bore tips), to a new microfuge tube. NB Do not disturb material at the interface. Use only debris free aqueous solution.
9. Add 900 µl (3 vol.) 92% ethanol. Mix by vigorous inversion of tubes 5 - 6 times. The DNA should precipitate out. Leave 2-3 mins. Carefully decant off most ethanol.
10. Add 1 ml 70% ethanol. Mix by gentle overend turning for at least 30 mins (preferably overnight at room temperature). Carefully decant off most ethanol.
11. Carefully remove remaining ethanol (50-100 µl) using a micropipette.
12. Allow the DNA to partially dry at room temp. (5-10 mins). DO NOT OVER DRY. Resuspend pellet in TE pH8.0. A final conc. of 500-1000 µg/ml is desirable. (For 70 mg muscle tissue try 50 µl TE initially.) Store at 4˚C. Total dissolution may take 1-2 days. For long term storage freeze @ -20˚C

Protocol Notes:
1. Though it has not been stringently assayed, proteinase K treated samples appear to give better yield and quality of DNA and thus is currently favoured over pronase digestion.
2. Digestion can be assisted by constant mixing of the sample; e.g. using a rotisserie style hybridization oven for overnight incubation.
3. Vigorous shaking after addition of RNAase, phenol, and chloroform helps to break up tissue aggregates and ensures efficient extraction. It does not adversely shear the genomic DNA.
4. Multiple shaking, and mixing and final DNA precipitation can be conveniently carried out using tight fitting microfuge tube racks.
5. Use of three volumes 92% ethanol (at room temperature) instead of the more usual two volumes of 99% ethanol helps prevent or reduce EDTA coprecipitation. We have routinely found that DNA precipitated with 92% ethanol dissolves quicker and more thoroughly and restricts more effectively.

15.2.3. SALT PRECIPITATION OF GENOMIC DNA FROM BLOOD (Lars-Erik Holm, Danish Institute of Agricultural Science, Denmark)

Extraction of DNA from whole blood.
1. 10 ml of whole blood (EDTA or citrate stabilized) (for fish 50-100 µl) is mixed with 40 ml of cold buffer A in a sterile Falcon tube.
2. The tube is turned around a couple of times and is placed at 4˚C for 30 minutes.
3. Centrifuge at 3 000 rpm (1500 g) at 4˚C for 10 minutes.
4. Discard the supernatant.
5. The pellet is washed with 15 ml of buffer B. With the use of a plastpipette (we use a Sarstedt transferpipette for this) the pellet is tumbled up and down a couple of times.
6. Centrifuge at 3000 rpm (1500 g) at 4˚C for 10 minutes.
7. Discard the supernatant. If the pellet is very big or very red, it can be an advantage to repeat this washing step.
8. The pellet is dissolved in 4.5 ml of buffer B until the pellet can be sucked up by a pasteurpipette.
9. Apply 500 µl of buffer C carefully at the side of the Falcon tube. If buffer C with Pronase is used, incubate at room temperature in a slowly rotating device over the night. If buffer C with Proteinase K is used, incubate at 50˚C in a gently shaking waterbath over night.
10. Apply 1.35 ml of 6 M NaCl. Shake the tube violently for some seconds (15 sec is sufficient).
11. The proteins are centrifuged down at 3000 rpm at 4˚C for 15 minutes without the use of the brake on the centrifuge.
12. The supernatant is carefully transferred to a new Falcon tube by the use of a pipette (for this we also use the Sarstedt transferpipette, but a pasteurpipette is also usable).
13. DNA is precipitated by the application of 2 volumes of absolute ethanol.
14. The DNA threads are caught and transferred to an empty Eppendorf tube. Wash with 70% ethanol.
15. Redissolve in 1xTE buffer (ex. 500 µl).
16. Measure the DNA-concentration and purity.

Solutions:

Buffer A:
0.32 M Sucrose (109.5 g/l), 1 mM Tris-HCl pH 7.5 (1 ml of 1 M Tris-HCl pH 7.5), 5mM $MgCl_2$ (5 ml of 1 M $MgCl_2$), 1% Triton X-100 (10.0 g/l),with dd H_2O until 1000 ml; Autoclave

Buffer B:
400 mM NaCl (100 ml of 4 M NaCl), 2 mM EDTA pH 8.0 (4 ml of 0.5 M EDTA pH 8.0), 10 mM Tris-HCl pH 8.0 (10 ml of Tris-HCl pH 8.0), plus dd H_2O until 1000 ml; Autoclave

Buffer C (with Pronase: mix just before use):
5% SDS (0.5 ml per ml buffer of 10% SDS), 2 mg/ml Pronase (0.1 ml per ml buffer of 20 mg/ml Pronase), sterile H_2O (0.4 ml per ml buffer)

Buffer C (with Proteinase K: mix just before use):
5% SDS (0.5 ml per ml buffer of 10% SDS), 2 mg/ml Proteinase K (0.2 ml per ml buffer of 10 mg/ml Proteinase K), sterile H_2O (0.3 ml per ml buffer)

Pronase 20 mg/ml:
1 g Pronase (1 ampoule from Boehringer-Mannheim), 50 ml sterile distilled H_2O. Incubate at 37 C for 2 hours. Store in 1 ml portions (use Eppendorf tubes) at -20 C.

Proteinase K 10 mg/ml:
50 mM Tris-HCl pH 8.0 (5 ml of 1 M Tris-HCl pH 8.0), 1 mM $CaCl_2$ (0.015 g $CaCl_2$, 2 H_2O), distilled H_2O until 100 ml. Sterile filtrate or autoclave this solution. Apply 1 g Proteinase K. Store in 1 ml portions (use Eppendorf tubes) at -20 C.

6 M NaCl (some salt remains undissolved):
35 g NaCl, dd H_2O until 100 ml

15.3 VISIT REPORT TO MELBOURNE ON RDA TECHNOLOGY

Report by Anne Reilly on visit to the Royal Women's Hospital, Carlton and University of Melbourne, Parkville for discussions on RDA technique
March 20, 1997.

15.3.1. Dr Eric Moses,Senior Research Fellow, Department of Perinatal Medicine, Royal Women's Hospital, Victoria

Eric provided a brief outline of his group's work in trying to find genes associated with the condition pre-eclampsia which is responsible for high blood pressure in pregnant women. The condition is believed to be genetic and RDA (Representational Difference Analysis) has been applied to cDNA (complimentary DNA) generated from mRNA (messenger RNA) isolated from women suffering from the condition (tester) and unaffected women (driver). It has been found that there is some variation in products when the experiment has been repeated using different tester individuals, but there are common products which are more abundant than in driver individuals. Difference products were excised and labelled as probes to be characterised by northern and southern blots. Their RDA work has isolated a difference product that appears to be more prevalent in patients suffering from the condition.

As a control to demonstrate effectiveness of the subtractive/enrichment procedure, the driver cDNA digest was spiked with lambda DNA digest material. In their case, due to working with cDNA which is far less complex than gDNA (genomic DNA), 4-base cutter restriction enzymes were used, which produced a smear in lambda, with a couple of major bands. Two of these major bands were excised and used to spike the driver sample, adding only femto-mole (10^{-15}) quantity of this material. These two major bands comprised all the material visible as difference product on the agarose gel. However, when the difference product was probed with total lambda digest material, additional bands were obvious, corresponding to other lambda digest bands. This indicated that the method is very sensitive, i.e. the very small amount of contaminating material in the gel excision was enriched through the successive selective amplification steps.

I briefly presented the aim of our research and explained our current progress with the RDA technique. I went through the first RDA experiment and showed results so far.

Outcomes:

Eric pointed out that my gels indicated only partial cutting of the genomic DNA, as evidenced by fairly large quantities of intact genomic DNA. I discussed our protocol for genomic extractions which includes a CTAB step. Eric recommended using a standard genomic extraction procedure that uses SDS, such as that in Sambrook, J., Fritsch, E.F. and Maniatis, T., 1989. "Molecular Cloning: A Laboratory Manual" 2nd edn. (Cold Spring Harbor Laboratory: New York) to clean up impurities that may be inhibiting restriction. He recommended using RNAase early in the protocol, rather than a final suspension in buffer containing RNAase. He also advised to extend proteinase K treatment from 1 hr up to ~ 4 hrs (or overnight), also to increase the number of phenol, phenol/chloroform, chloroform extractions and include ammonium acetate for precipitation to maximise yield.

Qiagen genomic clean-up kits were also recommended.

It was also suggested that I perform restriction digests for at least 4-6 hrs or overnight. The use of several 6 base-cutter enzymes was recommended since some may perform better than others.

If DNA is not restricted to completion then there will be little material in the amplifiable size range, i.e. ~150 bp up to ~1500 bp. This was evident in the amplicons, where there was poor amplification, apart from the amplification of some repetitive elements. The representation amplicon and the DP1 amplicon should consist of heavy smears in the amplifiable region, with perhaps a couple of heavier repetitive element bands visible. It was Eric's opinion that the quality of the initial digest was the main reason for lack of success in the first round experiment.

Digested driver DNA (i.e. female salmon) spiked with digested lambda DNA at a range recommended in the original protocol (i.e. 16 pg to 1 ug driver) was recommended as the positive control.

Other points of interest:

Once difference product is isolated, a large quantity can be produced by PCR amplification, instead of cloning. Difference product may still contain a complex mixture of products i.e. may appear as a single band but may contain other minor contaminating products which co-run with major band. If the band is excised and PCR amplified, these contaminations will also be amplified. Therefore to ensure only the major product is amplified, limiting dilution PCR should be performed to eliminate products due to the contaminant.

Gibco Taq has proven to be as efficient an enzyme in microsatellite work as Amplitaq Gold.

Use NuSieve agarose for good resolution.

Glycogen (Boehringer) can be used instead of tRNA as carrier in precipitations.

Prepare amplicons of driver separately. This is the approach they used when pooling 3 individuals for the driver.

Crude grade primers are satisfactory for both PCR amplification and as adaptors in ligation reactions (far cheaper than HPLC grade @ $2.95 for 200 nm synthesis instead of $5.95). Alternative supplier is DNAgency or Ransom - Hill Biosciences

Use barrier tips to guard against aerosols

It was mentioned that they previously had trouble due to mung bean nuclease buffer inhibiting subsequent amplifications, but have since rectified the problems. Appears it was one particular bad batch supplied by NEB. Subsequent batches have been OK.

For hybridization, be sure to overlay reaction with mineral oil to prevent evaporation.

15.3.2. Drs Jill Maddox (Group Leader) and Peter Roche, Centre for Animal Biotechnology, School of Veterinary Science, University of Melbourne, Parkville

Peter Roche explained the aims and methods of their research. This group is attempting to identify major genes that are important to parasite resistance in sheep. This would provide an effective means to select for genetically resistant sheep. They have been able to isolate a number of polymorphic ovine genetic markers by both RDA and GDRDA (genetically directed RDA). The RDA experiment used Romney "resistant" ewes as tester and Merino "susceptible" rams as driver. Many clones were isolated, with some 75% being Romney specific. An additional experiment using only Romney sheep which are either highly or poorly resistant to the parasite has produced several clones specific to the tester.

Clones are tested for specificity by dot blot analysis of tester and driver amplicon DNA, with subsequent analysis by southern blot.
These markers are all being further characterised to determine whether they are linked to parasite resistance.

Recommendations:

Jill Maddox and Peter Roche concurred that genomic digestion looked to be the most likely problem with my work. Peter Roche suggested getting around the problem by digesting more DNA than was required, say 10 ug, so that even if digestion is not complete, then there will be more total DNA in the appropriate small size range. Peter also suggested overnight enzyme digestion. He recommended then diluting this DNA to 1 ug/ul and using as a stock.

Peter always relies on ethidium stained gels to estimate DNA concentrations, comparing to various standards of different dilutions. (Especially for digests, since UV or fluorometer readings will not reflect completeness of digest). Jill suggested that DNA teststrips might be worth trying for quick quantitation of gDNAs.

Peter pooled his driver DNAs following DNA digests. That is, he digested equal amounts of each driver individually, estimated on a gel and then pooled.

Peter used 2 ug for ligation instead of only 1 ug as recommended.

This group modified the original method slightly by removing the adaptors from the tester using a 10% PEG/0.6 M NaCl fractionation instead of electrophoresis through agarose. They used this approach due to problems with DNAase contamination in gel cut-outs. (It was also mentioned that PEG fractionation for clean-up increases the efficiency of cloning and PCR amplification).

A basic protocol for the PEG fractionation is in Sambrook. An equal volume of 10% PEG/0.6 M NaCl is added to the DNA solution. The DNA is recovered by centrifugation. The supernatant, containing the adaptors is discarded, and the pellet resuspended in TE buffer. The solution is extracted once with phenol, once with phenol/chloroform and once with chloroform. The DNA is ethanol precipitated and resuspended in TE buffer.

Other comments:

I also showed my second experiment results where I had variation in the profile between replicates for HindIII digest. It was suggested that this result was due to problems with uneven heating within the PCR. Jill suggested running some RAPDs in replicate to test the PCR thermocycling profile. RAPDs will demonstrate any heating problems in the PCR since they are so sensitive to temperature changes.

They commented though that overall the amplicons of this second experiment with BglII looked more encouraging than the first, due most likely to better BglII digests.

15.4 SEQUENCE DATA

15.4.1 MICROSATELLITES

The microsatellite motif as per Table 5.1 is underlined

LOCUS *cmrSs1.2* 1093 bp E5A4291D checksum.

```
   1  GnTCTTCTTG GCCTTCCTGT GACACTGGGT GCTGTAGATG TCCTGGAGnG
  51  CAGGCAGTGT GCCCCTGATG TTGTGTTGGG CTGACTACAC CCACATCTGG
 101  ATAGCCCTGC GGTTGCGAAC GGTGCAATTG CCATATCAGG CGGTGATACA
 151  GCCCGACAGG ATGCTCTCAA TGGTGCATCT GTAGAAGTTT GTGAGGGTCT
 201  TAAGGGCCTA nATGAATTTC TTCAGCCTCC TGCTGCGCCT TCTTCACCGC
 251  ACTGTGTGTG TGAAGGGACT ATTTCATCTC TTTAGTGATG TGCACACTGA
 301  GGAACTTTAA GCTTTTGACC CTCTCTACTG CGACCCTGTC GATGTGGATG
 351  AGGGGTCGTG CTTTCTCTGC TGTCTCCTGT AGTCCACAAT CAGAACCTTC
 401  GTTTTGTTGA CACTGAGGGA GAAGTTATTT TCCTGGCACC ACTCCACCAG
 451  GACGTGCGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG
 501  TGTGTGTGTG TGTGTGTGTG TGTGTTTGCT TAATGGAGCA TGTGTAGACT
 551  CTTCAGACAA TAAGCGTGCA GCTTGCCTGA GAATGGACTC CCATAGGCCC
 601  ATTAAGGCAT CATTCCTACA GAAAATGAAT CCGCCTTGAT CTTCAGGGAC
 651  CTGCACGGGC GGGGTGTAGT GGTGTATATA GATGATATTC TAGTATACTC
 701  CACTACCCGT GCTGAGCATG TGTCCCTTGT ACGAATGGTG CTTGGTCGAC
 751  TGTTGGAGCA TGACCTTTAC ATCAAGGCGG AGAAATGTCT GCTCTTTCAA
 801  CAGTCCGTCT CCTTCCTTGG GTACCGTTTG TCCGCGTCAG GGGTGGAGAT
 851  GGAGACTGAC CGCAATTCAG CCGTGCATAA TTGGCCTACT CCAACCACGG
 901  TAAAGGAGGT GCAGCGGTTT TTAGGGTTTG CCAACGACTA CCGGAGGTTT
 951  ATCCGGGGAT TTGGCCAGGT GGCGGCTCCC ATTACCTCCT TACTGAAGGG
1001  GGGACCAGTG CGATTGCGTT GGTCGGCTGA GGCGGACAGA GCTTTTGGTA
1051  ACCTGAAGGC TCTGTTTACC ACGGCTCCAG TGCGGCGCAn CCT
```

LOCUS *cmrSs1.7* 997 bp 34482D94 checksum.

```
   1  GATCATATCA NCTCCACCCT NCCTGACNCC CTAAACCCAC TCCAANTNNC
  51  TNNCCGCCCC AANAGGTCNA CAAACGAGNC AATCGCCATC ACACTGTNCA
 101  NTGTCCTANA CCCNTCTGGA CNNNAGGAAT CCCTACAGNG GGGANAACAN
 151  CTATTTGTCN CNCTGCCNAN TTTNCNGGTT TTCCTACTTN CAAANCATGT
 201  NGAGGTCTGT AATTTCTNTC CTNNGTNCAN TNCAACTGTG AGAGACCGAA
 251  TCTAAAACAG GTTCCGGAAA ATCGNNTCGT GTAATGNGGN NTTAATTCAT
 301  TTGAATTATA TATATATATA TATATATATA TATATATATA TATATATATA
 351  TATATATATA TATATTTNTN TTTTAAAGCC AGCCCGGTGA ATAAACAGTT
 401  AGGTGAGGGA GAGGGCATTA CTAAGGGTGG AGACGGTAAA TANTACTGCA
 451  TANAATCTGT AATTCATCAG TNATACCATT AATCATTATA TACAGTGCAT
 501  TCAGACCCCT TGACCTTATC TACATTTTGT TACTTTACGG CCTTATTCTT
 551  AAATGGATTA AAACTTTTTT CTCCTCAATC TCTGATCTGA GANAACTTCC
 601  TCAATCCTGA CCTTGGGCAA CAAGTTCCCT TCCGTATGTC TGTGAAAAAT
 651  CCCNNAAAAA AGGCTCACTC CTACCNTNGA AGAAAAACCC TGGGNANTGC
 701  ACCNAAGGGT TCCCAAGGCA TNCCCGTNAA ANCNTTCTGA AAATTTTTAA
 751  ANNTGCATTG AACGGGCNCT TGGGGTTNTT CNTTNAATTG AAATTAACAN
 801  NTCCNCCCAA TGTTNAAAAA AAAAAANAAN CCCANTTTTC CANTNATTNC
 851  AAATCCTTTG GGTTCCCCAT NCCAANTCTT CCCCTTCTTA AAAAAGGGCC
 901  TTTTTTTTTT AAAAAAGNAT TNCTAAAGGT TTTNTTCCCT TCCCCCTCAA
 951  ATTAAANGGN GGGGGCCTTT TNATNTGGNA NCCNTTNCCN NCNNTTN
```

LOCUS *cmrSs1.10* 786 bp AEEE9C9D checksum.

```
   1  aaccaccaaa acccccccgg ggtttaatgc nccngtaaca gggggngtcc
  51  attngccatt cagGnTgcgc caacttgttg ggaagggcga tcggtgcggg
 101  ccttttttcgn tattacccca nctggcgaaa gggggatgtg ntgcnaggcg
 151  attaagttgg gtaacgccca gggttttccc agtcacgacg ttgtaaaacg
 201  acgcccagtg aattgtaata cgactcacta tagggcgaat tcgagctcgg
 251  tacccgggga tCAAACTGTT GCCTTGCTTA ATTACACTGT ATAGACTTAT
 301  GTTCCAGAGC TAGGTGCCAT AGAGAGCGAG AGAGTGTGCG CTTGTGCTGG
 351  TGTGTTAGAG TGTGTGCAGT CACAAGGGAC TACTCCAGGC CTCTGTCATA
 401  TnTATGGCCC CTCTCAGGAT GAATGGTCCT TTGAGTATTG GCCATGAGGC
 451  TGGGGTAATT GCCATTGTAC ACACACACAC ACACACACAC ACACACACCA
 501  CACATTACCT AACCATTAAT AGTCTTAATG GCGGGGAGGC GGACAGTTAA
 551  AAGGCGCTGG ATAACTACCT GGCTGGCTGA CCACAGTGTG TGTGTGTGTG
 601  TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGCATA
 651  kAttttctct acaggacagt aagtggtctg ctgttgcaat gctttctgaa
 701  ggaacacttt ttatgctgtt ataaggcatg tttacctttc tctttctgtt
 751  tcattttgtt tgatcctcta gagtcgacct gcaggc
```

LOCUS *cmrSs1.14* 671 bp E3A6DFF2 checksum.

```
   1  GCCTTCCTGT TNATCAGNNA CGTTCCTCGG GCAGAGGNAN TGCGTCGGGG
  51  TAGGTGTGGT GGTGATCACA GAGNAGCTTT NACCNATGTC NANTACGGTA
 101  GATTGNAGCA TTNNATTNTA TCGATTTATG ACACTATTNT GGTCATCNAN
 151  AGGTTTATTT AGTNTTNTAG GGCAACATAT NATGACAGAA GTGNAGCTGC
 201  GTGTTTTCTN ACTGTAGACA AGGTGCCGGG NNANNANATG GCCTACCNAN
 251  ATGTCAGANA CGACNAGANA GAATCTGCTT GGAAAAATCG TTCCGCCACA
 301  CAGGAGGTTA ATTGTGGAGG ACGGGCTTGT GGGAATGGCT GGAGCGGNAT
 351  AGTATCAAAT ACATCAAACA CGTGGTTTTT CCNATGCCAT TATTATGAGC
 401  CGTCCTCCCC TCAGCAGCCA CCACTGTTCC GTCGTCTCTG ACTGTAGCCT
 451  ACGGCGCATT TTCTATATTA GCTGGTTAGG GTCGGACGCT GGCCTCGGAT
 501  TTTCACATAT AGTCAGGCTG TTACGGATGG ATTGGTTATT AACACCACTA
 551  GCACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACGGTT
 601  CTTACCTGGC CCTGCCAGTC GGTGCCATTG ACCAGGATGA GACTCTCAGG
 651  CAGAGCCGCG TCAGACAGAA G
```

LOCUS *cmrSs1.15* 874 bp 3400261A checksum.

```
   1  TAACnCCnTA AAACGAAnCG GGnTACCCGC CGTCCCnCTT TTTTAAATAA
  51  AAAGGnACCC CCAACTTAAA TnAAGGGGGG CGGAAAnTTT CCCGAAAAAG
 101  CCTnCCGGGG TnAACCCCCG GGGGGGATAT CCCCGGGGCC CTAAAAAAAA
 151  GCCAATnGAA GGGGGTTTGG CGGGCAATGT TCTGGTCCTC CTCCTGGGAA
 201  AGTTGAAACT GGGGCCACAA GGGTCCCGGC AAGGGCCAAG CAATCCACTC
 251  CTCAATTTTG GCCGTTCCAA TTTTGGAATA AGAATTAACA CAGAATTAAT
 301  GGCAAGCTGA ACCTGAAGGC TGCACTGTCA ATGGGGAAGC TAAATTGGCC
 351  TACTTCcGTG TgTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG
 401  TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG
 451  TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG
 501  TGTGTGAGAG AGAGAGAGAG AGAGGCACCA ACTGCCTGGT ATGTGCTCAT
 551  TACTGGAGAT TATCTCACAA ACTGTCAGTT AATAAACTGT CTGTTCAAAA
 601  ATAATGAAAC CAAGTCTATT ATCCTCTTAA GGGTAGGCCT ATAGCTTGCA
 651  TTGCAGAATA AAGTCAGATG CTTGTTCATG AAAAATGGGC ACATTCGGTT
 701  CTGCTTAGAG GACAACCTCA TCCAAAGACT AGCCTATGAG CCCTGCAATG
 751  TAATGATTCT ACTCATCATG AACTATGACA AACTAAGAAT ACACAATGCA
 801  TGTATCACAT ACCCCAGTGA CAAGTCAACT ATTTGTAGAT ATCATCAACA
 851  ATGnCCCnCT TCACAACTAT nCTT
```

```
LOCUS      cmrSs1.16              585 bp          9CF9733C checksum.

       1  GGNGGATTGG aGTTcGGTAC CcGGGGaTca TcaTTaGTaG GTTcaTGACA
      51  TCCTTCNATT CacTaTGGGG aaGtACacTG CAGTcGAGCC aGTCaCCaTG
     101  GGATTcATAT TCCCTtTaTG aTTaTTaTTG NACaTGATTc TTCaGcTcTa
     151  ATaaACATcT GcTTTGACaA CGcTGaGCAG CCATTCCTCA CCATAGCaaC
     201  CaATCATGCA TGCCTTTcTc TTcTATTTCA TGTTTGTGAT AGCTTACCCC
     251  TagaATGTCA TATTGaCCGT TTTTGCCaAT GTTTTGCAGT CCGCaAGGTG
     301  aaATGGTCAG TTGTGTGCAT ACTTGTGTAC AGTTAAGGGG TCAGGGGTTA
     351  TGGTaACACA CATACATTGC TGAGAGTATG ACTTTGGACC ATGATGCAGT
     401  TGGGCGCTTG GAGaaaAAGC AGAGGAGGAG GaATACACAA CTGTCaATTT
     451  GTTGTCCTAT AGCTGGAGGA GCATCTTGAT GGGTAGAAGA GTTAAACCAC
     501  CCACACACAC ACACACACAC ACACACACAC ACACACACCT TAGGACCGCA
     551  AAGGACCTTA ATAACGTGTC AATACAAAAG CATCA

LOCUS      cmrSs1.18              821 bp      A1020BF2 checksum.

       1  Gggggggttt ccattttgcc catttcnggg nnngngnnaa ntnntggggg
      51  aggggcgatt cggtgcgggc ctttttttnt natnannccc agcttggcgg
     101  aaaggggggat gtgnttgcnn ggcgatttaa gttgggtaac gccaggtttt
     151  cccattcang acgttgtaaa cgacggccag tgnattgtna tacgactcac
     201  tatagggnga attcgagctc ggtacccggg gATCTGCTAC TTCTCTTTGC
     251  ACATCTGCAT TCACCAGCTT AATGACATnC ACCCATGCTG TGCCTTTTTA
     301  CACAGACGTA CCAGCTTATG TGATTGTGAA aCTTCTAGTG TATAGCACAG
     351  CCAGGTCAAA CATCAGTCCC CAATCTGAAA ACCTGGCCAC CAGTGTTAGG
     401  AAGAGAGCAC AGGCGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT
     451  GTGTGTGTGT GTGTGTGTgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt
     501  gagagagact cacttggtct ggatgaggat gtcagcattg gcggagttca
     551  gcatgaattt acagatgagc tcctgtgtca ctggaatggc cgtcttattg
     601  gacacacaca ggtctgacag atagtccagg aacctgattg gacagaagag
     651  gacaaccaaa cgtattggac taaatctgct caaagacatt gaacccagta
     701  cagtggctta ccaaacaata cctaataaat ataggcaaca caaatatagt
     751  gtaatataca taggtttttt gtttgtgTGG CACATTGTTC TTTATAGCCC
     801  CAGTGTGCAT ATCTATTTGT G

LOCUS      cmrSs1.22               550 bp      62B1C63D checksum.

       1  GATCAATAAA AAAAGATAGT TATGTTGTCT AACTTTCTGC ACCGCGAGGA
      51  CTGTnTTACG AACACTGTGT GAGTGAGTGG TTACCTTTTG GGCGGGCCCT
     101  TGACAAGGTT GGTGACTTCT CTGTAnTCCT CGGCCAGCTG GTTCTTTAGC
     151  CGCAGCACCC TGCATCTCTG GCCCACACAT TCCTCACAGA GAGTTGACCC
     201  TGGACACAGG GTCACACACA CACACACACA CACACACACA CACACACACA
     251  CACACGAGGT CAGAGGTCAC AGAGAGGAGC TGAGACTATC CATGCATACA
     301  AAAAAAAAGA GACTATCCAT GCGTACACAT TTTTTTTTTG TTAAATGCAT
     351  GAAATGAAAT GTATGCATTC ACTACTGTAA GTCGCTCTGG ATAAGAGCGT
     401  CTGCTAAATG ACTAAAATGT AnATATCAGC AATGTCTTTT ACCATCTAGT
     451  CTGGGTCCTC CGCCATATCT GTCAAAGAGA AGnTnTGCTG CTTTnnCGGA
     501  GAnCCnCTAG AGTCGACCTG CAGGCATnCA AGCTTGAGTA TTTCTTATAG

LOCUS      cmrSs1.28               295 bp          CHECK: 3203

       1  NNNNCNCACA CACACACACA CACACACACA CACACACACA CACACACACA
      51  CACACACACA CGTGTCCGTG TCCTGTCCCT TATAAGCAGG TAAAGGCCAT
     101  GCAATCTGAG TTGAAACCTG AGTGAGGTGC ACATGTACAG TACATAAACA
     151  GATACATGCA CCATATCTTT ACGTGGTGGA CACTTTATAC GGTCACATGA
     201  TATTGTTGTG ATATGTCACA AAATCATATT ACGACCACAC GTATCAATAT
     251  TCATTTTTAT ATACCAATAT TTTGCTAAGN CTTGCTAACT GAATG
```

LOCUS *cmrSs1.30* 740 bp 2830E5BB checksum.

```
   1  TCCAAGGGGA TTAAGTTGGG TAACGCCGGG TTTTCCCCAT TCACGACGTT
  51  GTAAAACGAC GnCCAGnGAT TGTAATAnGA CTCACTATAG GGGGAATTCG
 101  AGTTCGGnAC CCGGGGATCA AAATGTAnTC GGGGGTGGTC CAAAATGnAC
 151  TCGGGGGTGG TCCcAAATGT ACTCGGGGGG TGGTCCAAAA TAGGGGnaGG
 201  gTTGTCAAAC GTTTAAAAAT ATATATATTT TTACAAATAT TTTTTATTGG
 251  GCACAGGGGA aGGgTAGTAA GTTTTCCCCT TGGTTTACAT TTGCTCATCT
 301  CCTTGTATTT TCTCTATAAA CAATGGGTTG ACTTACTTTT GATATTACCC
 351  TAATAAAGTT CAGAAACGTT TGTGAAGGTT TGCTATGGTG TGGCTGTTAT
 401  TAAGCTTTAG CTACTGCAGG TCAACTTGAG GTGAGGAAGA CTATATGCAT
 451  GTCGGTGTCA TAGCATGCCA CGGCATATCT CTATCTCTCT GAGTTAGGCC
 501  TAAGCTTCTC TTACTGTTAT ATAGGCTATA TACAGCGTGT GTGTGTGTGT
 551  GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT
 601  GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTTTGTAG TTTTGTGTAT
 651  GTTGACTTGT TGAGAACTAA ATCCCCTAAA GGCATCATTC TTCTGCTAAC
 701  CCGAGTCTAC CCTACAAGAC AACCTGAGCC TTAACTGCAA
```

LOCUS *cmrSs1.35* 1210 bp A2BEB4FD checksum.

```
   1  GATCGGTGGC CCCTCATGAT GAATTCAGAA TTTTTGTGGC CCCCACTCCC
  51  ATCAAAGTTG CCTATCCCTG ATGTAAATAG TACGTGCAAG AAGTTAAAAC
 101  ACAAATATAG CACTGCAGGC GTAACACCGG TACACTGCTG TATTGCATTT
 151  ACTGTACGTA CTATTGCACA CAAGAAGTTT CCTCACAGGA ACATTTTTAT
 201  TTGAAGTGAA TCGAATCATT TCAATAATCA CAGTGCTAAA GAGTTGCAGT
 251  ACCTCTTTTC TCCAGGAGGG CGGAGACCCC CTGTGAGCCA GGCTTGACGG
 301  TGGATACATC CAGTGTTCCT TTGCCAATGT CCTGTAGCAG AGTCCTGGCT
 351  GTGTCAAAGC TGTCGTTCAT AGTGGTGGCT ATCACCCCCG TGGGACCCCT
 401  CTTCACCCAG CCACTGCAGT ACAGACCTGC CGACACACAC ACACACACAC
 451  ACACACACAC ACACACACAC ACACACACAC ACACACACAC GTTAAAAGGG
 501  GACAGTTCAA TCCAAAGGGG ATAGACgTTC TTCTCAgACc TCAACAGTGG
 551  TTTGAgGTCG ATCCTCTAAA NTCGAACTGC AGGCATGCAA GCTTGANTAT
 601  TCTATAGTGT CACCTAAATA GCTTGGCGTA ATCATGGTCA TAACTGTTTC
 651  CTGTGTNAAA TGGTTATCCG CTCACAATCC ACNCAACATN CNAACCGGAA
 701  ACTTNAANTT TTAAGCCTGG GGTGCCTAAT GAATGANCTA ACTCCCTTTA
 751  ATTGGGTTGG GCTCAATGCC CGCTTTCCAA TCGGGAAAAC TGTCTTGCCA
 801  ACTGCATTAA TAAATCGGCC AACCCCCGGG AAAAGGGGTT TGCTTTTTGG
 851  GCCCCTTNCN CTTCNCCCTN AATNAAANCC TNNNCCCGGT CTTCNGGTNG
 901  GGNAAAAGNT TTNNNCNCCC CCAAGGGGGT TTTNNGGTTT CCCNAAAANC
 951  GGGGGANANC CCGGGAAAAA AATTTTNACN AANGGGCCNN AAAGGNCNNA
1001  AACTTNAAAG GGNCGTTTTN GGGGTTTTNC CAAGGGCCCC CCCCCCAAAA
1051  AAAANNNAAA NTNNANCCCN TTTTNNNGGG GGGAACCCCC NGGGTNTTNN
1101  NAANNNNNTT TNCCCCNAAA AANCCCNNNN NTNNNTTTTN AANNNNNNNN
1151  NNAAAATNTT TTNCCTCCCC CNNAGGNGGG NNNNNNNNCN NNNNNTTTTN
1201  NNGNGNGGNN
```

LOCUS *cmrSs1.39* 1014 bp 9F28C486 checksum.

```
   1  GnTCAGnTTA CTACAGTATT AGAGTATAAA GTCAGnTTAC CACAGTAGAA
  51  ACAGGCCCAG GTCAGATTAA CACAGTAGTA AAGTTACAGT TCAGATTAAC
 101  ACAGTAGTAG AGTCCCAGGT CAGATTACCA CAGTAGTAGA GTCCCAGGTC
 151  AGATTAACAC AGAAGAAGAG TCCCACTTCA GATTACCACA GTAGTAGAGT
 201  CCCAGTTCAG ATTAACACAG TAGTAGAGCC CCAGGTCAGA TTACCACAGT
 251  AGTAGAGTCC CAGGTCAGAT TACCACAGTA GTACAGTCCC AAGTTAAATT
 301  ACCACAGTAG TACAGTCCCA GGTCAGATTA CTACAGCATT AGAGTCTCAG
 351  GTCAGAATAC CACAATTGTA CAGTCCCAGG TCAGATTAAC ACAGTAGTGG
 401  AGTCCCAGGT TCAGATAACC ACAGAAGTAG AGTCCCAGGT CAGATTACCA
 451  CAGTAGTAGA GTCCCAGATC ACATAAAGGA CTTAACAGAT GGGAGTTTAT
 501  ACAGGAATCT GTCTCTGTCT CTGTGTGTGT GTGTCTCTTT CTCTCTCTGT
 551  GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT
 601  GTGTGTGTGT GTGTGTGTGT CnGTGTGTGT GTGTGGGGGT TGGTTCTTCT
 651  ATCCTTGCTG GGCCTTAAAA TCCCCAAAAG TCCCAACAAG GACAGTAAAA
 701  TAAGGAACAT TCTCCCTCGT GGGGACATTG TCTGCACAAA CTTGAGACAT
 751  CTGTGGCATT GTGTTGTGTG ACAAAACTGC ACATTTTAGA GAGGCCTTTA
 801  TTGTCCCCAG CGCAAGGTGC ACTTGTGTAA TGAACATGCT GTTTAATCAG
 851  CTTCTGAATA TGACACACCT GTCAGGTGGA TGGAATATCT TGACAAAGGA
 901  GAAATTCTCA CTAACAGGTA TGTAAACGTA TTTGTGCACA ACATTTGAAG
 951  GAAATAAGAT GTTAGTGAGT AAGGAACATT TCTGGGATTT TTTATTTCAG
1001  CTCTTGAAAC ATTG
```

LOCUS *cmrSs3.4* 1002 bp B9BB47BF checksum.

```
   1  GATCAACTTC TCTACCACCT ACACACACAC ACACACACAC ACACACACAC
  51  ACACACACAC ACACACACAC ACAAAGGAGA GATGTTACAG AGACATGTGA
 101  GTGTGTATAC ACTGAGTATA CTAAACATTA GGACACCTTC CTAATATTGA
 151  GATGTCCTCA GAACAGCCTC AATTCATCAG GACATGGACT CTACAAGGTG
 201  TTGAAAGCGT TCCACAGGGA TGCTGGCCCA TGTTGACTCC AATGCTTCCC
 251  ACAGTTGTGT TAAGTTGGCT GGATGTCCTT TGGGTGGTGG ACCATTCTTG
 301  ATACACACAG GAAACTGTTG AGTGTGAAAA ACCCAGCAGC GTTGCAGTTC
 351  TTGACACAAA CCGGTGCGCC TGGCACCTAC TACCACACCC CGATCAAAGG
 401  CACTTAAATA TTTTTCCTTG CCCATTCACC CTCTGAATGG CACACAAACA
 451  GAATCCACGT CTCAATTGTC TTAAGGCTTA AAAATCCTTC TTTAACCTGT
 501  CTCCnCCCCT TCAnCTACAC TGACTGAAGT GGATTTAGCA AGTGACATCA
 551  ATAAGGGAnC CnCTAgAann CnACCTGCAG GCATGCAAGC TTGAnTATTn
 601  cTtATtAnTG nCaCggCCtA AaTAnCnTGG CGTtAngnAa TGGTCCATAA
 651  CTGTTTCCCT GTTGTnAAAA TTGGTTATCC CGCTCCCnAA TTCCCCCCCC
 701  AAACATTACC AAACCCCnGG AAAACCTTTA AAAATTTnTT AAAAACCCCT
 751  GGGGGGGTTG CCCTAAAATT GnAATTTGAA ACCCAAAnCT nCCCCAnTTT
 801  AAAnTTGGCC TTTTGCCCCC CCCACTGGnC CCGCCTTTTn CCCAnTTCCG
 851  GGGAAAAACC TGTTCCTTGn CCCACCTGGC ATTTAATnAA ATCCGGCCAA
 901  CCCCCCGGGG nAAAAGCGGn TTGCTTTTTT GGGGGCTCTT CCGCTTCCTC
 951  CGTTCAATnA AACCCTGCCC nCGGTCTTTC GGCTGCGGGA AACGTTTTTT
1001  TT
```

```
LOCUS      cmsSs3.5              1014 bp         6E8CDEEB checksum.

       1  GAAGAAAATG GAAAAAGGGG TGGGGGGAGC AAGGGGGGCC CtAACAAACC
      51  tGAATTTCAG TAAcAACCAG GTTTTTTTCA GGGGGGAATG GGNCCCAAAA
     101  TTTAAACCCA AAATTTATTN GGGGGGAGGN TTNGGGGAGG GTAACCCGGA
     151  AAAAATTGAA CCCAAAGTTC AAGGAATTTA AAAGGCAATG TTACCAAATA
     201  TTAATTGAGT GTATGTAAAN GTTTGACCCA CTGGGAATGT GATGAAAGAA
     251  ATAAAAGCTG AAATAAATCA TTTTNTNTAA TATTATTCTG ACATTTCACA
     301  TTGTTAAAAT AAAGTGGTGA TCAGTGTGCA GGGTACGAGG TATTTGAGGT
     351  AGATATGTAC AAGAAGGCAG GGTAAAGTGA TTAGGCATCA GGATATATAA
     401  TAATAATAGC AAAATAAAGA ACAGAGCAGC AGCTGCAAAT GATGTGTGTG
     451  TTTGTGAGTG TGTAATGTGT ATGCATGTGT GTTTGGGTTG TGTCGGTTTG
     501  TGTGTGTGTG TGTGTGTGTG TGTGGGAGTG TGAATGTAGT GTGTGTGAGT
     551  GTGCATATGC TTTGTATACA GTcTAGTGAT TGTGCGTAGG GTCAGAGCAA
     601  GACCCTTTAA CTTTTTGAGG ATGCGAGTGC CCATGCCAAA TATTTCAGCC
     651  TCCTGGCGCG CCCTCTTCAC GACTGTGCAG GTGTGTGTGG GCCATGTTAG
     701  GTCGCCAAGG AACAGCCCTG TTGATGTGGA TGGGGACGTG CTCTCCCCTC
     751  TTGCTCCTGT AGTCCACGAT CAGCTCCTTG GTCTTACTGG CATTGAGGGA
     801  GAGGTTGTCC TGGCACCACA ATGCCAAGTC NCTGACCTCT CANCGCCTTC
     851  GTGNCGTCAG CAAACTTGAC ACACTGGGNC CCGACACACT GGGNCCCAAC
     901  CCACTGGCTC CAACCNCCTG NCTCCAACCC NCTGGATCCT CTAGAGTCGA
     951  CCTGCAGGCA TGCAAGCTTG AGTATTGCTA TAGTGNCACG NCANATAGCT
    1001  TGGGTANGNA ATGG

LOCUS    cmrSs3.7              412 bp              CHECK: 1554

       1  ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC
      51  ACACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACAC
     101  ACACACACAC ACACACACCA ACCACAGATA CCACTGTGTT GGGAnGATTT
     151  CTGCAACAGT GGnTACGTTA TTCTCTAAAT CTGATGATAT CCCTCAACGG
     201  GCTATAATCG TTGCATTTAA AGTAGTTTTT TCTCCCCGTG TTAAAAACCA
     251  GTGACTTGCA TAATTAGAAG AGTAGTAATG GAAATCAATG GAATGGAAAG
     301  ACAGCGGTTA ATTATACACA CAGTCTTCTC CCAAAATGGC ACCCTATTCC
     351  CTATATAATG CACTACTTTT GACCAGGGTC CATAGGGCAT TGGAGCACTA
     401  TAAAGGGAAC AA

LOCUS    cmrSs3.8              656                 CHECK: 2747

       1  AATTGTAATA nGnCTCAnTA TAGGGsGnAT TCGnGCTCGG TACCCGGGGA
      51  TCAGnCTAAA ATTGnGCTTT TGGCCATCAA GGnAAAACGC TATGTyTGGC
     101  GCAAACCCAA CACyTyTCAA CACyTCGAGn ACAACATCCC CACAGTGAAG
     151  CATGGTGGTG GCAGCATCAT GCTGTGGGGn TGTTTTTCAT CGGCAGGGAC
     201  TGGGAAACTG GTCAGnATTG AAGGAATGAT GGATGGCGCT AAATACAGAG
     251  AAATTCTTGA GGGAAACCTG TTTCAGTCAT CCAGAGnTTT GAGACTGGGA
     301  CAGnGGTTCA CCTTCTAGCA GGACAATGAC CCTAAGCATA CTGCTAAAGC
     351  AACACTTGAG TGGTTTAAGG GAAAACGTTT AAATGTGTGT GTGTGTGTGT
     401  GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT
     451  GTGTGTATAA tcattttata cagctgcaga gttttaaaac agcttgaatg
     501  ttgttgctat aaatcagtgt gcacgcgagc acgttcncnc tttccatcaa
     551  ttcaatttaa attgcatgca aaatagatcc tctagagtcg acctgcaggc
     601  atgcaagctt gagtattccn atagtnacac gccaaatagc ttggcgtagg
     651  gaatgg
```

LOCUS *cmrSs3.9* 947 bp 4E238D7B checksum.

```
     1  NAANGCCNGN TAATTTNAAN NGAAANCAAA NGGGGNTNAA NTTAAANGGG
    51  NNAAAGGACC CNNTTTTTTN ACGGGGGGGG GNATTAAGCC CCGGNTAANN
   101  NTTGGTGTTT TNGGGNGGGG GGNAAAATTN GATTTGGCAA CCCCAGATTA
   151  AANNGCCAAT TNTGGAAAAC CCCTTNGGGC AATGGGANGT GTGNTTNAAA
   201  ATTAAATTAA TAATAATTTT TTGGGAAAAT AGAATGCCTG ATAAGACATA
   251  ATTAACATCA TATGTTTCAG CTTAAATTAA ANTAATTTAG TATTGGACTG
   301  GGGAGGCGGC GATGGAGAAA GGCCATGACC CGTGATTNTT TGNGTGAACG
   351  AGAATGGCCA CTGACCTTGT GATCACCCCA TACCTGGACA TCACACACAG
   401  TCCGACCATG CGGTAATGCC TTTGACCTTC AACCTCCCAC TCCCAGTCAG
   451  CCAAACATAA CAAGGCTTAC CTAGTTTTTA CAGCAGGGAA GATAGTAAGA
   501  CCTTCAGAAT GAAGAACTTT CACACTACAT CATCCAGTCA CATCTATTTA
   551  CTTCAGCATC CACCCCTGAC CCCAACCCAC TGCTCAGAGC TCCCCTCAAA
   601  CAAACACACA CACAATCCCA CTCCTATGTT TACCCGAAGC CAATGGTGGT
   651  CCAGTGGTTG CGTGTGCCCT GGTGCGGTCT GACCATGGGC AGGAGGTCAG
   701  GGGGGGGAGG TGATTGGTCC AGACACAATG TTGATGATGT CAGCATTCTT
   751  CAGCACAGGT ACCGGCTCCA TGGCCATCTC CACATGGTCC ATGATACGGT
   801  TCAGGTCTGA CTCAAACAGC TCCTTACCAA AACATGTACA ACACACACAC
   851  ATACACACAC ACACAGAGCA GTACAACTGA GGTTAGGCAT ACAGAGAGGG
   901  AGAGAGAGTC TGTGTGAGAG AGTGGGACAG TGAGAAAGCG AGAGAAA
```

LOCUS *cmrSs3.12* 1010 BP Check: 8808

```
     1  TTTAAGNGTA TTAANATGGG GAGCCTTCAG GNCCAGNAAN AANAGAGAGA
    51  TGTTCCNAAA CAAACCNGTC ANTGGGGAAN ACCCAG-GCC CTTTTCCNGT
   101  NTGNGGANGG ATNGTTAACC GGGGNTTGCA TGCATGGANT TTGAACANTC
   151  CTATGCTTTA CTNGNGAATT AATAGGGCTG NGGCGNTTTT GGCGCCTATT
   201  TTTTGTTTGC AACAGGAGGA AGGACCATTN TTGCAGTTTT CAGAGAAAGC
   251  TGGATATTTG TTTTAAAACA GTAGAGTATG TATGGGAAAC ATACTGACAA
   301  TAGTATGCCA CTAACTTCAT TAATTGATGG AGTAGTTGGT CAGTATAGCA
   351  CACATATGCT TGGTGTTTGT CACATCATTT GCTTAAAACA GCAGATNTGC
   401  --TGAGGAGA TGAGTGAAAT GTAATGTCAC GTAATGTGTG TGTGTGTGTG
   451  TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG
   501  TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTAGAGCTT
   551  GAGTAGATTT AGGCTTTTTT GGAAAACTCA GAGGGACTCC ACTGTGACCT
   601  CCCGCTGCCG GCTCAAAGTC TATTACTGGA AAATGAAATG GTGGCATTTT
   651  AAGACGATTT AAGGGATATG AGGGGATTTT TGAGCCCTTT TTGGATTTAT
   701  CTCTGGCTGT GATAAGTGAT ATGAAAATTC CTCCCATTAA TGAATGAATA
   751  GAGAGTGACA ACGTACACTG AACAAAAATA TAAACTGAGC TACAGTTCAT
   801  ATAAGGAAAT CAGTCAATTG AAATAAATTC AGTAGGACCT AATCTCTGGA
   851  TTTCACATGA CTGGGAATAC AGATATGCAT CTGTTGGTCA CAGACACTTA
   901  AAAAAAAAGG TAGGGGCGTG GANCCNCTAG AGTCGACCTG CAGGCATGCA
   951  AGCTTGAGTA TTCCTATAGT NNCACGCCAA TAGCNTGGCG TANNAATGGN
  1001  NNNNAANGNA
```

15.4.2 RDA CLONES

```
LOCUS       RDA3.1              264 bp        BDEA591E checksum.
ORIGIN
       1  CTAACTTcGC CTCGGCACGG TGGGCCGtTT TGTGGAACTC CTCCTCGCGA
      51  GCGGcCTGCG CCCTCCGGAC GAGTTCCCTC TCCCTCTCAG CCAGCACCTC
     101  CTTCTGTCTC TCCACCGTCC CCTTCAACAC CTCCACATCA GACTGCATAC
     151  CTGACAGACA GCCAGAGAAG AAAGGAAGGG AGAGAGAGAG AGTTGAAGGA
     201  GAGAGACGGG AAGGAAAGAG TTAAGTTTGC AGACTAAAAT GAGAGTGATG
     251  TACAGTACAT GTAT

LOCUS       RDA3.2           379 bp           AC0BF29E checksum.
ORIGIN
       1  TGCCTTACAA CAGGCCTACA CGCCCTCAGT CCAGACTCTC TCAGCCTATT
      51  TTGGACAGTC TGAGCACTGA TGGAGGGATT GTGCCTTCCT GGTGTAACTC
     101  TGGCAGTTGT TGTTGCCACC CTGTACCTGT TCCGCAGGTG TGATGTTCAG
     151  ATGTACCGAT CCTACATCTG TATGTAATGG AGGCAGGTCA GACAGGAGGG
     201  GTGCTGGCTG GCACGGAGTC AGTCAGGCTG TGGCCTGCTC AGCTCCATGC
     251  TGTAGATGTA GCTGTGGGAG ACAGGCCTCC TGCTCCTCTC TCAGACGCTG
     301  CACTGCTGAG GAGACACCTG CAGCAACATC ACAGTCAGGA CCTGTCACAG
     351  CCAATGATGG GTGGCCGTTG GCAGAGAAG

LOCUS       RDA3.4            260 bp           40CEBA03 checksum.
ORIGIN
       1  CAGGCTTTTC GCTGGCCTTG GCAGAACACT GACATTCCTG TCTTGCAGGG
      51  AATCACGCAC AGAACGAGTA GTACGGCTGA TGGCATTGTC ATGCTGGAGG
     101  GTCATGTCAG GATGAGCCTG CAGGAAGGGT ACCACATGAG GGAGGAGGAT
     151  GTCTTCCCTG TAACGCACAA CGTTGAGATT GCCTGCAATG ACAACAAGCT
     201  CAGTCCGATG ATGCTGTGAC ACACCGCCCC AGACCATGAC GGACCCTCCA
     251  CCTCCAAATC

LOCUS       RDA3.5       256 bp              3D525748 checksum.
ORIGIN
       1  TCCTCCCAGA CCTGGACTAA AGCATCCGCC AACTCCTGGA CAGTCTGTGG
      51  TGCAACATGG TGTTGGTGGA TGGAGCGAGA CATGATGTCC CAGATGTGCT
     101  CAATTGGATT CAGGTCTGGG GAACGGGCGG GCCAGTCCAT AGCATCAATG
     151  CCTTCCTCTT GCAGGAACTG CTGACACACT CCAGCCACAT GAGGTCTAGC
     201  ATTGTCTTGC ATTAGGAGGA ATCCAGGGCC AACCGCACCA GCATATGGGG
     251  TCTGAG

LOCUS       RDA3.6           266 bp           BD327181 checksum.
ORIGIN
       1  CTTACACCCC TTGTGAGAAC ATATGCTGGT GCGGTTGGCC CTGGGTTCCT
      51  CCTAATGCAA GACAATGCTA GACCTCATGT GGCTGGAGTG TGTCAGCAGT
     101  TCCTGCAAGA GGAAGGCATT GATGCTATGG ACTGGCCCGN CCGTTCCCCA
     151  GACCTGAATC CAATTGAGCA CATCTGGGAC ATCATGTCTC GCTCCATCCA
     201  CCAACACCAC GTTGCACCAC AGACTGTCCA GGAGTTGGCG GATGCTTTAG
     251  TCCAGGCCTG GGAGGA

LOCUS       RDA3.7       258 bp              9D872018 checksum.
ORIGIN
       1  TCCTTTACGC CACCAACCCC TGTGGCCGTG TTTAGCAACA ACAAAGCCTC
      51  CCTAATGGGC TCCCATTTCC CACTACTAAC GCAGGTTTAA TAGCAGGGTG
     101  TGTGTTCTCT CCCTGCTCCC CATCTCAGTG GGACACCCTT CCCCTCTCTC
     151  CCTGTGGGCT CTCTCTCACA CAGAGGCAGT CTGGGAAGGC ATCTCCCCTC
     201  CGCCTCGCTG CGCAGTGATG TCATGGGGAG CGCTCTCTTT TTACGAGCGC
     251  AGCTCTGT
```

```
LOCUS       RDA3.9       295 bp                80F64CBC checksum.
ORIGIN
       1  ATGTTTCCGG CAGCGCCAGG AAGTCTAGGG ACTGGAGGGT AGCATAGGCT
      51  GAGATGAACT CTGCCTTGTt GGCCGCAGAC CGGCAGTTCC AGAGGCTGnT
     101  GGAGACCNGG AACTNCACGN GGNTCGTGCG CGCTGGGACC ACCAGNATAG
     151  AGTGGCAGNG GNCACGCGGC GNGGAGCGCT TGNANGGCCT GNGCAGAGGA
     201  GAGAGAACAG GGATAGGCAG ACACACAGNA GACAAGCTAC AGAAAGAGGC
     251  TACNCNAAAT GCAAANGANC CGNNCATGGA TAGACGACGC CGNCA

LOCUS       RDA3.10          286 bp           651CDD80 checksum.
ORIGIN
       1  GAGTGTAGGT TTTTTCAGAA GGGGTGCAAC TCTCGCTCTC TTGAAGACGG
      51  AAGGGACGTA GCCAGCGGTC AAGGATGAGT TGATGAGCGA GGTGAGGAAG
     101  GGGAGAAGGT CTCCGGAAAT GGTCTGGAGG AGAGAGGAGG GGATAGGGTC
     151  AAGTGGGCAG GTTGTTGGGC GGCCGGCCGT CACAAGACGC GAGATTTCAT
     201  CTGGAGAGAG AGGGGAGAAA GAGGTCAAAG CACAGGGTAG GGCAGTGTGA
     251  GCAGGACCAG CGGTGTCGTT TGACTTAGCA AACGAG

LOCUS       RDA3.11       272 bp               CFEC8138 checksum.
ORIGIN
       1  AGAACTTGAA AGCCATCGCG GACCAAGTGA TAGCGGTGGA TGGCCAGCTT
      51  CTTCCGGGGG TGAGTGACTG GTGATAGCCC CATTTCCAAA AATAACCTCA
     101  TGTATCAGGT GACGCACAAA CAGAGGTCAC TTCGAGAGGT GCTGTTGCTG
     151  CCGCCCCGTC CCCGGTACGT CAGAATCGTT CGTCAGCCGC TGGCCTGGGA
     201  TGAGGCGGGA CATGGAAGAC TATTGCCGCA GCTGCCCGGA GGTGCCAAAG
     251  CACACCCGCT GGTTCCCCTA CG

LOCUS         RDA3.13         328 bp            CHECK: 6443
ORIGIN
     1  CTGTAGGTTC ATGCTCTACA ACATTCGCAn AGTACgACCC TGCCTCACAC
    51  AGGAAGCGGC GCAGGTTCTA ATCCAGGCAC TTGTCATCTC CCGTCTGGAT
   101  TACTGCAACT CGTTGTTGGC TGGGCTCCCT GCCTGTGCCA TTAAACCCCT
   151  ACAACTCATC CAnAACGCCG CAGCCCGTCT GGTGTTCAAC CTTCCCAAGT
   201  TCTCTCACGT CACCCCGCTC CTCCGCCCTC TCCACTGGCT TCCAGTTGAA
   251  GCTCGCATCC GCTACAANAC CATGGTGATT GCCTACGnGA GCTGTGAnGG
   301  nAnCGGCACC TCCATACCTT CAGGCTCT

LOCUS          RDA3.14          387 bp              CHECK: 8962
ORIGIN
     1  ACCGACGTCg ACTATCCATG AACAGATCAg AAGGTGCCCC GGACCTCTCC
    51  GCGGAgTACC AGGACCTACG ANAGGTGTTC AACACTTTGC TTCCGCCGCA
   101  CTGACCCTAC gACTGCGAAA TTGACCTTCT CCCTAGCACC ACACCACCCT
   151  GGGGACGTCT GTACTCTCTG TCGGGACCGA AgACAAAGGC TATGGAAACC
   201  TACATTGGGG TCTCCCTAgC TGCAGGATTT ATCCGTCCCT CTTCCTCTCC
   251  CGCAGGnTTC TTCTTCATGG AnAnAnAAGG ACAAGTGCAT TGACTACCGG
   301  GGATGCAATG ACATTACTGT TAAnAACGTT ACnACTCGTT TCCTCGGCnT
   351  TCAACTGCTC CANGGGGCnA CTGTNTnTTn CnAnCTG

LOCUS          RDA3.15          216 bp              CHECK: 5439
ORIGIN
     1  ATTTCACCCA CATGTAGTTG TTCTCTGCAC ACAAGCTTGG CATGTTACTA
    51  GTGTGGGATA AAATGGACTG GGTGCCTGTT TTTAACGTAG GGTTTTATAA
   101  gAACACATTG ACAACACAAT TTTTGTAATA GGAGCAgAAG CAGGGTTTTG
   151  TAAATCTGGT TTATCAATAA ATTAACGTGA AACCTGGTAA CACTGCTGTT
   201  TGCAACACAC CGTGTG
```

```
LOCUS           RDA3.16          244 bp            CHECK: 5359
ORIGIN
     1  GAATCCNCNA CCATGGCGTT NTTAGCACCA CGCTCTAACC AACTGAGCTA
    51  ACCGGCCGCT AGCTACGCTG CACAATTGTT GAAATGCACC GAGCATCTGC
   101  AANTGCTGGC GTNTNCACTA ACAAGTCCAC CAACAGCATA NCCGGACNAN
   151  ACTGGACCCT CTTGGTTGCT CATAGTGGCG TTTCTACCCA ACTGGGGAAT
   201  TAGCTCAAAT GGTAAANTGC TTGCTTAACA TGTGANAAGT ATGG

LOCUS       RDA4.1          157 bp            1D64E14E checksum.
ORIGIN
       1  GGTAAGGaaA TCGAcCAAcA GATACGAGCA AGGCCGTNGN GGAACAGGTA
      51  AGGGGTGAAG CTTACCTCTG GGCAGGTGCC TAGGAGCCTT ACACTGGGCG
     101  CACACCGAGC AGGAGGAAAC ATAAACCGCA CGAATttGCC AAGGNGGGCC
     151  ANCAGNA

LOCUS       RDA4.2       225 bp             2629EDC5 checksum.
ORIGIN
       1  TGACTCATNA ACTTTCCTTA ACATGCATTG TTCTAACATG CCAGAGGATG
      51  TTGCCCTTGG AGACCTGCTG CGGATATGGG TACGGCCCGA CGTGAGATTT
     101  ACACCCTCTC TTACGGATTT TCAAGGGCCA GCAAGAGCTC ACCGGACGCC
     151  GCCGGAACTG CAACTCTTTC CAGGGCTTGG GCCCCTCTCT CGGGTTTCCC
     201  GCCCAGCTTC CGTGCTCACC GCGGT

LOCUS       RDA4-3      186 bp            4C39A437 checksum.
ORIGIN
       1  CACCTGCCCC TCAACTACCC CCCGGGCAGC CCCGACCTGG GCAGACACTA
      51  CCGCTCCAAC TCCCCACTGC CCTCCATCCA GCTGCAGCCC CAGTCCCCCT
     101  CTGCCTCCAA GAAgCACCAA GCCGTTCAGG ACCTGCCTGC CACCAACACC
     151  TTTGTGGGCA CAGGTGACAA CAACTCTACA GGCTCC

LOCUS       RDA4.4       230 bp             E89B4226 checksum.
ORIGIN
       1  ACATCAACAT GGTTTCTCTC CAATTCAGCG CTCACCAAGA TGGACGACTT
      51  TAGGGAATTC CACGTGTCCA CTACGACCAT ACTTATTGAT GCAGCAGACG
     101  CGGAAGCGGT AATCGCCCTC gCACGGTACA CTGTCGCCCA TGACCTCTGC
     151  GGAGGTCGCC GTGTCGGANG TCGNACACCT CANNCAGTCC TGNGATCTTC
     201  CTTTGTATAG NACAGTTGAC TAATCACTTT

LOCUS       RDA4.5       224 bp            F023751F checksum.
ORIGIN
       1  ACCGCGGTGA GCACGGAAGC TGGGCGGGAA ACCCGAGAGA GGGGCCCAAG
      51  CCCTGGAAAG AGTTGCAGTT CCGGCGGCGT CCGGTGAGCT CTTGCTGGCC
     101  CTTGAAAATC CGTAAGANAG GGTGTAAATC TCACGTCGGG CCGTACCCAT
     151  ATCCGCAgCA gGTCTCCAAG GGCAACATCC TCTGGCATGT TagAACAATG
     201  CATGTTAANG AAAGTTGATg ATCA
```

```
LOCUS       RDA4.06              587 bp              CHECK: 3162
ORIGIN
     1  CGCAtaGCAC TCTCCATCCT CTCACCGTAA CACTCTCCAT CCTCTCACCG
    51  CAAATCGTTG TGGGCTATNc TtcTGtcCTT TGTCTCACGA CgGTAATTTG
   101  GTGGTTGGAA ACATCTCTCT ACTGGTTgTN GGGCNTGTGC TTTGGCAANT
   151  TGGTTGGGGT TNATACCCTN CCTGTTTGGT CCTTTCCCNG GGGTTTCANC
   201  TTATTGGGNC CANCACTTTT TNCNTTATTT TCCCCCCCGA NNACCANAAT
   251  TTCCCTNTTC CCTTAATTTC CGTCCCCTTC NTGTTCCAAC TATTTTCNNA
   301  ATACCCCCCC ACCCCCTTTT NANTTCCATN AATTNNTATT TTTCCNTCNN
   351  TTTTTTCCNC TANANTAACC TTCCCTTTNT TNCAGNTCNT NCCNCTTTTC
   401  CCCCTTTTAA ATTTTTTCCC CNTCCNNATT CCNCCCTNNT TTTTAACCCC
   451  AAAACTTTAT TTTTTAACCN CGNTTCCCAA TNAAATTNAN TTCATCCCNT
   501  TTNNATTNTT TCCCTCTCTC TCCCCTTTCT CTTCTGAAAA NCTTTTCCCC
   551  CCTTTTTATT NAATNATTAT CCCCCNAAAA ATTNTCC

LOCUS       RDA4.7       268 bp          9754EAF2 checksum.
ORIGIN
     1  CATCAGGAAG TCCACCGACA GGTGCGACCA CgGCCGTTGT GGAACGGGTA
    51  GGGGTTGTAA CTTCCCTCTG GGCAGGTGCC TGGGAGCCTT GCACTGGGca
   101  cacaCCGAGC AGGAGGAAAC ATAAACCCTC ACGTCCTTGG CCAAGGTGGA
   151  CCACCAGTAC TTCCCACTAA gACAGCgCAC TGTCTGACTG ATGCCAggAT
   201  gACCACANGA GGTGACgTGT TGGCCCAATA TATcTTCnCT tCGGATACAC
   251  AGTTGCCTAA TCacCTAg

LOCUS       RDA4.8        118 bp          1875E0FD checksum.
ORIGIN
     1  ACTCTCCAGC ACTCTCCAgC CTAgCACTCT CCAGCCTCTC ACCGCAGATC
    51  ACCACCTCAA GCTGAACCTC GGCAAgACGG AGCTGCTCTT CCTCCCGGGG
   101  AAgGACTGCC CGTTCCAT

LOCUS       RDA4.9          215 bp           CE600965 checksum.
ORIGIN
     1  CGCAAGCACT CTCCAGCCTC TCACCGTAGC ACTCTCCAGC CTCTCACCGC
    51  AGATCGTTGT GGGCTATACT CGGCCTTGTC TTANGACGGT AAGTTGGTGG
   101  TTGGAGACAT CTCTCTAgTG GTGTGGGGGC TGTGCTTTGG CAAAGTGGGT
   151  GGGGTTATAT CCTGCCTGTT TGGCCCTGTC CGGGGGTTTC ATCgGATGGG
   201  GCCACAGTGT CTCCT

LOCUS       RDA4.10           387 bp          5D7EF46D checksum.
ORIGIN
     1  TTTGTGGGCT ATACTCTGCC TTGTCTCAGG ATGGTAAGTT GGTGGTTGAA
    51  gATATCCCTC TAGTGGTGTG GGGACTGTGC TTTGGCAAAG TGGGTGGGGT
   101  TATATCCTGC CTGTTTGGCC CTGTCTGGGG GTATCATCGG ATGGAGCCAT
   151  AGTGTTTCCT GACCCCTCCT GTCTCAGCCT CCAGTATTTA TGCTGCAGTA
   201  NTTTGTGTCG GGGGCTAnGG TCAgTATGTT ATATCTGGAG TATnTCTCCT
   251  GTCTTATCCA GTGTCCTGTG TGAATTTAAG CATGCTCTCT CTAATTCTCT
   301  CTTTCGGTGG ACCTGAACCC TAGGACATGC CTCAGnAcTA CnTGGCATGA
   351  TANTCCTTGT TGTCCCCAnT CCACCTGGCC GTGCTCT

LOCUS       RDA4.11       326 bp          8C9DB10F checksum.
ORIGIN
     1  CTGTAGGTTC ATGCTCTACA ACATTCGCAG AGTACGACCC TGCCTCACAC
    51  AGGAAGCGGT GCAGGTCCTA ATCCAGGCAC TTGTCATCTC CCGTCTAGAT
   101  TACTGCAACT CGCTGTTGGC TGGGCTCCCT GCCTGTGCCA TTAAACCCCT
   151  ACAACTCATC CAGAACGCCG CAGCCCGTCT GGTGTTCAAC CTTCCCAAGT
   201  TCTCTCACGT CAnCCCGCTC CCCCGCTCCT CCACTGGCTT CCTGTTGAAg
   251  CTCGCATCcG CTACAAgAnC ATGGTGCTTG CCTACGGAnC TGTNAAGGGA
   301  ACGGCACCTC CATACCTTCA GGCTCT
```

```
LOCUS       RDA4.12            164 bp         8B6AB558 checksum.
ORIGIN
       1  CCTCCGATGT CAAGAACTAC AGACCAGTAT CCCTTCTTTT TTTTCTCTCC
      51  AAAACTCTTG AGCGTGCCGT CCTTGGCCAG CTCTCTTGCT ATCTCTCTCA
     101  GAATGACCTT CTCGATCTGC GGTGAGAGGC TGGAGAGTGC TACGGTGAGC
     151  GGCTGGAGAG TGCT

LOCUS      RDA5.02              131 bp             CHECK: 556
ORIGIN
     1  CTCAgACCCC TTGTGAGACC ATATGCTAgT GCGGTTGGCC CTGGGTTCCT
    51  CCTAATGCAA gACAACGCTA GACCTCATgT GGCTGGAGTG TGTCAGCAgT
   101  TCCTGCNAgA GAAGCATGAT GCTATGGNcT G

LOCUS        RDA5-03            167 bp            CHECK: 2207
ORIGIN
     1  CCACCTCAGC CGGCGCGCGC CGGCCCTCAC TTTCATTGCG CCACGGGGTG
    51  TgTTCGGAAA AAACCCTCTG ACTTGCGCGT GCGTTAgACT CCTTGGTCCG
   101  TGTTTCAAgA CGGGTCGGGT GGGTTGCCGA CATCGCCGCT GACCCCTGGC
   151  GCCAGTTTAC gTGAGCC

LOCUS        RDA5.04            377 bp            CHECK: 8417
ORIGIN
     1  CACACTCCAC GTCATTTATT TAAAGTAGGC TGAGGCTAAT TGTCATGTAA
    51  CACACCAAAA GTGAAgACgA TCTGCGGTGA gAGGCTGGAg AGTGCTAATC
   101  ACTAGTGCGG CCGCCTGCAG GTCgACCATA TGGGAgAGCT CCCAACGCGT
   151  TGGATGCATA GCTTGAGTAT TCTATAGTGT CACCTAAATA GCTTGGCGTA
   201  ATCATGGTCA TAGCTGTTCC TGTGTGAAAT TGTTATCCGC TCACAATTCC
   251  ACACAACATA CGAGCCGGAA GCATAAAGTG TAAGCCTGGG TGCCTAATGA
   301  gTGAgCTAAC TCACATTAAT TGCGTTGCGC TCACTGCCGC TTTCCAgTCG
   351  GAACTGTCGT GCCAgCTGCA TTAATGA

LOCUS         RDA5.05            162 bp          CHECK: 9998
ORIGIN
     1  AgATGGTGAC AGGTGTCCCA gCAGGGTGGT GACAGGTGTC CCAgCAGGGT
    51  GGTGACAGGT GTCCTAACTG GGTGGTAACC GGTGTCCCAg CAGGGTGGTG
   101  ATAGGTGTCC CAATAGGTGG TGACAGGTGT CCCAgCAGGT CAgANGTGAC
   151  AGTGTCCAgC AG

LOCUS         RDA5-06             245 bp          CHECK: 8565
ORIGIN
     1  CCACTACCTC TCGCATGCTA AGCAAGCACT CTACCATTTG AGCTAATTCC
    51  CCAGTTGGGT AgAAATGCCA CTATGAGCAA CCAAgAGGGT CCAGTCTCGT
   101  CCGGCTATGC TGTTGGTGGA CTTGTTAGTG TACACGCCAG CACTTGCAgA
   151  TGCTCGGTGC ATCTCAACAA TTGTGCTGCG TAgCTAGCGG CCGGTTAGCT
   201  CAGTTGGTTA gAGCGTGGTG CTAATAACGC CATGGTCGTG GATTC

LOCUS         RDA5.07            135 bp            CHECK: 515
ORIGIN
     1  gAGGCTCACC CCGTGGACGG TGTGAGGCCG GTAACGGCCC CCGTCGCGCC
    51  GGGGTCCGGT CTTCTCGGAg TCGGGTTGCT TGGGAATGCA gCCCAAGGTG
   101  GGTGGTAAAC TCCATCTAAG GCTAAATACC GGCAC
```

```
LOCUS         RDA5.08            225 bp           CHECK: 2896
ORIGIN
     1  TGACTCATCA ACTTTCCTTA ACATGCATTG TTCTAACATG CCAgAGGATG
    51  TTGCCCTTGG AgACCTGCTG CGGATATGGG TACGGCCCGA CGTGAGATTT
   101  ACACCCTCTC TTACGGATTT TCAAGGGCCA GCAAgAGCTC ACCGGACGCC
   151  GCCGGAACTG CAACTCTTTC CAGGGCTTGG GCCCCTCTCT CGGGTnTCCC
   201  GCCCAGCTTC CGTGCTCACC GCGGT

LOCUS        RDA5.09              197 bp          CHECK: 8763
ORIGIN
     1  CTGGACTTCC TGACGGGCCG CCCCCAGGTG GTGAGGGTAG GTAGCAACAC
    51  ATCTGCCACG CTGATTCTCA GCACTTGAGC TCCCCAGGGG TGCGTGCTCA
   101  GTCCCCTCCT GTACTCTCTG TTCACCCACg ACTGCATGGC CAGGCACgAC
   151  TCCAACACCA TCATTAAgTT TCCAgACgAC ACAACAGTGG TATGCCT

LOCUS         RDA5.10             169 bp          CHECK: 4389
ORIGIN
     1  TGCTCAATTG GATTCAGGTC TGGGAACGGG CAGGCCAgTC CATAGCATCA
    51  ATGCCTTCCT CTTGCAGGAA CTGCTGACAC ACTCCAGCCA CATGAGGTCT
   101  AGCATTGCCT TGCATTAGGA GGAACCCAGG GCCAACCGCA CCAGCATACG
   151  GTCTCACAAG GGGTCTGAG

LOCUS        RDA5.11             177 bp          CHECK: 723
ORIGIN
     1  TGGCTGGTAC CCCAGTATGC ACTGGAAGGG GGAgAGGTTA GTGGAGGAgT
    51  GGCAAAGCgA GTTCTGGGCC ACCTCTGCCC AGGGCACgAA CGCCGCCCAC
   101  TCCCCCGGCC GGTCCTGGCA ATAAgACCTC AgAAACCTAC CCACATCCTG
   151  GTTCACTCTC TCCACCTGCC CGTTACT

LOCUS       RDA5-12               266 bp         CHECK: 5698
ORIGIN
     1  TCCTCCCAgA CCTGGACTAA AGCATCCGCC AACTCCTGGA CAGTCTGTGG
    51  TGCAACGTGG CGTTGGTGGA TGGAGCGAGA CATGATGTCC CAgATgTGCT
   101  CAATTGGATT CAGGTCTGGG GAACGGGCGG GCCAGTCCCT AgCATCAATG
   151  CCTTCCTCTT GCAGGAACTG CTGACACACT CCAGCCACAT GAGGTCTAgC
   201  ATTGTCTTGC ATTAGGAGGA ACCCANGGCC AACCGCACCA GCATaTGGTC
   251  TCACAAgAAG TCTGAG

LOCUS        RDA5.13               266 bp        CHECK: 234
ORIGIN
     1  GATTTGGAGG TGGAGGGTCT GTCATGGTCT GGGGCGGTCT GGGGCGGTGT
    51  GTCACAGCAT CATCGGACTG AGCTTGTTGT CATTTCAGGC AATCTCAACG
   101  CTGTGCGTTA CAGGGAAAAC ATCCTCCTCC CTCATGTGGT ACCCTTCCTG
   151  CAGGCTCATC CTGACATGAC CCTCCAGCAT GACAATGCCA CCAGCCATAC
   201  TGCTCGTCTG TGCATGATTT CCTGCAAAAC AGGATGTCaG TGTTCTGCCA
   251  TGGCCAGCGA AgCCTG

LOCUS        RDA5.14               259 bp      CHECK: 7712
ORIGIN
     1  CTCAgACCCC TTGTGAGACC ATATGCTGGT GCGGTTGGCC CTGGGTTCCT
    51  CCTAATGCAA GACAATGCTA gACCTCATGT GGCTGGAGTG TGTCAGCAGT
   101  TCCTGCAAgA GGAAGGCATT GATGCTATGG ACCGCACGTT CCCCAgACCT
   151  GAATCCAATT GAGCACATCT GGGACATCAT GTCTCGCTCC ATCCACCAAC
   201  GCCACGTTGC ACCACAgACT GTCCAGgAgT TGGCGGATGC TTTAGTCCAG
   251  GTCTGGGAA
```

```
LOCUS         RDA5.15                  224 bp     CHECK: 2445
ORIGIN
     1  ACCGCGGTGA GCACGGAAGC TGGGCGGGAA ACCCGAAAgA GGGGCCCAAG
    51  CCCTGGAAAG AGTTGCAGTT CCGGCGGCGT CCGGTGAGCT CTTGCTGGCC
   101  CTTGAAAATC CGTAAgAgAG GGTGTAAATC TCACgTCGGG CCGTACCCAT
   151  ATCCGCAGCA GGTCTCCAAG GGCAACATCC TCTGGCATGT AgAACAATGC
   201  ATGTTAAGGA AAGTTGATgA gTCA

LOCUS          RDA5.16                  168 bp     CHECK: 7208
ORIGIN
     1  CCACCTCAGC CGGCGCGCGC CGGCCCTCAC TTTCATTGCG CCACGGGGTg
    51  TGTTCGGAgA AAACCCTCTg ACTTGCGCGT GCGTTAgACT CCTTGGTCCG
   101  TGTTTCAAgA CGGGTTGGGT GGGTTGCCTA CATCgCCGCT GACCCCTNGC
   151  gcCAGGTTAC TTTgAGCC

LOCUS         RDA5.17                  157 bp    CHECK: 4303
ORIGIN
     1  gATTTGGAGG TGGAGGGTCC GTCATGGTCT GGGGCAGTGT GTCACAGCAT
    51  CATCGGACTG AGCTTGTTGT CATTGCAGGC AATCTCAACG CTGTGCATTA
   101  CAGGGAAAAC ATCCTCCTCC CTCATGTGGT ACCCTTCCTG CATGCTCATC
   151  CTGACAT

LOCUS         RDA5.18                  167 bp    CHECK: 1895
ORIGIN
     1  CCACCTCAGC CGGCGCGCGC CGGCCCTCAC TTTCATTGCG CCACGGGGTG
    51  TGTTCGGAAA AAACCCTCTg ACTTGCGCgT GCGTTAgACT CCTTGGTCCG
   101  TGTTTCAAAA CGGGTCGGGT GGGTTGcCGA CATCgCCGCT GACCCCTGGC
   151  GCCAGTTTAC GTGAGCC

LOCUS         RDA5.19                  267 bp    CHECK: 8891
ORIGIN
     1  CTCAgACCCC TTGTGAGACC ATATGCTGGT GTGGTTGGCC CTGGGTTCCT
    51  CCTAATGCAA GACAATGCTA gACCTCaTGT GGCTGGAGTG TGTCAGCAGT
   101  TCCTGCAAAA GGAAGGCATT GATGCTATGG ACTGGCCCGC CCGTTCCCCA
   151  gACCTGAATC CAATTGAGCA CATCTGGGAC ATCATgTCTT GCTCCATCCA
   201  CCAAAGCCAC GTTGCTCCAC AgACTGTCCA TGAgTTGGCA gATGCTTTAg
   251  TCCGGGTCTG AAAAGGA

LOCUS         RDA5.20                  165 bp    CHECK: 9409
ORIGIN
     1  AAATGGAGAA ACAGATGGTC GGAGGCGATT TCGACCCCCG GGCGTCTTTT
    51  GTCGAGTTGG AAAAgAATTT CGACCCGAGT GAAATGGAag ACGGCGAGAC
   101  GGGAGACGCG AGACGANAGA CGCGAGACGG GAnnAAGTGG ANnGGAgCAT
   151  GGTTACGATC TCCAG
```