A Comparison of Categorical vs. Fractional Parental Allocation Based on Microsatellite Markers to Estimate Reproductive Success and Inbreeding Levels over Three Generations of Selective Breeding in a Closed Population of Rainbow Trout (*Oncorhynchus mykiss*)

by

Kathryn Anne MacEachern

Submitted in partial fulfilment of the requirements

for the degree of Masters of Science

at

Dalhousie University

Halifax, Nova Scotia

December 2011

DALHOUSIE UNIVERSITY

DEPARTMENT OF BIOLOGY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "A Comparison of Categorical vs. Fractional Parental Allocation Based on Microsatellite Markers to Estimate Reproductive Success and Inbreeding Levels Over Three Generations of Selective Breeding in a Closed Population of Rainbow Trout (*Oncorhynchus mykiss*)" by Kathryn Anne MacEachern in partial fulfilment of the requirements for the degree of Master of Science.

Dated: December 5, 2011

Supervisor: _____

Readers: _____

_____

DALHOUSIE UNIVERSITY

DATE: December 5, 2011

AUTHOR: Kathryn Anne MacEachern

TITLE: A Comparison of Categorical vs. Fractional Parental Allocation Based on Microsatellite Markers to Estimate Reproductive Success and Inbreeding Levels Over Three Generations of Selective Breeding in a Closed Population of Rainbow Trout (*Oncorhynchus mykiss*)

DEPARTMENT OR SCHOOL: Department of Biology

DEGREE: MSc CONVOCATION: May YEAR: 2012

_____

Signature of Author

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The aim of this project was to assess three DNA-marker based pedigree reconstruction approaches and their associated challenges, strengths and weaknesses by conducting a retrospective analysis of a real, three generation rainbow trout (*Oncorhynchus mykiss*) pedigree from the SPA hatchery. Molecular genetic data at as few as three or four loci was used to infer relatedness among individuals and between generations in the reconstruction of the three full pedigrees. Parentage and pedigree reconstruction was estimated, for the quasi-categorical (exclusion-based and LOD-based) approaches via the program CERVUS 3.0 and for the fractional approach via a software (PIPEDIGREE), developed for this project. The fractional pedigree method appeared superior, particularly for the estimation of inbreeding levels. This retrospective analysis was able to demonstrate, under different pedigree reconstruction approaches, that the semi-selective, on-farm breeding scheme implemented at the time was successful in limiting the level of inbreeding increase and identifying possibly superior broodstock.

LIST OF ABBREVIATIONS USED

ACOA	Atlantic Canada Opportunity Agency
AD	Adipose clip
DFO	Department of Fisheries and Oceans
DNA	Deoxyribonucleic Acid
h^2	Heritability
IRAP	Industrial Research Assistance Program
LBD	Loch Bras D'Or lake
MCMC	Markov Chain Monte Carlo
MGPL	Marine Gene Probe Laboratory
NC	No clip
N_e	Effective population size
NRC	National Research Council of Canada
PCR	Polymerase Chain Reaction
PT	Probabilized trios
QTL	Quantitative trait locus
R	Response to selection
RP	Right Pectoral clip
S	Selection differential

SPA	Salmonid Propagation Associates Coop Ltd.
SSR	Simple sequence repeats
WSS	White Silver Springs

ACKNOWLEDGEMENTS

I thank my supervisor, Dr. Christophe Herbinger for all of his patience and support in allowing me the opportunity to pursue this graduate degree on a part-time basis. This project turned out to be much larger and took much longer to complete than we ever thought it would. I also thank Philippe Fullsack for all of his patience, support and encouragement during the many hours spent reviewing data and his help in giving me a clearer understanding of the software "PIPEDIGREE" used for this project. I also thank Christophe Herbinger and Philippe Fullsack for their suggestions and help in putting pen to paper in the writing of this thesis. Lastly, I would like to thank my husband Shaun for his unending patience and support during this very long journey.

CHAPTER 1: INTRODUCTION

1.1 Pedigree, Inbreeding Control and Selective Breeding in Aquaculture

1.1.1 Background

Artificial selection can be thought of as the modification of a species for a desired trait over time through human intervention. In aquaculture breeding programs, selection is usually based on desirable traits such as size or age at maturity among others (Dube and Mason, 1995; Herbinger *et al*., 1995) in an attempt to produce a more robust strain better suited to a particular environment. The selection of broodstock is then based on choosing potential parents with superior genetic merit (or breeding value) for the additive genetic effects of that desired trait, whose genes will then selectively be passed on to the next generation. At the same time, a sufficient level of genetic variability must be maintained in the population to avoid a genetic bottleneck and inbreeding increase and allow the population to adapt to the selective pressure over generations (Weaver and Hedrick, 1989; Falconer and Mackay, 1996; O'Connell and Wright, 1997).

Inbreeding is a measure of the proportion of alleles that are identical by descent and is recognized by an increase in the level of homozygosity observed in a population with a corresponding reduction in heterozygosity (Weaver and Hedrick, 1989; Falconer and Mackay, 1996). The inbreeding coefficient of an individual is the probability that two homologous alleles are identical by descent and can be calculated from pedigrees by tracing paths to its common ancestors (Weaver and Hedrick, 1989; Falconer and Mackay, 1996). The method of estimating the inbreeding coefficient, by counting the number of common ancestors, is straightforward but the accuracy of the estimate is dependent on the accuracy of the pedigree information. However, obtaining complete and accurate pedigree information in aquaculture populations is generally very difficult (see below).

In closed aquaculture populations under selection, mating between relatives is very likely because the number of actual contributing parents may be relatively small. This is related to the notion of effective population size N_e (Falconer and Mackay, 1996).

When strong differences in fertility exist, only a small number of individuals may actually produce most or all of the progeny for the next generation, reducing the effective number of breeders. As a result, continual breeding of the same strain with limited parental lineage and limited introduction of new genetic material will lead to an increase in the rate of inbreeding from generation to generation and potentially limiting the response to selection (Weaver and Hedrick, 1989; Falconer and Mackay, 1996). Empirical studies by Pante *et al*. (2001) on three populations of rainbow trout under selection for six generations, found that a moderate effective population size (N_e) of 25 - 94 individuals was large enough to maintain an acceptable level of inbreeding increase. However, rates of inbreeding calculated from N_e were generally lower than the inbreeding rates calculated from pedigree information. Rates of inbreeding of 1.3% per generation or 0.4% per year were obtained when calculated from pedigree information, while the rate of inbreeding calculated from N_e was ~ 0.9% per generation or 0.3% per year. These levels were found to be below the level of 3 - 5% per generation previously observed in commercial salmonid farms (McKay *et al*., 1992; Pante *et al*., 2001). For most species, an increase in the rate of inbreeding will result in inbreeding depression (i.e. a reduced fitness for important performance traits such as growth, viability and survival) and may cause potentially deleterious abnormalities (Dube and Mason, 1995; Falconer and Mackay, 1996; Pante *et al*., 2001; McLean *et al*., 2005). In addition to the possibility of inbreeding depression, increased inbreeding is also indicative of a loss of genetic variability which could be associated to decreasing genetic gains under artificial selection.

The response to artificial selection (R) can be measured as the heritability (h^2) times the selection differential (S), where the selection differential is the difference between the mean of a population for a particular phenotype and the mean of the individuals selected to be parents of the next generation. It is an estimate of the expected gain that can be achieved in one generation of selection (Weaver and Hedrick, 1989; Falconer and Mackay, 1996). A rate of 10 - 15 % genetic change per generation has been observed for some fish species (Gjedrem, 2000) and has been as high as 25% for weight gain in brown trout, *Salmo trutta* L. (Vandeputte *et al.*, 2004). Observed differences in progeny growth among dams and sires were observed where complete factorial crosses

between 10 sires and 10 dams was performed in a small rainbow trout farm (Herbinger *et al.*, 1995). Some females did produce better surviving and better growing progeny than other females. These effects were attributed to genetic effects. The parent fish that produced superior offspring (in terms of growth and survival) had presumably higher breeding values and could have been preferentially chosen to be the next year broodstock. However, one of the major concerns in a closed aquaculture breeding programs is the conservation of an adequate level of genetic variability that can sustain long term genetic improvement goals and prevent inbreeding (Dube and Mason, 1995; Herbinger *et al.*, 2003). The sustainability of a long term genetic improvement initiative then becomes a delicate balance between genetic gain (selection for desirable phenotypic trait) and genetic loss (decreased genetic variability and increase in the rate of inbreeding) leading potentially to inbreeding depression.

For any species, a central part of running a selective breeding program is the knowledge of the pedigree of the individuals under selection. Such pedigree knowledge is necessary to calculate individual and population inbreeding levels and as well to evaluate the breeding values of the individuals under selection. However, populations of aquatic organisms under culture are generally characterized by poorly known pedigrees. This is a consequence of the very small size of most aquatic species at hatching, which prevent the identification of the minute progeny with any sort of external marking (Dube and Mason, 1995; Herbinger *et al.*, 2003). Typically the female broodstock are monitored until they are ripe with eggs. The eggs from one or several females are manually stripped into containers, and fertilized with milt from one or several males (Dube and Mason, 1995). Most aquaculture breeding programs then rely upon rearing the different genetic units (e.g. families or groups of fry from particular crosses) in separate environments such as tanks, until the individuals are large enough that they can be marked. This is an approach that must take place in a specialized facility and it is a time consuming, costly and labour-intensive process (Dube and Mason, 1995; Herbinger *et al.*, 2003). Furthermore, genetic effects are confounded with environmental tank effects at least initially. Alternatively, all individuals can be reared in a common environment from birth and the assigned to their family later on based on their genetic profiles. Many different types of genetic marker can

potentially be employed to achieve such a goal but the most commonly used markers over the past two decades have been microsatellite markers.

1.1.2 Microsatellite Markers and Genetic Diversity

Microsatellites or simple sequence repeats (SSR's) consist of short (1-6 base pair) tandem arrays (O'Connell and Wright, 1997). They are ideal for population genetic studies because they are abundant in the genome, exhibit high levels of allelic variation, are co-dominant, are inherited in Mendelian fashion and are selectively neutral (Wright and Bentzen, 1994; O'Connell and Wright, 1997; Liu and Cordes, 2004). Microsatellite polymorphism is based on size differences due to varying numbers of repeat units observed in different alleles at a given locus. The highly variable nature of microsatellites makes these markers particularly suited for studying both population genetics (e.g. stock structure, effective population size) and inheritance of traits important to aquaculture (O'Connell and Wright, 1997; Ferguson and Danzmann, 1998; Liu and Cordes, 2004). The level of genetic variability in a population can be measured by calculating the average observed heterozygosity (the frequency of heterozygotes averaged over the number of loci tested) as well as allelic richness (the number of alleles observed at a particular locus in a specific population) (Falconer and Mackay, 1996; Weaver and Hedrick, 1989). Microsatellite molecular markers have been used to study genetic diversity in several marine species for over a decade (e.g. Wright and Bentzen, 1994; Garcia de Leon *et al.*, 1995; Morris *et al.*, 1996; Tessier *et al.*, 1995; Estoup *et al.*, 1998; Neilson, 1998; Perez-Enriquez *et al.*, 1999; O'Connell and Wright, 1997; Borrell *et al.*, 2004 and Liu *et al.*, 2005). In the context of aquaculture, a lower level of genetic diversity was reported for three hatchery strains of Polish sea trout (*Salmo trutta*) due to a smaller number of contributing parents relative to the natural ancestral population (Was and Wenne, 2002). Similarly, a study by Liu *et al.* (2005) identified a reduced level of genetic diversity (fewer alleles per locus, a smaller number of low-frequency alleles, a smaller number of unique alleles, a smaller number of genotypes and a larger number of the most common alleles) in two hatchery stocks of Japanese flounder (*Paralichthys*

olivaceus) as compared to a common (wild) population. The lower variability could be attributed to the low number of parents typically used in hatcheries. These studies provide empirical evidence of the need to monitor and maintain an adequate level of genetic variability in selective breeding programs.

1.1.3 Microsatellite Markers in Parentage and Sibship Reconstruction

Microsatellite markers have also been used successfully to establish kinship, paternity and parentage relationships in aquaculture (O'Connell and Wright, 1997). Several studies have already shown that it is possible to successfully determine the pedigrees of communally raised fish using microsatellite markers when parental DNA samples are available. For instance, Herbinger *et al.,* (1995) were able to determine pedigrees in a mixed farmed population of rainbow trout where the progeny of a full factorial mating of 10 sires by 10 dams were reared in a common environment for one year. Using four microsatellite loci, over 91 % were successfully traced back to one or two crosses out of the 100 potential crosses. Similarly, a parentage analysis of 792 Atlantic salmon from 12 full-sib crosses, reared communally, successfully assigned 98.4 % of offspring to one set of parents using four microsatellite markers (O'Reilly *et al.,* 1998). In a 48 X 2 factorial cross in rainbow trout, Fishback *et al.,* (2002), successfully assigned 91 – 95% of the progeny to one parental pair using 14 multiplexed microsatellites. A parentage analysis of 550 common carp offspring from a full factorial cross of 10 dams X 24 sires, using eight microsatellite markers, successfully assigned 95.3 % of the offspring to one set of parents (Vandeputte *et al.,* 2004)

Microsatellite markers have also been used successfully in pedigree reconstruction in aquaculture and natural populations when parental DNA samples were not available. Smith *et al.* (2001) were able to accurately partition individuals from the same large Atlantic salmon data set into full-sib families using as few as four microsatellite loci. Credible assignment to families was also seen among abalone offspring that had been

produced in a hatchery captive breeding program using a similar approach (Lemay and Boulding, 2009). An empirical study by Wilson *et al.* (2003) examined the marker-assisted estimation of trait heritabilities and genetic correlations in three strains of rainbow trout from 2 generations. They compared genetic parameter estimates generated from a regression-based model using estimates of pairwise relatedness and the same Markov Chain Monte Carlo (MCMC) procedure as in Smith *et al.* (2001) to reconstruct full-sibships and infer the fish pedigree when no parental information was available. However, they did observe some downward bias of genetic parameter estimates attributable to errors in pedigree reconstruction due to the failure to distinguish between full and half-sibling relationships in the population.

One aspect that has not been explored to a large extent either in parentage studies (when parental DNA samples are available) or in sibship reconstruction studies (when no parental DNA samples are available) is that of the inherent uncertainty that are associated with such studies. Assignment of individuals to families is generally probabilistic. Unless a very large number of loci are used, there could be several parental pairs that could have generated a specific offspring genotype. Furthermore, DNA profiles are subject to errors, and some putative parents may not have been sampled, which can also lead to difficulty or uncertainty in establishing the pedigree of the population. Potential sources of error that may be encountered when using microsatellites in parentage assessment and pedigree reconstruction are: 1) the mis-scoring of microsatellite alleles, often due to the interference of stutter bands, or more generally from human errors, 2) the presence of null alleles resulting from the non-amplification of an allele at a heterozygous locus such that it is incorrectly scored as a homozygous locus and 3) incompatibilities between parent and off-spring which may arise from mutations at a single locus. These sources of error must be taken into consideration during parentage determination and sibship reconstruction from microsatellite data, as each will have the same effect in causing false exclusions and can cause severe biases in sibship inference if they are ignored (O'Reilly *et al.*, 1998; Wang, 2004).

1.1.4 Genetic Evaluation with Uncertain Parentage

Pedigree reconstruction using molecular markers has been used to study many breeding systems (reviewed in Jones and Arden, 2003). There are three types of commonly used methods of pedigree reconstruction: simple exclusion, identification of the most likely set of parents, and fractional assignment of progeny among potential parents. Simple exclusion compares the multi-locus genotype of the progeny with all putative parental genotypes (Ellstrand 1984; Hamrick and Schnabel, 1985; Devlin *et al.*, 1988) in an attempt to eliminate incompatible trios. However, this may lead to exclusion of all parental trios (false exclusion) or on the contrary to ambiguities when several parental pairs are still compatible (incomplete exclusion). Identification of the most likely set of parents is an extension of the simple exclusion method but it incorporates a likelihood approach to determine the most likely set of parents from a pool of non-excluded parents (Smouse and Chakraborty, 1986; Devlin, 1988; Meagher and Thompson, 1987).

One advantage of the likelihood method, compared to exclusion is that it incorporates a model of genotyping error and allows occasional mismatches between parents and progenies. Either the likelihood is calculated and the trio maximizing the likelihood is chosen, or alternatively one tests the hypothesis of parentage compared to the hypothesis of unrelatedness through a likelihood ratio method. In contrast, the fractional method (Brown *et al.*, 1985) does not attempt to identify a single pair of parents but instead assigns some function, between 0 and 1, for each offspring to all non-excluded candidate parents (Devlin, 1988). Parent-offspring likelihoods are calculated in the same manner as in the categorical allocation methods, that is using Mendelian segregation probabilities. From a biological perspective the fractional method is not accurate in that an offspring can have only one mother and father but for studies of reproductive success from a population perspective, the fractional method may be better. Devlin *et al.* (1988) demonstrated that the fractional method of paternity analysis was the most accurate method to estimate reproductive success because it eliminated downward bias towards parental individuals with homozygous genotypes that are compatible with offspring. Smouse and Meagher (1994) compared estimates of reproductive success in *Lilaceae*

from a likelihood and fractional approach and determined that the fractional approach generated a more accurate distribution of male reproductive success than the likelihood approach because it made more complete use of the data set.

In contrast, very little has been published in the area of the impact of pedigree uncertainty for quantitative genetic parameters estimation such as breeding values or even less so for inbreeding levels. Perez-Enciso and Fernando (1992) noted that uncertain parentage would be associated with lower accuracy of evaluation and that the mis-identification of parents could bias downwards estimate of heritability compared to the true pedigree. They also showed with simulations that better accuracy and greater response to selection would be obtained using the average numerator relationship matrix ($\bar{A}$) introduced by Henderson (1988) which integrates over the pedigree uncertainty as compared to coarse phantom parent techniques which were commonly used at the time. Dodds *et al*. (2005) extended these results to the case where DNA markers were used to calculate the probabilities associated with fractional parentage assignment and showed with simulations that this method allowed much of the genetic progress that could have been made if the true pedigree had been known. It is notable however that the few studies that were found involved simulations only. The present work intends to address this gap by analyzing a real three generations data set collected in a rainbow trout farm which ran a selective breeding program using DNA information.

1.2 The SPA Selective Breeding Program

From December 1989 to December 1996, a novel farm based genetic improvement program was undertaken for an aquaculture facility in Cape Breton, Nova Scotia using DNA typing technology. The aim of the program was the development of a robust superior strain of rainbow trout (steelhead) (*Oncorhynchus mykiss*) adapted to local conditions of production specific to aquaculture facilities in Nova Scotia. The plan was to achieve genetic improvement for the aquaculture stock, on the farm, in real time, using DNA markers to provide the necessary pedigree information. This approach was

the first of this type in the world. The project was developed in collaboration between the Salmonid Propagation Associates Co-op Ltd. (SPA) fish hatchery and the Marine Gene Probe Laboratory (MGPL) at Dalhousie University. SPA was an aquaculture facility located in St. Peter's Fish hatchery in Cape Breton, Nova Scotia where they produced and hatched fish eggs and grew them to fry, fingerlings and smolt to supply aquaculture requirements in Nova Scotia. The hatchery produced seedstock for four species of commercially important fish in Nova Scotia: Atlantic salmon, Speckled trout, Arctic char and Rainbow trout. The SPA genetic improvement program was focused on rainbow trout raised in sea water (steelhead trout) which was at the time the subject of intense interest as an aquaculture alternative to Atlantic salmon. The Marine Gene Probe Laboratory (MGPL) was established in 1989 at Dalhousie University, Halifax, Nova Scotia, Canada as a pioneering laboratory dedicated to the management, enhancement and conservation of marine resources, using genetic analysis and what was termed at that time "recombinant" DNA technology.

SPA and the MGPL each had complementary interests in this project. If the selective breeding program was successful, it would provide SPA with a better, more consistent quality of genetically superior rainbow trout broodstock and allow the SPA facility to be more competitive as a supplier of seedstock in the aquaculture industry. The focus for SPA was to obtain basic information on its rainbow trout stock (such as the amount of genetic variability and the level of inbreeding), 2) obtain improved first generation fingerlings through re-spawning of the best parents identified by the MGPL, and 3) obtain improved second generation fingerlings through intense selection without an associated increase in inbreeding level. The MGPL interests were related to the development of the selection schemes, the necessary recombinant DNA tools and the computational programs needed to achieve the objectives outlined for SPA and the establishment of the SPA genetic improvement program as a pilot demonstration of the potential of a DNA fingerprint based genetic program for fish farms (Herbinger, 1993a; 1993b).

1.3 Objectives of the Present Study

While running from 1989 to 1997, the SPA selective breeding program managed to establish a number of pioneering techniques and approaches. Initially, DNA probes using the Southern blot/hybridization technique were developed and used, but they proved to be very unpredictable when they were tried on a large scale. In the summer of 1992, the MGPL switched to PCR based probes (microsatellite markers) which were a newer, cheaper and more reliable alternative. The first microsatellite markers for salmonids, and possibly for any fish species were developed for this program. To exploit the sort of co-dominant genotype information generated by these markers, what appear to have been one of the first exclusion-based parentage program and one of the first pair wise relatedness estimation program were developed as well. Finally the SPA program itself was the first aquaculture breeding program based on the idea of using pedigree information derived from genetic markers when fish were reared mixed from birth. In the course of its 9 year / 3 generations life span, a large database of 2396 fish genotypes was accumulated. Despite these achievements, the information contained in this large genotype collection was never fully analyzed. In particular, the main objective of the SPA breeding program, to avoid inbreeding accumulation in the population, was achieved by using the genotype information in a fairly simplistic manner. Genotypes of putative pairs of parents were contrasted to estimate if a specific pair of fish might be related, but the rest of the genotype information in that same generation and most importantly in the preceding generation was never utilized. In other words, assembling a proper multi-generation genotype database, using this information to reconstruct the multigenerational pedigree of the fish and assessing the evolution of individual and average inbreeding levels in the population under selection was not done.

To that end, the objectives of this study are:

1) To assemble all the genotype and mating information that accumulated in the program;

2) to use three methods to exploit the above information and construct a list of trios describing the pedigree structure, each with different degree of pedigree uncertainty;

3) to compare the impact of the type of list of trio/pedigree on estimation of reproductive success and individual inbreeding level;

4) to assess retrospectively the efficiency of the mating recommendations that had been provided to the hatchery to minimize crosses of related individuals;

5) to assess retrospectively whether it would have been possible to identify broodfish producing a high proportion of large offspring.

CHAPTER 2: MATERIALS

2.1 The Data Base

Genotypic and phenotypic data was made available from 3 successive generations of rainbow trout for this retrospective analysis. These fish had been produced and grown for the most part at the SPA hatchery in Cape Breton, NS. The data set was comprised of:

1. Genotypic data from three generations of fish at several single locus microsatellite markers: Omy 2, Omy 38, Omy 77 and Omy 105, for selected fish from each of three generations. Two other loci Omy 27 and Pupupy, were initially used as well for a few fish but were subsequently abandoned because of reliability issues (Pupupy) or because it was not very informative (Omy 27).
2. Phenotypic data from three generations of fish:
    - juvenile fresh water performance (i.e. large or small sized fish)
    - location of origin of original broodstock (i.e. Loch Bras D'Or or White Sulfur Springs)
3. Mating Design over the three generations

There were enormous challenges in compiling this multi-generation data set. Many hours were spent in compiling the data base, finding and associating the genotypic data from each group and understanding and associating the mating designs that were provided. In effect, one of the first tasks of this project was to reconstruct the "narrative" of this breeding program that was spread almost over a decade, to understand how the various files were associated and to understand what had been the logic behind various aspects of that program. Each aspect had its own set of challenges.

2.1.1 Compiling the Database

Bringing all of the separate pieces of information, the genetic data files, the mating design files and the phenotypic data files, together in order to compile the data

into a useable format for the purposes of this project represented a substantial challenge and involved a large amount of work. From the original data, a separate DNA fingerprint file with associated identifier was saved for each offspring group in various archived data bases at the time the DNA fingerprints were being developed. The genotypic information was stored separately from the phenotypic information and these files had to be married up in order to figure out associations between individuals. In addition, the mating design files were also archived separately from the DNA fingerprint and phenotype files. Most files had been archived on older mediums such a floppy disk, zip disk and older Iomega Bernoulli drive and had to be recovered and extracted from the sources. The file format was also obsolete (Paradox for DOS and Dbase III) and the information had to be imported into a more modern format (mostly Excel and text format). Even the particular way the genotype information had been collected and assayed had to be translated into a modern fragment size in base pairs. Allele scoring was performed at the time by running radio-labeled PCR amplified fragments on polyacrylamide gels with a 4 lane M13 vector sequence interspaced to provide a reference ladder. Each microsatellite fragment was then aligned to a particular base position and scored based on that position in the vector sequence. This product (raw data) had to be translated by subtracting the vector size from the raw data providing the actual size of the DNA fragment, ranging in size from 80 to 280 base pairs. The genotype files also had to be converted to a more conventional format for use in the various programs used in this study, such as CERVUS and PEDIGREE (referenced in the Methods section).

A great deal of time was spent at the beginning of this project to pull out the relevant data from the large archive of documents that had been created as the project progressed from year to year. Some information was retrieved from files created by various participants in the project over the period of years of data collection. Many of the files were duplicates or were new editions of previous files with minor additions, for example, the addition of iinformation about length or weight over several years of monitoring, or the confirmation of sex as the fish aged. In the end, some of the information was not pertinent to this project and was not used. In many cases, it was necessary to go back to the field notebooks or even the series of faxes that had been

exchanged between SPA and Dalhousie to try to figure out specific aspects of the program. This part of the project took a great amount of time to sort through the different files and folders of information provided and determine which files were the most current and accurate sources of information in order to proceed with the data analysis.

2.1.2 Genotypic Data

At the onset of the project in 1989-90, minisatellite hybridization based probes were developed but were later discontinued. In 1992-93, the project switched to PCR based microsatellites with 2 loci available, Omy 77 and Pupupy (Morris, 1993). In the following year, four more loci (Omy 2, Omy 27, Omy 38 and Omy 105) were added to the pool, and the "Pupupy" locus was discontinued because it was unreliable (Table 1). DNA typing profiles were provided at four single locus rainbow trout microsatellite markers: Omy 2, Omy 38, Omy 77 and Omy 105. The genetic information at each single locus marker was combined to create a four locus DNA typing profile for each individual in the study. Genetic data at a fifth less variable locus, Omy 27, was originally included for some of the generation 2 fish at the beginning of the project; however this locus was discontinued as it proved to be not very informative and the incremental cost of generating an additional locus was quite substantial at the time.

The genotypes were developed using PCR-based probes to amplify the areas of interest, each in separate PCR amplifications. The microsatellite markers used for this data set are co-dominant autosomal markers inherited in Mendelian fashion, where each individual is assumed to have two alleles at each locus, one inherited from each parent (Herbinger *et al.*, 1995; Wright and Bentzen, 1994). The probes used to isolate each locus were developed in the Marine Gene Probe Laboratory, Dalhousie University (Morris, 1993). The microsatellite alleles used in this study were amplified using polymerase chain reaction (PCR), resolved on vertical denaturing polyacrylamide gels and visualized using autoradiography initially, and non-radioactive DIG-labeling and detection subsequently (Herbinger *et al*., 1995, McConnell *et al*, 1997, Morris *et al.*, 1996). Allele scoring was

done manually by measuring the size of the PCR product relative to an M13 reference sequence which was run in 2 or 3 lanes on each gel. Allele scoring for all groups was performed by the same two technicians for consistency throughout the length of the project, with the exception of the last group (Parents 3Y9). For this group, the largest Omy 2 bands seemed to have shifted by 2 base pairs. This discrepancy was discovered at the time and was initially thought to have resulted from inconsistent scoring by a new technician; The scores were therefore revisited by one of the original technicians who concurred with the readings. Around the time the problem was first noticed, there had been a slight modification to the PCR reaction and it was concluded that this might have led to the slight shift. Subsequently, all the scores for this group were verified against a known steelhead reference sample and corrected accordingly in order to maintain continuity in scoring among the different groups of fish. DNA typing profiles with genetic information at three or more loci were available for most individuals; however, some individuals had profiles with genetic information at fewer than three loci, most probably due to technical problems during PCR. A summary of distribution of useable genotypic data for each offspring group per generation can be seen in Table 2. The various offspring groups are described in detail in the next section.

Table 1: Rainbow trout (*Oncorhynchus mykiss*) microsatellite markers.

Locus	Repeat sequence	Size range (base pairs)	Number of Alleles	Observed Heterozygosity	Source
Omy 2	$(GT)_n$	110 - 164	10	0.85	Herbinger *et al.*, 1995
		84-172	34	n/a	Heath *et al.*, 2001
Omy 27	$(GT)_n$	n/a	4	n/a	Herbinger *et al.*, 1995
		137-201	4	0.76	McConnell *et al.*, 1997
		99-111	5	0.69	Fishback *et al.*, 1999
		98-120	7	n/a	Heath *et al.*, 2001
		n/a	6	0.65	Wilson *et al.*, 2003
Omy 38	$(GT)_n$	94-130	6	0.70	Herbinger *et al.*, 1995
		139-247	50	0.88	McConnell *et al.*, 1997
		92-138	7	n/a	Heath *et al.*, 2001
Omy 77	$(GA)_n$	98-140	8	0.80	Herbinger *et al.*, 1995
		96-140	9	0.72	Morris *et al.*, 1996
		98-142	7	0.88	Fishback *et al.*, 1999
		98-134	13	n/a	Heath *et al.*, 2001
		n/a	10	0.74	Wilson *et al.*, 2003
Omy 105	$(GT)_n$	165-213	10	1.00	Herbinger *et al.*, 1995
		147-217	30	0.91	McConnell *et al.*, 1997
		109-271	40	0.91	Heath *et al.*, 2001

Potential sources of error that may present problems in parentage assessment and pedigree reconstruction were: 1) mis-scoring of microsatellite alleles often due to the presence of stutter bands, or more generally from other human error; 2) the presence of null alleles resulting from the non-amplification of an allele at a heterozygous locus such that it is incorrectly scored as a homozygous locus. Recognition of null alleles is a critical point in pedigree inference, since frequencies above 5% are considered to compromise pedigree determination (Marshall *et al.*, 1998; Castro *et al.*, 2004), and 3) incompatibilities between parent and offspring which may arise from mutations at a single locus. A study by O'Reilly *et al.* (1998) determined that the overall impact of mutations on the accuracy of assigning parentage was minimal with mutation rates at four Atlantic salmon loci on the order of ~ 10^{-3} and 10^{-4}. These sources of error must be taken into consideration during parentage determination and pedigree reconstruction from microsatellite data, as each will have the same effect in causing false exclusions and can cause severe biases in sibship inference if they are ignored (O'Reilly *et al*., 1998; Wang, 2004).

Table 2: Summary of genotypic data per offspring group.

Generation	Offspring group	No. of individuals	Data at ≥ 3 loci	Data at < 3 loci
1	Parents 1	224	224	0
2	Adipose clip (AD)	224	183	41
	AD PILOT	874	874	0
	No clip (NC)	140	119	21
	Right Pectoral clip (RP)	132	113	19
3	Group 1 Year 5	144	96	48
	Group 3 Year 5	141	138	3
	Group 5 Year 5	46	46	0
	Group 11 Year 5	40	40	0
	Parents 3 Year 8	309	308	1
	Parents 3 Year 9	122	120	2
Total:		2396	2261 (94%)	135

2.2 Chronology of the SPA Genetic Improvement Project

This section is essentially a narrative, describing the chronological development of this program, as it was reconstructed from different sources. Documentation included the project funding proposals "High-yield rainbow trout for aquaculture in Eastern Canada using recombinant DNA pedigree" (Herbinger, 1993a) and "Development of a farm-based genetic improvement program for the SPA hatchery using DNA fingerprinting technology" (Herbinger, 1993b) which had been submitted to the Industrial Research Assistance Program (IRAP) of the National Research Council of Canada (NRC) and the Atlantic Canada Opportunity Agency (ACOA). Interim and final progress reports related to the proposals submitted to the Department of Fisheries and Oceans (DFO) and the IRAP were also analyzed (Herbinger, 1994, 1995, 1997a and 1997b). The background information and the timing of events presented below is a compilation of the information obtained from these documents in conjunction with many personal discussions with Dr. Herbinger. It is presented in chronological fashion to illustrate the sequence of events, the evolution of the genetic improvement project and the derivation of the complex multi-generation pedigree data set used for this project. At the end of this section, all the information is condensed in Figure 1 which focuses on the timeline of events and Figure 2 which focuses on the succession of generations and the genealogy of individuals.

2.2.1 Generation 1

Year 1 (1988-1989)

For the purpose of keeping chronology, this year was considered as year 1 of the project which consisted mainly of the planning and organizing phase for the SPA genetic improvement project which would take place over the next several years (Figure 1).

Year 2 (1989-1990)

In order to begin the genetic improvement project, it was necessary to establish a genetic baseline and obtain basic information about the make-up of the rainbow trout

broodstock at the SPA facility. The broodfish were selected from two distinct groups of fish that were held at the farm at this time, the Loch Bras d'Or strain (LBD) and the White Sulphur Springs strain (WSS). These two strains of fish were thought to have separate origins (Herbinger, 1993a). However, no documentation was available to confirm the past history of these fish. In mid December 1989, two hundred and twenty four broodstock fish (generation 1 fish designated as parent 1 or "P1") were available and marked individually with opercular tags (125 from Loch Bras D'Or, 75 from White Silver Springs and 24 of unknown origin). Blood samples were collected from these fish for further DNA work.

Spawning took place in three separate episodes. A first spawning event of selected "P1" individuals was actually performed in early December, 1989 about one week before fish marking and sampling. All the females used in this spawning episode (n=14) were put aside and were therefore identified, marked and sampled later the following week as candidate mothers. The identity of the males used in this first set of crosses was not clearly documented and the males were not put aside. Consequently, for the purposes of this project, all possible males that had been selected by the SPA personnel for use in "P1" crosses were included as candidate fathers (n=21). In the absence of information on the specific crosses that were performed, the assumption had to be made that the spawning took place as one group. In other words, every male had an equal chance to fertilize the egg of any female. The progeny from this group did not receive a fin clip. One hundred and forty offspring from this spawning group were later on selected as broodstock and identified as "No Clip" broodstock fish (NC). These individuals were analyzed as generation 2 individuals (designated as "P2" NC or parents of generation 2 from the NC group).

At the same time as marking and sampling of broodstock fish took place in mid-December 1989, status of breeders was assessed and a simple mating plan was developed for subsequent crosses with the goal of avoiding crossing parents originating from the same strain to minimize the risk of crossing potentially related fish to minimize inbreeding. The crosses that were actually performed were designated as "CRP1Y2" (crosses of generation 1 fish, designated as Parents 1 (P1), performed in year 2). The

offspring of these crosses would become potential next generation broodfish (P2) that would be subject to a high selection strategy developed for the next phase of the genetic improvement project. The “CRP1Y2” spawning took place between the end of Dec. 1989 and mid-Jan. 1990 (year 2 of the program) in nine separate spawning groups. The spawning involved 40 candidate mothers and 17 candidate fathers in total. Each spawning group, representing the spawning that took place on one specific day, further represented a collection of small “pools” where the eggs from a small number of selected females (generally 2 or 3) were pooled and fertilized with the milt from a small number of selected males (generally 2). Thus the nine spawning groups represented a total of 25 pools. Following this design, the number and identity of candidate parents was very specifically documented and traceable. All progeny from this spawning episode were later tagged with a Right Pectoral fin clip (RP). One hundred and thirty two offspring were sampled later as broodstock and analyzed as generation 2 individuals and designated as “P2” RP individuals.

A third spawning of selected “P1” individuals was performed in late January, 1990. This spawning took place as one group hereinafter referenced as the “10 X 10” crosses. Ten females and 10 males were selected and crossed in a complete factorial set. All offspring were later tagged with an adipose clip (AD). Two hundred and twenty four offspring from this spawning group were later selected in May 1992 as broodstock. These individuals were analyzed as generation 2 individuals and designated as “P2” AD broodstock. In addition to the data from the 224 “P2” AD broodstock individuals, a large number of young AD offspring of this last factorial cross were also sampled and measured as part of a “pilot” experiment being conducted at the time (see section below). These younger individuals were also included for evaluation as part of the large multi-generational pedigree in the present project.

Year 3 and 4 (1990-1992)

Following the crosses performed in year 2, the resulting progeny were reared in the SPA facility. Juveniles from the last spawning (“10x10” crosses) were later sampled as part of a pilot study. The "PILOT" project took place in the SPA hatchery from

February 1990 to February 1991. In this experiment, ~4100 offspring from 100 possible parental pairs of the "10 X 10" crosses were reared communally for 1 year. In February 1991, the 443 largest and 430 smallest fish were then measured and sampled. As there was not good DNA marker system available at that time, these samples were archived for development of DNA typing profiles at a later date.

Spawning also took place in the 1991-1992 season. The resulting offspring were purely for production and were not supposed to be part of the breeding plan. As a result, no specific information on parental identity or crosses performed was collected, but hatchery records indicated that only a few parents were used. Most of the effort during the period 1990-1992 was devoted to generating genotypic information for the various fish that would allow estimating the relationship among individuals. As mentioned earlier, Southern blot DNA probes (minisatellites) were initially used. However; these hybridization probes proved to be very unpredictable when they were tried on a larger scale and the decision was taken to switch toward single-locus PCR-based probes (microsatellites). Two such probes were available by 1992; Omy 77 and Pupupy, (Morris and Richard, 1995). Additional markers were developed later (see next section, Year 6).

2.2.2 Generation 2

As mentioned above, 140 NC offspring (from the first spawning), 132 RP offspring (from the "CRP1Y2 "spawning) and 224 AD offspring (from the "10 x 10" crosses) were selected as potential broodstock for the next generation (P2). The genetic improvement part of the project actually started with the breeding selection scheme that began with the 1992-1993 crosses that took place in year 5 (when the fish were 3-year old) and progressed through years 6 and 7. One of the objectives of the farm-based genetic improvement project was to be able to, "at the end of a generation cycle, select the best fish in the whole production and determine the family make-up of these select fish to maintain a genetic basis as large as necessary to avoid subsequent inbreeding depression problems" (Herbinger, 1993a). A second objective was "the estimation of the

parent fish giving the best and the worst progeny, re-spawning of the former and elimination of the later" (Herbinger, 1993a). Finally minimizing inbreeding increase due to crossing of related fish was also a primary objective. The 1992-1993 crosses were designed to address these objectives.

Year 5 (1992-1993)

A selection scheme was implemented as part of the mandate of the SPA genetic improvement program to create a non-inbred line of fish. The initial DNA work was performed on selected fish representing the "best choice", a decision based on the size of the fish relative to the rest of the group and on other external appearance aspects as evaluated by the SPA team. Development of the breeding scheme was a two step process. The first step was the selection of the best fish (i.e. larger adults) by the SPA team from the entire production (thousands of fish) to be used as the next generation broodstock (496 fish, i.e. the 140 NC, 132 RP and 224 AD fish already mentioned). The second step was an evaluation of the potential level of relatedness of these fish using DNA markers to avoid sib mating and potential inbreeding. In the summer of 1992, tissue samples were collected from the 496 "P2" broodstock, all fish were measured and were identified with an individual PIT tag. The data was collected 6-8 months before the actually spawning to allow time to devise a mating strategy.

In preparation for the up-coming spawning season, all the "priority" fish (fish that seemed to be maturing) were scored with the locus Omy 77 and a second probe (Pupupy) for as many individuals as possible. All of the data was entered into a database. Potential crosses were then evaluated based on two parameters: 1) the group identity of the fish (RP, NC or AD) and 2) sharing of alleles detected by the two (or one in some case) DNA probes. Crossing within groups was discouraged as there was a more likely chance of common parentage. Crossing fish with different alleles was encouraged as it was less likely that they were related. Specific information on the scoring scheme will be provided in the Methods section. All possible combinations of all females and all males were evaluated and spawning recommendations were returned to the SPA hatchery for the actual crosses to be performed. The "CRP2Y5" (Crosses of Parent 2 in year 5 of the

program) crosses were performed in late 1992 and early 1993 in 11 spawning groups. The performed crosses were the result of the decisions made by SPA at the time of spawning and did not necessarily correspond to the recommendations that were submitted to them as some of the recommended fish may have died in the interim or were not ready for spawning as was expected. This was identified as one of the risks associated with the program. From the documentation, ~ 260 crosses were actually performed. Genetic information was available for 206 of the crosses. As to the remaining 54 crosses, either a pit tag number for one of the parents did not match with spawning recommendations list or no DNA was available.

In addition, large and small offspring from various spawning groups from the 1992-1993 spawning season (CRP2Y5) were later collected in the summer of 1993. DNA typing profiles were obtained from large and small offspring from spawning groups 1, 3, 5 and 11. The intention here was to assess which "P2" parents gave high proportion of large juveniles, to be able to re-use these superior "P2" parents later on (i.e. in year 6 and beyond). These profiles were added to the multi-generation pedigree as a sub-set of the generation 3 individuals. When the offspring of "CRP2Y5" mating reached the age of maturity and would become the next generation potential broodstock (P3 in 1995-1996 or Y8), the next phase of the genetic improvement program, the high selection phase, was to be implemented. Using their DNA profiles, all putative broodfish could be traced back to their parents, which had been sampled at the beginning of the program.

Year 6 (1993-1994)

Of the original 496 "P2" breeders, only 250 had survived and were used again as broodstock. The main change that took place that year was related to the mating recommendation protocol based on DNA profile information. Changes occurred concerning the loci used and concerning how the genetic information was used to generate mating recommendations. Four more microsatellite loci were developed: Omy 2, Omy 27, Omy 38 and Omy 105 (Herbinger *et al*. 1995, Heath *et al*. 2001). The genetic marker "Pupupy" was discontinued due to technical difficulties that had been encountered in getting this locus to work well in 1992. Omy 27 was used on a few hundred

individuals, mostly from the PILOT project (see below) but was found to be generally not very informative as it exhibited only a few alleles and was discontinued as well. All surviving "P2" broodstock fish were genotyped at the four target loci, Omy 2, Omy 38, Omy 77 and Omy 105 as this would refine considerably the estimation of relatedness between the broodfish and help in controlling inbreeding accumulation. In following years, genotyping with the additional loci was extended to the "P2" and the "P1" fish that had previously died, in order to create a genetic profile database as complete as possible.

The second major change that took place concerned the way genetic information from the loci was processed to arrive at mating recommendations. The preferred solution would have been to determine which broodstock fish might be related by first establishing their parentage through parent-offspring genotype comparisons. Hence two "P2" broodstock fish tracing back to the same parental pair of "P1" parents or to one common "P1" parent would be full sib or half sib. However this approach was not easily applicable here as parental DNA profiles were often not available, either because some parental fish had not been sampled, or because the previous cross information was incomplete or because previous parental fish had not been yet genotyped at all the new loci. In addition, there was generally little lead time between obtaining the ID (pit tag) of the broodstock that may be used and the time when they would have to be used and therefore generating the missing information was not feasible in such a short time. It was thus decided to only use the genotype information of the current potential broodstock fish to estimate on a pair wise basis which pair may be full-sib, half-sib or unrelated. To that effect, an algorithm was developed and a program was written allowing the calculation of the likelihood ratios of a specific pair being Full-sibs versus Unrelated (FS/Unrel) or Half-sibs versus unrelated (HS/Unrel) based on the multi-locus genotype information (Herbinger *et al*. 1995, Herbinger *et al*. 1997). This appears to have been the first developed pair wise relatedness estimator to use with microsatellite genotype information

The surviving "P2" fish were assessed in November of 1993 for suitability for breeding. In practice, the two likelihood ratio scores (FS/Unrel and HS/Unrel) were generated for every potential pair of broodstock fish based on genotype data at the four Omy loci, and a file was sent to the SPA hatchery. At spawning time, the likelihood ratios

of the specific males and females available for mating that day were extracted from the file and mating was designed so that males and females crossed had a low likelihood of being related. Based on these recommendations, the “CRP2Y6” crosses were performed in late 1993 and early 1994 (year 6 of the program) in fourteen groups. As with the CRP2Y5” crosses, the crosses file “CRP2Y6” is the documentation of the result of the decisions taken by SPA based on the spawning recommendations from MGPL.

In the spring of 1994, the samples collected for the “PILOT” project were also analyzed using the four PCR-based genetic markers. The "PILOT" project took place in the SPA hatchery from February 1990 to February 1991. In this experiment, 4100 offspring from 100 parental couples of the “10 X 10” crosses were reared communally for 1 year. The 443 largest and 430 smallest fish were then measured and sampled. These large and small juvenile fish were therefore from the “P2” generation and were sibs of the “P2” AD parents used to generate crosses in year 5, 6 and 7. Although these large and small juveniles had been sampled much earlier in year 3, this experiment had not been yet analyzed due to the absence of an appropriate DNA typing system. This was now possible with the four available microsatellite markers. In this specific case, DNA profiles were available from all parental fish (10 “P1” males and 10 “P1” females) and from all 874 offspring. It was feasible in theory to determine the pedigree of the juvenile fish through the use of Mendelian laws to compare the offspring genotypes to the putative parental genotypes. A computer program was developed allowing matching of offspring profiles with putative parental profiles. It became possible to reconstruct the pedigree of each selected fish to a large extent, using enough different probes until only one male and one female parent could have given the pattern of alleles observed in that particular offspring. This appears to have been one of the first exclusion-based parentage programs using genotype information. Approximately 91% of the 874 offspring were matched to 1 or 2 parental couples of the 100 possible parental pairs (Herbinger *et al*. 1995). This computer program was seen as a central part in achieving the SPA objectives of being able to determine the family make up of selected broodstock fish in a “walk-back” selection scheme (Doyle and Herbinger, 1994) and to determine parentage of the large and small offspring using only DNA profiles. A high selective pressure could be applied after one

generation and the parents producing the best progeny could be identified for re-spawning in subsequent years to obtain immediate genetic gains. The principle underlying the "walk-back" selection scheme was presented at the Aquaculture Association of Canada meeting in early June 1994. Using the data collected on the 874 largest and smallest fish of the PILOT project as well as some computer simulated data, the feasibility and efficiency of the "walk-back" approach was tested and presented at the International Symposium on Genetics in Aquaculture that was held in late June 1994 in Halifax, Nova Scotia (Doyle and Herbinger, 1994; Herbinger *et al*. 1995)

The final event that year relevant to the present project was the following. In the winter of 1993-94, rainbow trout originating from the 1991-92 spawning season (Year 4) were overwintering in a cage in the Bras d'Or lake. Many fish died during a severe "superchill" episode but some fish survived. Tissue samples were immediately collected from one hundred fish that died (mortalities). Tissue samples were also collected from 95 survivors later in November 1994. The intent was to see whether there were family differences in survival rate, hence to see if selection for cold resistance could be meshed in the SPA program, using the survivors as broodstock fish (Herbinger, 1994). Microsatellite genotypes of the mortalities and survivors were generated, from which 24 among the mortality fish were actually identified to be salmon. The data for the salmon was not included in the multi-generation data base. Using this information turned out to be challenging as these fish came from unknown matings which were not originally part of the breeding program, and no information was available on their parents (see section above, Year 3 and 4). The potential family structure among the dead and surviving fish will be evaluated as a part of this project.

Year 7 (1994-1995)

For this next spawning season, 1994/1995, the remaining "P2" broodstock fish from the 1992-1993 spawning that were still alive were used (n=150) as well as the "new" superchill survivors (n=95). All potential broodfish had been previously genotyped. As in the previous year, a file containing likelihood ratio scores for every potential pair of broodstock fish was sent to the SPA hatchery. At spawning time, these

likelihood ratios were used to ensure that mating took place between males and females that had a low likelihood of being related. The crosses (CRP2Y7) were performed in late 1994 and early 1995 in 12 spawning groups. No subsequent samples of offspring were taken as this was the last time the "P2" parents were used.

2.2.3 Generation 3 and Beyond

The objective for SPA, from the beginning of the project, was to produce its first generation of highly selected fish during the Fall 1995 - Winter 1996, in year 8. This was what this program had been building toward for the past 7 years.

Year 8 (1995-1996) and Year 9 (1996 -1997)

The initial strategy for the real scale high intensity selection scheme was to effect a "walk-back" type selection among best performing 3 year-old offspring from the 1992-1993 spawning (Year 5). The pedigree would be established by matching offspring to parents using existing spawning records and DNA profiles of both offspring and parents. DNA profiles were developed for these "P3" individuals (n=309) that were selected to be spawned in year 8 – designated Parents 3Y8. The origin of these fish will be evaluated as part of the scope of this project.

This operation was replicated the following year to identify best performer 3 year-old offspring from the 1993-1994 spawning (Year 6). In November 1995, 220 steelhead trout selected as the largest among several thousand fish, were brought from another site in Shelburne, NS to the SPA hatchery also to be used as next generation "P3" broodstock. The Shelburne site did raise SPA stock in their sea cages and other stock as well. The sampled fish were supposed to have originated from the 93/94 spawning. In December 1995, all fish were pit tagged and samples were collected. Unfortunately, many fish from Shelburne did not adapt well to their new environment and only 58 of the 220 survived the transfer. In addition, about 90-100 fish from the SPA/Bras d'Or sea cage site were also repatriated in the hatchery. They did not transfer locations well either and only 66 survived. To complement the breeders for this season, some offspring from the 1992-1993 spawning (4 year olds) were added to the pool. The 4 year olds were not the prime

choice to be selected breeders because they had matured earlier. DNA profiles were developed for this second set of "P3" individuals (n=122) that were selected to be spawned in year 9 - designated Parents 3Y9. These fish will also be evaluated as part of the scope of this project.

Despite the fact that fewer selected broodfish survived than expected, the SPA hatchery used the provided information to generate optimal crosses in Year 8 and 9. They applied for funding to continue the breeding program and to compare the production performance of the produced progenies to the best available commercial strain at the time (imported from the State of Washington). Unfortunately, they did not manage to obtain such funding. They used again the mating information for year 10, but the breeding program did not continue beyond this, as this small operation did not have the financial means to operate such a program.

Figures 1 and 2 summarize graphically all of this information, with Figure 1 focusing on the timeline and Figure 2 on the succession of generations. Table 3 provides summary parentage information for the various groups of fish which were analyzed and constituted the database.

Table 3: Offspring groups and associated spawning sessions.

Crosses	Spawning dates	Spawning groups	Spawning pools	Candidate Mothers	Candidate Fathers	Associated Offspring group(s)
CRP1Y2	Dec. 89 to Jan. 90	9	25	40	17	Right Pectoral clip (RP)
14 x ?	Dec. 1989	unknown	unknown	14	all possible (21)	No Clip (NC)
10 x 10	Jan. 1990	1	n/a	10	10	Adipose clip (AD)
CRP2Y5	Dec. 1992	7	1	17	8	Group 1 Y5
	Jan. 1993	10	1	39	11	Group 3 Y5
	Jan. 1993	4	1	6	4	Group 5 Y5
	Feb. 1993	1	1	7	1 (sex reversed)	Group 11 Y5
	Dec. 92 to Mar. 93	11	57	126	24	Parents 3 Y8
						Parents 3 Y9?
CRP2Y6	Nov. 93 to Mar. 94	14	84	116	56	Parents 3 Y9
CRP2Y7	Nov. 94 to Mar. 95	12	81	89	22	none

Figure 1: SPA genetic improvement project time line by year

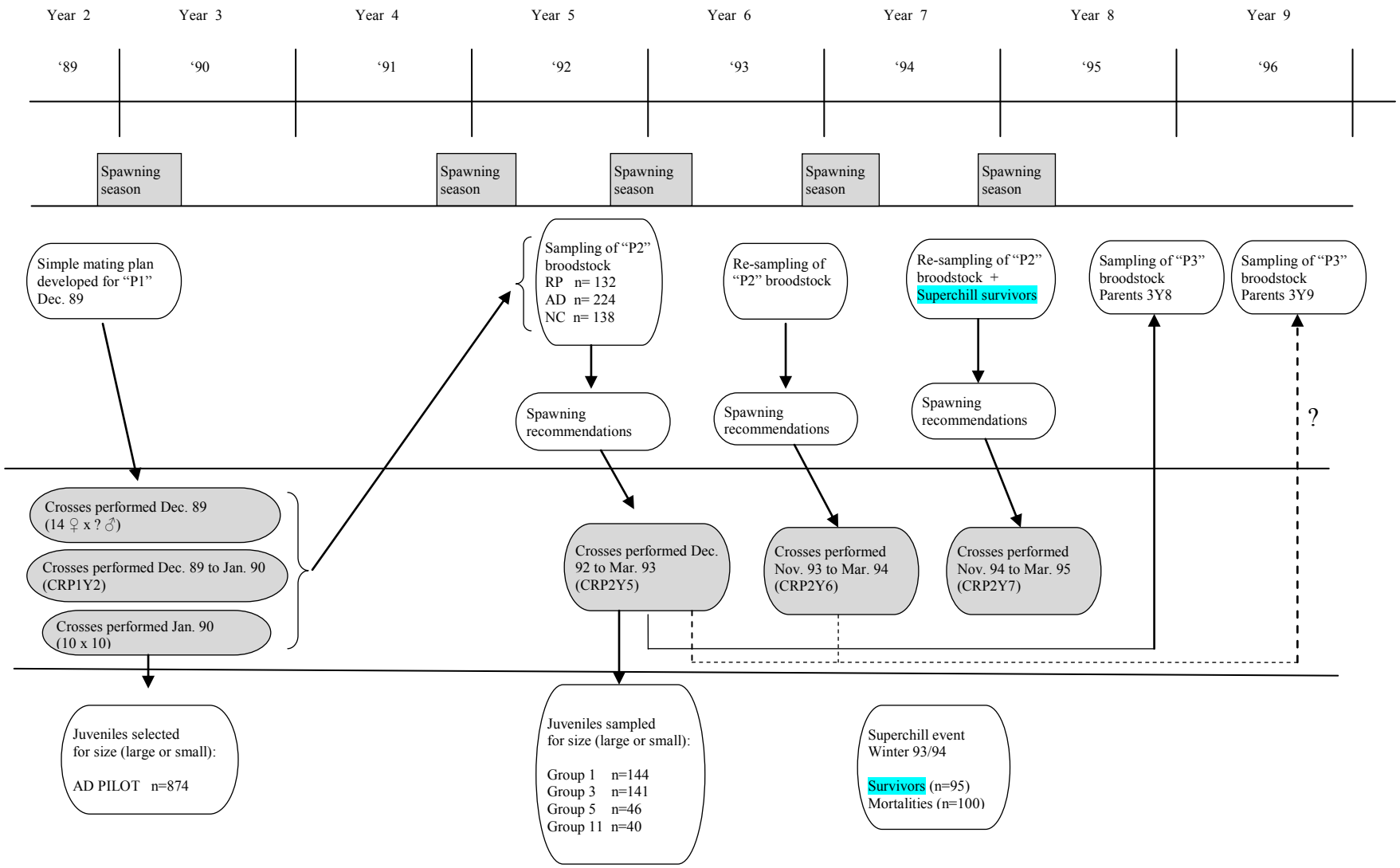
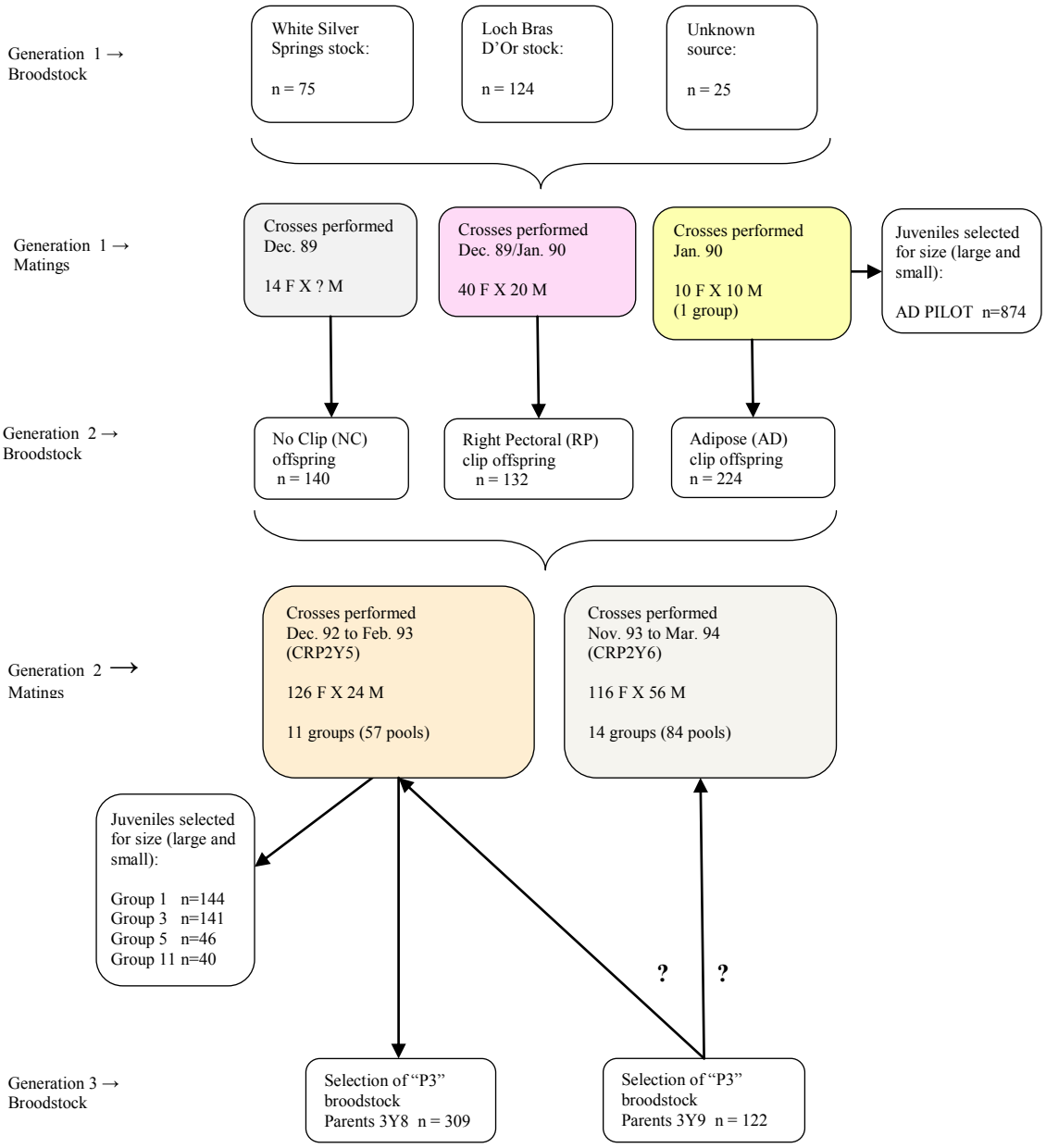


Figure 2: SPA genetic improvement selective breeding program flow diagram.



CHAPTER 3: METHODS

3.1 Assessment of Parentage and Pedigree Reconstruction

3.1.1 Parentage Analysis Based on Genetic Markers and Mating Information.

Since the advent of PCR, it has become commonplace in many fields of scientific study to use microsatellite loci to infer genealogical relations. In the case of paternity (or maternity) estimates in humans for example, where one of the two parents (usually the mother) is known, the challenge is typically to determine the most probable other parent to complete the trio. These approaches are also relevant to the case of aquaculture genetic programs, as exact genetic/family relationships between individuals are often unknown. Fish fry are so small that it is impossible to tag them right away at birth, which create problems to recognize family units. One possibility is to rear families in separate tanks until the animals are large enough to tag, but this may lead to confounding differences between family effects and environmental tank effect (Herbinger *et al*. 1999). Alternatively, the fish may be reared communally and their family identity determined from matching offspring genotypes to putative parental genotypes (Herbinger *et al*., 1995; Fishback *et al*., 1999; Perez-Enriquez *et al*., 1999; Norris *et al*., 2000). This is a daunting task. There are many potential sources of error that could generally result in a 'mismatch', or lack of ability of allocating a fish to any of its putative parental pairs:

1) the group of offspring is not the one that we think it is;

2) some fish may have jumped from another tank;

3) the parental crosses information was not complete (e.g. a fish without tag was used) or not completely accurate (e.g. some males and females were not crossed as noted);

4) there could be errors in the DNA profile of some of the parents;

5) there could be errors in the DNA profile of some of the offspring.

For this data set, the task of assigning parentage was complicated because neither parent was known with absolute certainty. What was known was that a particular group of offspring were derived from a particular spawning group in a semi-controlled manner based on a breeding scheme. The link between offspring and possible parental combination was based entirely on the documentation provided describing each mating design and its associated offspring group.

Several approaches exist for calculating parentage and pedigree determination: exclusion of genetically incompatible individuals from parentage, categorical allocation, fractional allocation and parental reconstruction (Neff, 2001; Jones and Arden, 2003). Each approach has its own advantages, disadvantages and range of applications. Each approach is vulnerable to the presence of genotyping errors, null alleles and/or mutations. Genotyping errors are avoidable only to some degree, and may occur at any step of the DNA profile development process (i.e. sampling, DNA extraction, amplification, allele size scoring and/or data analysis). Genotyping errors can greatly influence allele frequency estimates and consequently the ability to distinguish between individuals (O'Reilly and Wright, 1995; Bonin *et al*., 2004). Alleles that lack a proper flanking PCR primer binding site will not amplify during PCR and can result in interpreting a null-heterozygote genotype at some locus (i.e. heterozygote genotype that appears to be homozygote because only one allele has amplified) as a homozygote genotype at this locus. The presence of null alleles may have ramifications in parentage analysis by causing false exclusions (Jones and Arden, 2003), and blurring the estimation of the number of individuals which are homozygotes or heterozygotes, with an associated loss of accuracy of the (coarse/population-level) estimation of inbreeding.

The simplest approach to parental allocation is based on elimination by exclusion. Parentage is determined by comparing the alleles at a given locus from each offspring, with all four possible sets of alleles from the putative parents. Candidate parental pairs with alleles incompatible with those of the offspring, at one or more loci, are excluded as possible parents. The object of exclusion testing is to eliminate as large a portion of the population as possible as putative parents, leaving in the end, categorical assignment to only one possible parental pair. However, sometimes more than one parental pair cannot

be excluded as putative parents and there is no justification to exclude one set of possible parents over the other. In this situation (incomplete exclusion), both/all non-excluded parental pairs are assigned, but this stretches the definition of categorical assignment. This scenario can be loosely considered "nearly" categorical or "quasi" categorical. Another weakness of a complete exclusion approach is that genotyping errors, null alleles and mutations may contribute to false exclusions, which may result in no parental pair being allocated to a particular offspring. A likelihood (or likelihood ratio) approach can sometimes resolve parentage when exclusion cannot, by being more tolerant of genotyping errors and limiting the impact of false exclusion. Likelihood or likelihood ratio approaches can also help in the case of incomplete exclusion by providing a way to rank the various non-excluded parental pairs and selecting the most likely one.

Categorical and fractional allocation methods both use a likelihood-based approach to select the most likely parent or parents from a pool of non-excluded parents (Jones and Arden, 2003). True categorical allocation assigns one offspring to exactly one pair of parents and uses a likelihood-based approach to identify the most likely parent from a pool of non-excluded parents (Marshall, 1998; Jones and Arden, 2003). On the other hand, fractional allocation also uses a likelihood-based approach in assigning a fraction of a parental pair to one offspring (Devlin *et al*., 1988). One of the concerns with fractional allocation from a biological perspective is that assigning a fraction of parent to an offspring is not realistic because an offspring can ultimately only come from one set of parents, however, for populations where matings are not certain, or when the genetic information is limited (due to a low number of markers/high rate of genotyping errors) then fractional allocation is perhaps the most robust representation of parentage for the pedigree (Perez-Enciso and Fernando, 1992; Dodds *et al*., 2005).

One of the main goals of this project was to assess these pedigree approaches and their associated challenges, strengths and weaknesses by conducting a retrospective analysis of a real, three generation pedigree from the SPA hatchery using a limited number of genetic markers available at the time (early 1990s). This retrospective analysis would assess, under different pedigree reconstruction approaches, whether the semi-selective, on-farm breeding scheme implemented at the time was successful in limiting

the level of inbreeding increase and in identifying possibly superior broodstock. Molecular genetic data at as few as four loci was used to infer relatedness among individuals and between generations in the reconstruction of the full pedigree. Parentage and pedigree reconstruction was estimated, for the quasi-categorical (exclusion-based and LOD-based) pedigrees via the program CERVUS 3.0 (Marshall *et al*., 1998; Kalinowski *et al*., 2007), and for the fractional pedigree via a software (PIPEDIGREE), developed for this project (Fullsack and Herbinger, unpublished).

CERVUS 3.0 uses a "maximum likelihood" approach to statistically distinguish a non-excluded candidate parental pair and identify the candidate parental pair that is most likely to be the true one, based on co-dominant markers. Likelihood ratios can be used to contrast several types of conditional hypotheses, and are formed to test single parent-offspring allocation, and parental pair-offspring (or joint) allocation. When one is interested in testing single parents –male or female- for parentage on a giving offspring, the hypothesis that the candidate parent is the true parent (given the observed genotypes) is compared to an alternative hypothesis that it is not the true parent (among those parental candidates that have not already been excluded by some prior knowledge of the mating design, or some prior analysis). A similar method is used for trios – i.e. a set of three individuals, offspring-male-female, tested for the joint parentage relation. The likelihood of each hypothesis is calculated from the probability of obtaining the observed genotypes.

Likelihood Ratio: $\frac{P(E/H1)}{P(E/H2)}$

Where: E = DNA evidence based on allelic values (multilocus genotype) and allele frequencies

H_1 (Hypothesis 1) = The candidate parent/parental pair is the true parent/parental pair.

H_2 (Hypothesis 2) = The individuals involved in the comparison are unrelated.

A number of assumptions are made about the genetic marker data: 1) the species is diploid, 2) the genetic markers are autosomal, co-dominant and are inherited independently, 3) there is no linkage between loci, and 4) all of the Mendelian laws of

heredity/allele transmission apply. The CERVUS program compares, at each locus, the set of alleles for each offspring, with all four possible sets of alleles from each of the potential parental pairs. It evaluates whether there is (or not) a mismatch between the offspring and the male candidate parent, the offspring and the female candidate parent, and similarly for this trio of individuals. CERVUS can accommodate genotyping errors. If there were no errors in the genotypes then a mismatch between offspring and candidate parent(s) would imply non-relationship or an exclusion of parentage. However, if any of the genotypes contain errors, a mismatch may be due to non-relationship or may be due to a scoring/typing error in either the offspring or either of the parents. An advantage of CERVUS is that it uses likelihood equations that take into account typing error (including possible mutation). Details on the calculation of pairwise (parent-offspring) and trio likelihoods can be found in Kalinowski *et al*., 2007.

3.2 Assemblage of Pedigrees

Three pedigrees (QCEP, QCLODP, FP) were assembled, based on the preliminary parentage analysis described in the previous section, one generation at a time:

1. QCEP: Exclusion based (quasi- categorical) allocation approach with proportional probabilities equally distributed over all non-excluded parental pairs.
2. QCLODP: Exclusion based (quasi-categorical) allocation approach with proportional probabilities derived from an overall likelihood score (LOD) distributed over all non-excluded parental pairs.
3. FP: Fractional allocation approach.

Each pedigree is represented as a set of 'probabilized' trios (PT) of individuals. Each trio is composed of three individuals, the first being an offspring, the second and the third being the male and female (possibly fractional) parents of the offspring. More details on this pedigree representation will be provided in the section on fractional pedigrees. "Fractional parent" means that there is some "non-zero" probability that the individual in

question is actually a member of such a trio, where each trio expresses the joint parentage of the parental pair over the offspring. Only trios compatible with the mating design were listed in each pedigree reconstruction. In this "PT" representation, several trios may exist for the same offspring individual (in first position). This corresponds to a situation where several parental pairs cannot be excluded for this specific offspring.

Each of these pedigrees was extended, in a second step, by the addition of a hypothetical reconstructed layer "P0" of phantom parents of "P1" individuals to see how inclusion of this information would or would not influence the relationship matrix, thus creating a total of six pedigrees for evaluation.

3.2.1 QCEP pedigree: Exclusion Based (Quasi-Categorical) Allocation Method with "Equi" Probability among Multiple Non-Excluded Parental Pairs.

3.2.1.1 The Process

The determination of parentage for each offspring was a multi-step process. The first step was to determine, from the documentation provided, which offspring group was associated with which set of crosses by generation. The second step was to derive credible trios for each offspring, according to the mating design, using CERVUS to derive all non-excluded parents. Parental allocation for each offspring group in the 3 generation pedigree had its own unique set of challenges but the decision making process for evaluating the assignment of parentage was the same for each group. The decision making process is outlined in Figure 3. Depending on the complexity of the mating scheme a number of different allocation scenarios arose. Some individuals were allocated to only one parental pair, some to more than one parental pair from the same spawning group, some allocated to more than one parental pair from different spawning groups and some could not be assigned to any parental pair. The last part of the process was to assign a confidence value, or statistical significance, to each credible trio. The following decision making process was followed for each CERVUS run:

1. Prepare 4 data files according to offspring group and its outlined associated mating design defining: offspring IDs file, mother IDs file, father IDs file and genotypes for all individuals including all offspring and putative parents
2. Perform CERVUS run using the following criteria:
   - allele frequency estimates (includes the genotypes of parents)
   - simulation run on same genetic data defining: 100% of mothers sampled, 100% of fathers sampled, 100% of loci typed, 10% error rate and the minimum number of loci typed = 3.
3. Evaluate the trios defined by CERVUS where there are no ("0") trio loci mismatches. Filter the CERVUS output file for "0" trio loci mismatching. If no mismatches at any locus then the trio is accepted.
   Note: only trios with "0" mismatches were allowed for individuals with information at only 3 loci.
4. Add trios where there are no mismatches at any locus to the pedigree.
5. For the remaining offspring for which no trio with zero mismatches had been found, from the CERVUS output file, evaluate the trios allowing a mismatch at one locus. Filter CERVUS output file for "1" trio locus mismatch.
6. Identify locus where there is a mismatch and evaluate each possible trio combination to determine if allowing a 2 - 6 bp change at one of the alleles for either the offspring or one of the parents would allow the trio to work, and in that case accept the trio.
   Note: the correction could apply to the genotype of the offspring or to the genotype of one or the other of the putative parents (unless the parent had previously allocated to several other offspring without correction). If allowing a 2-6 bp change does not provide a credible assignment then the trio is rejected.
7. Add those trios where a correction was accepted to the pedigree.
8. Any remaining offspring that were not assigned a trio after evaluating the "0" and "1" trio loci mismatches (allowing for scoring errors and/or

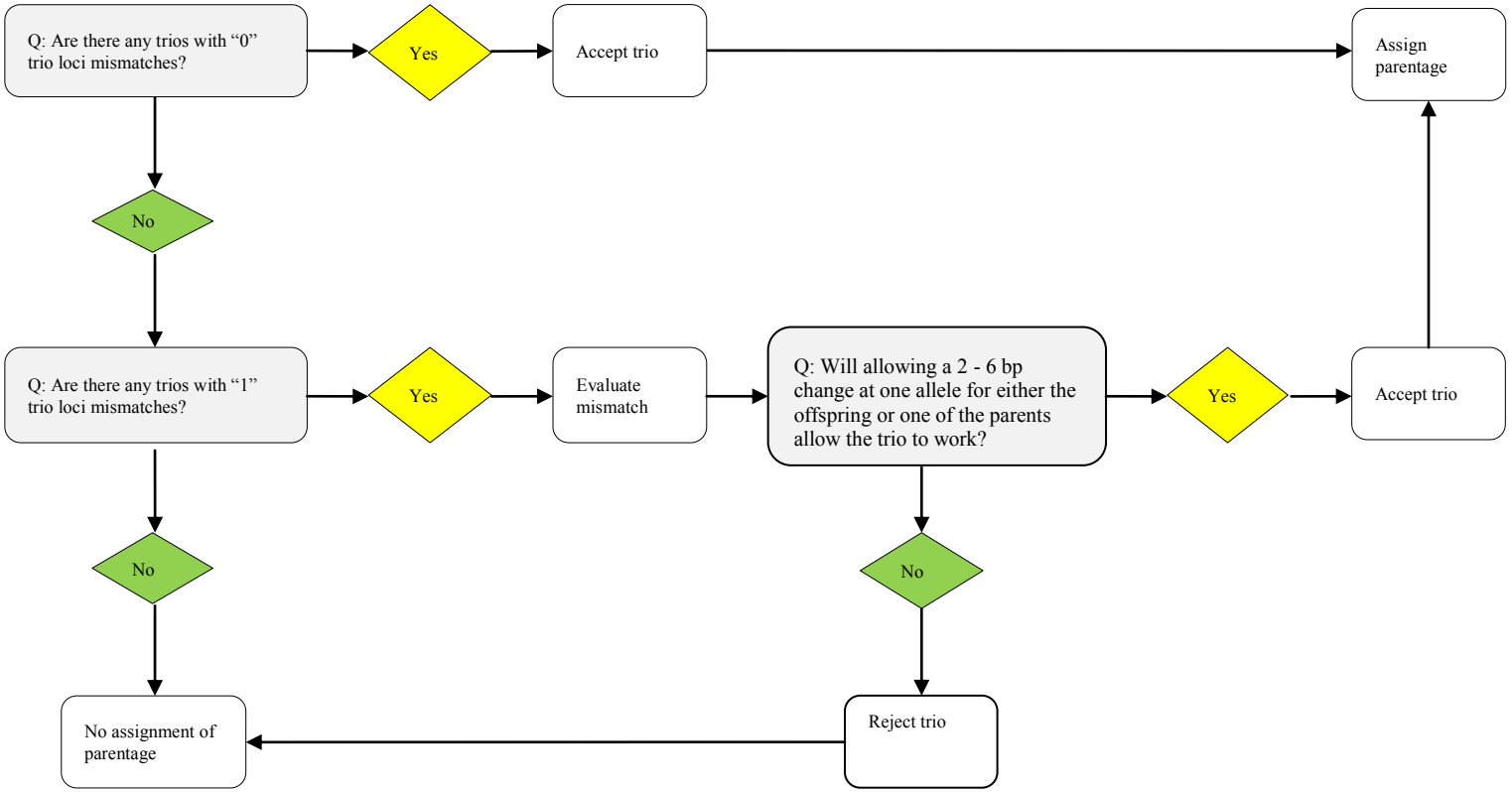
mutations) from the CERVUS output files were designated as “no parental allocation” and were treated as founders in the pedigree.

9. Assign a confidence value, or a statistical significance, to the trio: when only one parental pair was genetically compatible (i.e. when one offspring was only found in one trio), proportional probability = 1. If multiple pairs were genetically compatible (i.e. when one offspring was found in several trios), then a proportional contribution was estimated for each compatible pair simply based on the number of non-excluded parental pairs for that offspring, proportional probability = 1/n. In other words in that last case proportional probabilities were allocated equally to all possible trios. If there are no parental pairs that could be compatible with an offspring genotype, proportional probability = 0.
10. Offspring with information at less than 3 loci were not evaluated; however they were included in the pedigree with no parental pair assignment, proportional probability = 0.

Steps 3 to 8 are graphically depicted in Figure 3. The overall strategy for the QCEP was thus to try to arrive at a pedigree as categorical as possible using exclusion rules, supplemented by simple extensions 1) to account for multiple non-excluded parental pairs and 2) using a simple one locus “repair” rule to try to salvage offspring that may have been falsely excluded.

Figure 3: Decision making tree illustrating the exclusion based parental allocation assessment process.

For each CERVUS run: evaluate trio [Individual_Mother_Father]

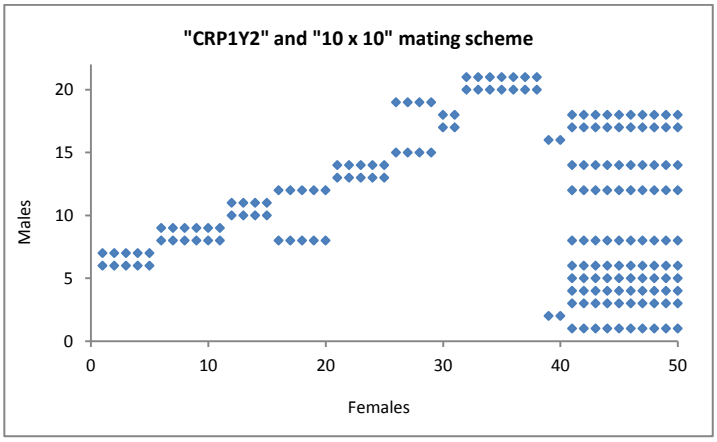


3.2.1.2 Challenges using CERVUS to Determine Trios.

Each offspring group in the 3 generation pedigree had its own unique set of challenges in using CERVUS for parental allocation. The simplest group to assess was the AD offspring group that was associated with the "10 x 10" factorial crosses performed in January of 1990. This very large group of 1098 individuals was processed as one CERVUS run, as any mother/father combination was possible. Similarly, the NC offspring group was associated with 14 possible female parents, but the number and identity of the males used was not specifically documented. For this reason, all 21 males that appeared to have been available in the year 2 crosses were used and this group was assessed as a "14 x 21" factorial cross allowing all possible combinations and processed as one run.

Things were a bit more challenging for the RP offspring group. This group was associated to parents used in crosses performed in year 2 (CRP1Y2) where the spawning groups were defined by smaller pools (acting as partially disconnected diallels). The spawning took place in nine groups where the eggs from two to six selected females were fertilized with the milt from two selected fathers at a time, further defining specific parental combinations. In this situation, if one large CERVUS run had been performed, incorporating all mothers and fathers, as was done for the AD and NC groups, some parental combinations might have been observed that were not actually possible and false inclusions could have occurred. To avoid this, evaluation of this group was done as nine CERVUS runs restricting possible parental combinations to the ones that could have really occurred. However, this meant that it was necessary to compile data from many different runs for the more complex groups making this exclusion based approach to the assessment of parentage very time consuming and labour intensive. To further complicate things, some males were used in more than one spawning group and in addition, some males (6) were also used in the "10 x 10" crosses (AD) compounding the interpretation (Figure 4).

Figure 4: The “CRP1Y2” (RP) and “10 x 10” (AD) mating scheme (# females x # males).



Parentage for each of the offspring groups from generation 3 (Group 1, Group 3 and Group 5 from Year 5 and the Parents 3Y8 offspring groups) were assessed in similar fashion to that described for the “RP” offspring group in that many runs were required. As explained above with the “P1” crosses, some males and females were used in more than one spawning group further complicating the interpretation of the results.

3.2.1.3 Challenges to the Assessment of Parentage

One of the biggest challenges with this data set was the acknowledgement of the possibility of genotype errors in the data, mainly due to scoring and/or transcription errors. Always keeping the possibility of genotyping errors in mind as the primary reason for a mismatch between offspring and parent, in instances where there were no “0” trio loci mismatches from the CERVUS output file, it was important to take the next step and evaluate each CERVUS run where a mismatch at one locus could be repaired using the simple rule described in section 3.2.1.1 and illustrated in Figure 3. For this large data set, there were many instances where it was necessary to re-evaluate the trios allowing a mismatch at one locus. The correction was applicable either to the genotype of the offspring or to the genotype of one or the other of the putative parents. This however was not allowed if that parent had been allocated to several other offspring without correction.

Table 4 illustrates the evaluation process for assignment of parentage for an RP offspring (7F7E501C5F) where there were no "0" trio loci mismatches for this individual after evaluation of all CERVUS output files according to the "CRP1Y2" mating design. It was then necessary to evaluate all instances/trios where a "1" trio locus mismatch was observed from any of the nine CERVUS runs. After reviewing each of the output files it was determined that there were 4 putative parental pairs with a mismatch at one locus. Table 4 also illustrates how allowing for a genotyping error at one locus could permit parental allocation for an individual. In the end, two sets of parents (s2198 x s2212 from spawning group 2 and s2237 x s2243 from spawning group 4) were assigned to fish 7F7E501C5F after allowing a 2 bp change at one locus. Based on the set of rules described in section 3.2.1.1, either set of parents was considered equally likely to be the true parents and a probability of 0.5 was assigned to each. In this manner, the exclusion based approach underlying the construction of QCEP can loosely be referred to as a nearly or quasi-categorical approach, when more than one parental pair cannot be excluded as possible parents. For the purpose of this project, the exclusion based approach described here and the one described in next section will further be referred to as the "quasi-categorical" methods.

In general, exclusion based methods for parentage analyses tend to assume pristine, error-free data sets which are largely unrealistic. As was already illustrated, this was not the case with this complex data set making this exclusion based (quasi-categorical) approach very time consuming and labour intensive.

Table 4: Assignment of parentage allowing a mismatch at one locus.

Possible parents	Pit tag/ID	Omy 2		Omy 38		Omy 77		Omy 105		Comments
	7F7E501C5F	112	164	98	118	98	104	177	189	1 change at Omy 2 of 2 bp – from 112 to 110
Group 2										
? Mother:	s2198	110	146	98	118	98	104	187	189	Trio allowed (with 2 bp change at Omy 2 for offspring)
? Father:	s2212	136	164	110	118	98	98	177	187	
Group 4										
? Mother:	s2225	110	144	98	118	96	104	177	189	Trio allowed (with 2 bp change at Omy 2 for offspring)
? Father:	s2212	136	164	110	118	98	98	177	187	
Group 5										
? Mother:	s2234	110	144	98	118	96	104	187	189	Trio not allowed – based on mismatch at Omy 2 -> no allele 164 in either parent
? Father:	s2243	110	134	118	130	98	104	177	199	
Group 8										
? Mother:	s2010	136	146	98	118	98	104	189	205	Trio not allowed – based on mismatch at Omy 2
? Father:	s2329	136	170	118	120	98	98	177	199	

3.2.2 QCLODP Pedigree: Exclusion Based (Quasi-Categorical) Method with "Weighted" Distribution of Probabilities among Multiple Non-Excluded Parental Pairs

The second pedigree created for this project (QCLODP pedigree) consisted of exactly the same trios that were derived for the first pedigree (QCEP). The only difference concerned the probability associated with trios, when there was more than one trio associated with a given offspring (i.e. when there were multiple non-excluded parental pairs). Trios with a probability of zero (i.e. offspring associated with no parental pairs) or with a probability of 1 (offspring categorically assigned to exactly one parental pair) were the same under QCEP or QCLODP. For the multiple trios associated with one offspring, a probability was now calculated for each trio derived from the CERVUS trio LOD score (hence the name QCLODP), instead of simply been equally allocated among all possible trios as under QCEP.

The LOD score is the natural log (log to base e) of the overall likelihood ratio of the trio. Again, this ratio divides the probability that the 3 individuals are linked by joint parentage of the male (second individual) and female (third individual) over the offspring (first individual) by the probability that these three individuals are unrelated. The following conversion formula was used to normalize the LOD score and estimate the trio probability:

CERVUS Trio LOD score = X_1

$$e^{(LOD)} = e^{(X1)} \rightarrow \text{trio probability} = \frac{e^{(X1)}}{\sum (e^{(X1)} + e^{(X2)} + ...)}$$

This conversion is actually equivalent to the calculation of probabilities done under the fractional pedigree (see next section).

The overall strategy for the QCLODP was thus to try to arrive at a pedigree as categorical as possible using exclusion rules and using a simple one locus "repair" rule to try to "salvage" offspring that may have been falsely excluded, just as was the case for QCEP. The main difference was in accounting for multiple non-excluded parental pairs, where a higher probability of parentage would now be allocated to a parental pair that appeared more likely to be the true one among all non-excluded parental pairs.

3.2.3 FP Pedigree Based on Fractional Allocation

The last pedigree was based on an extension of the idea embodied in QCLODP for the multiple non-excluded parents. It would now be applied to all possible parental pairs and exclusion rules would not be used any longer. As for the other pedigrees, the goal was to calculate a list of 'probabilized trios' (PT) of individuals. Fractional parentage corresponds to some non-zero probability that a parental individual, used in one of the crosses present in the mating design, is actually a member of such a trio. Each trio expresses the joint parentage of the parental pair over the offspring. Only trios compatible with the mating design were listed in each pedigree reconstruction. In the 'PT'

representation, several trios may exist for the same offspring (individual in first position). This corresponds to a situation where an ambiguity exists in the parentage allocation.

The program PIPEDIGREE (Fullsack and Herbinger, unpublished) was used to calculate the PT list of FP. This program allows the user to specify the mating design information as a set of admissible crosses per spawning group, the genotype database of multilocus genotypes, and various heuristic parameters such as the minimum number of valid (non-missing) alleles, the maximum numbers of trios and parent-offspring alleles mismatches, and the genotyping error rates. An error rate of 10% was used, matching the error rate used in the CERVUS runs for parentage allocation. The method of calculation of parentage likelihoods is similar to the one used in CERVUS. Various sets of parameters were tried and a set was chosen which corresponded to the limit of keeping nearly all the trios with positive LOD, in order to contrast the quasi-categorical pedigree approach (QCEP and QCLODP) to a quasi-all-probable-trios approach (FP).

It is important to note that an offspring that had been categorically assigned to one specific parental pair under QCEP/QCLODP would most often be assigned to many possible parental pairs under FP. Even though that sole parental pair might have been the only possible one following exclusion rules, the inherent error model in the FP approach would permit other potential parental pairs to be included in the list with non-zero probabilities. The unique parental pair identified under the quasi-categorical approach was most often (but not always) the pair with the highest probability under FP.

3.3 Applications of the Pedigree Reconstructions to the Retrospective Analysis of the SPA Breeding Program.

3.3.1 Estimation of Gene Diversity

Although gene diversity estimation does not require parentage analysis, it was processed as a byproduct of my analysis, with the aim of quantifying the genetic diversity, and potential loss of diversity in later generations. Genetic information for 2396 individuals from three generations of Rainbow trout were genotyped at four microsatellite loci; Omy 2, Omy 38, Omy 77 and Omy 105. Summary statistics were estimated including: allele frequency analysis observed (H_0) and expected (H_e) heterozygosity, deviations from Hardy-Weinberg equilibrium and null allele frequency estimates using CERVUS 3.0 (Marshall *et al*., 1998; Kalinowski *et al*., 2007).

3.3.2 Estimation of Reproductive Success

Reproductive success was estimated for each parent for each of the three pedigrees (i.e. the three lists of trios) as the sum of the number of offspring x the probability of that parent for that offspring. To allow a comparison of estimated reproductive success under the different types of pedigrees for the groups which had large differences in the number of allocated progeny, the relative success of the different parents was also estimated in addition to the absolute success. The relative success was estimated as sum of the number of offspring x the probability for that parent, normalized over the number of offspring assigned to trios per pedigree. The Chi-square Goodness-of-Fit test (χ^2) was used to test for differences in reproductive success for each parent in different progeny groups, except when expected cell count were below 5. Bonferroni corrected significance levels were used where required. In the specific case of the AD PILOT and AD broodstock group where reproductive success of the same parents could be evaluated in the two different offspring groups, Spearman rank correlation coefficients

(r) were estimated to test for differences in ranking of parental reproductive success. All test were performed with Minitab 16 (Minitab Inc, State College PA USA)

3.3.3 Estimation of Individual and Average Inbreeding

One of the rationale for using DNA markers in the original SPA breeding program was to avoid crossing/mating genetically related individuals, that is avoid adding inbreeding to the initial (unknown) inbreeding of the founders of the pedigree. The other aim was to help with the identification of superior broodstock for selection. Inbreeding can accumulate over time by accumulation of loops in the pedigree. This accumulation can be quantified either at the individual level, or, by averaging, at higher levels (group of individuals, generation, population. The inbreeding coefficient of an animal is the probability that the two genes at a locus in an individual are identical-by-descent. Two alleles are identical by descent when they are copies of the same allele transmitted to the individual carrying these alleles by a common ancestor. The inbreeding coefficient describes the relationship of genes within an individual (Falconer and MacKay, 1996). Wright (1922) proposed one way to calculate the inbreeding coefficient of an animal, the so-called path method. However, this method is very laborious. Individual inbreeding coefficients can also be derived from the numerator relationship matrix "A". This is a square matrix the size of the number of individuals. The inbreeding coefficients of any individual (i) can be computed from the diagonal elements of the matrix by:

$f_{(i)} = A_{(i,i)} - 1$

For example, if $A_{(3,3)}=1$, then $f_{(3)}=(1 - 1)=0$. The inbreeding coefficient of individual 3 is 0 so individual 3 is not inbred. Whereas, if $A_{(17,17)}=1.015$ then $f_{(17)} = (1.015 - 1) = 0.015$. The inbreeding coefficient of individual 17 is 1.5 per cent

The matrix element $A_{(i,j)}$ for individuals i and j is called the genetic relationship coefficient between these two individuals, and is equal to twice the coancestry between these two individuals. In turn the coancestry between a pair of individuals is equal to the

inbreeding of the progeny that could be produced by that pair. The numerator relationship matrix "A" thus contains information which not only can help calculate individual inbreeding of every individual in the pedigree but can as well predict future inbreeding resulting from any possible cross. In addition this matrix is central to quantitative genetic parameter estimation such as breeding values and heritabilities.

The numerator matrix is typically calculated (see for example Lynch and Walsh 1998) from a knowledge of the categorical pedigree, e.g. through a simple recurrent method which updates the matrix elements of pair of individuals from previously computed values for parents and ancestors of these individuals. For strictly categorical pedigrees, each progeny has at most one parental pair (progeny without parents are either ignored if they are not themselves parents, or treated effectively as founders). The only requirement is that individuals should be numbered in a way such that progenies always appear after their parents. A number of software programs can calculate the "A" matrix. In the case of a fractional pedigree, this classical approach can be extended to a list of "probabilized" trios where again, every offspring has to come after their parents, but where now a specific progeny may be represented in several trios (i.e assigned to several possible parental pairs), each associated with different probabilities. An average relationship matrix (or Abar or $\bar{A}$) can now be computed, integrating over all possible pedigrees, weighted by their probabilities. This idea was introduced by Henderson (1988) in the case of uncertain paternity and extended to uncertain (joint) parentage by Perez-Enciso and Fernando (1992). This matrix was calculated with the software PIPEDIGREE (Fullsack and Herbinger, unpubl.), which takes as input any pedigree in PT format, and computes the average relationship matrix and individual and average inbreeding values. Each of the three estimated pedigrees was used as an input for such computations, and compared to evaluate whether the SPA genetic improvement project was successful in limiting the increase in the level of inbreeding over the three generations.

The first three scenarios corresponded to the two "nearly categorical" pedigrees and the fractional pedigree previously described (QCEP, QCLODP, FP). In addition, sibship reconstruction was performed among the very first broodstock (P1) fish in order to assess the impact of transforming these three generation pedigrees into four generation

pedigrees where possibly the founders might have originated from a small pool as is often the case in aquaculture populations. A hypothetical "P0" layer was created using COLONY 2.0 (Wang and Santure, 2009), complemented by PEDIGREE 2.2 (Herbinger, 2006). COLONY uses a maximum likelihood based approach to infer parentage and sib-ship among individuals using multi-locus genotypes. COLONY was used to reconstruct hypothetical parents of "P1" (P0) adding a 4th level to the pedigree to see how/if the addition of "P0" would influence the measure of inbreeding in the pedigree. PEDIGREE uses the family partitioning method of Smith *et al*. (2001) to estimate the number of family groups among the "P1". The Smith *et al*. (2001) method uses a Markov Chain Monte Carlo algorithm approach to identify the most likely configuration of full-sib and/or half-sib family groups within a sample. This list of hypothetical trios involving the real "P1" fish and their hypothetical "P0" parents was then added on top of the three lists of trios. It should be noted that list of trio involving the "P1" and their hypothesized "P0" parents was perfectly categorical as there is at present no method for probabilizing sibship reconstruction. In other words, each "P1" parent was only assigned parentage to one trio with a probability of 1, after creation of one phantom "P0" male and one phantom "P0" female per sibling group. New inbreeding values were then determined for the three new pedigrees and compared to see how (or if) the addition of a fourth level to the pedigrees would influence the inbreeding values observed in generations 2 and 3.

3.3.4 Retrospective Evaluation of the Mating Recommendations that had been Provided for Inbreeding Control.

In order to limit the risk of consanguinous matings, several recommendations were made to SPA on the basis of available genotypes and genetic groups, as to which crosses of two individuals should be avoided. A first, coarse, strategy was developed for the evaluation of a series of "P2" crosses in 1992/1993, where this series of crosses led to the generation 3 fish in this study. This strategy was later refined into a potentially more

discriminanting one used for the crosses in 1993/1994. However the later strategy will not be evaluated as no fish resulting from these 1993/1994 crosses was available. In 1992/1993, all the "priority" fish (fish that seemed to be maturing) were scored with the locus Omy 77 and with a second probe (Pupupy) for as many individuals as possible (Herbinger, 1993a). The classification of the potential crosses was based on two parameters: 1) the group identity of the fish (RP, NC or AD) and 2) sharing of alleles detected by the two (or one in some cases) DNA probes. Crossing within groups was discouraged as there was a more likely chance of common parentage. Crossing fish with different alleles was encouraged as there was a less likely chance that they were related. The crosses were rated as either "OK", "Grey" or "Black" for each locus (left panel, below) and then the information from both loci was combined (right panel, below). A scoring of "OK" for both loci indicated probably non-related fish. "Black" at both loci, indicated probably related fish so the recommendation was that mating between these two fish should be avoided if at all possible. A combined ranking of "OK" was the preferred recommendation while "Grey" was an acceptable second choice when "OK" choices were not possible." The following scoring scheme was used for potential "P2" crosses:

1 Locus:		
	Group	
# Common alleles	Same	Different
0	Grey	OK
1	Black	Grey
2	Black	Black

2 Loci:			
	Locus 1		
Locus 2	OK	Grey	Black
OK	OK	OK	Grey
Grey	OK	Grey	Black
Black	Grey	Black	Black

In 1992/1993, Omy 77 genotype information was available for 470 "P2" fish but Pupupy genotype was only available for 112 fish. The cross ranking information was therefore derived mainly from the fish group origin and first locus marker information. The scores were re-calculated for all the potential "P2" crosses. In order to verify the quality of the recommendations, the scores for all pair of "P2" fish were compared with the level of consanguinity of each pair estimated as the genetic relatedness or double of the coancestry of the pair (that is, the element of the average numerator relationship matrix). The matrix used was that estimated under the fractional pedigree with the "P0" layer added. This analysis was then repeated for the subset list of the actual crosses

realized in 1992/1993, which was of course constrained by the availability of mature fish at spawning time.

3.3.5 Retrospective Evaluation of Individual Genetic Worth

Breeding programs have typically several objectives, which are often conflicting, at least to some degree. The basic tenets are the selection of the most profitable individuals, in terms of their own economic value or of the economic value of their descendants, and the maintenance of genetic diversity and limitation of the potentially adverse effects of inbreeding. To these primary goals must be added the constraints imposed by any specific facilities and species targeted. The difficulties specific to breeding in aquaculture in regards of the identification of individuals and maintenance or estimation of the pedigree information have been mentioned in section 1.1.1.

The genetic worth of an individual, or its breeding value, is the average additive values of the genes received by this individual from its parents. If this individual is itself used as a parent, it will in turn contribute for one half to the (additive) breeding value of each of its offspring. In the simplest model (infinitesimal model of Fisher), the value of a phenotype is decomposed/regressed into a genetic and a non-genetic (or environmental) effects. In order to estimate the breeding value, or genetic worth, of an individual, one can assess the value of the parent based on the phenotypes of its progenies, a procedure named “progeny testing”. More generally, if phenotypic (trait) measures are available on a pedigreed population, breeding values can be predicted from the Animal Model, a mixed linear model that requires the numerator relationship matrix “A” or equivalently the average relationship matrix (or Abar or $\bar{A}$) in the case of uncertain pedigree (see section 3.3.3).

In aquaculture however, the labor involved in sampling and measuring phenotypes of individual fish and associating it with the fish family identity is quite intensive and expensive. These questions were anticipated as far back as the early 1990s in the SPA breeding program, and it was recognized that a sub-sampling procedure might be

considerably cheaper and easier, and yet, able to recover at least in part the genetic worth of individual parents. The sub-sampling strategy chosen consisted of ONLY sampling offspring of the tested individual in the upper and lower tail of the distribution of some phenotype of interest, in this case the progeny size (see Herbinger *et al*. 1995 for an example). Hence a count of the number of LARGE progenies (L), sampled from the upper tail, and SMALL progenies (S), sampled from the lower tail originating from a specific parent was obtained in order to evaluate to which degree these categories might have a genetic origin reflecting e.g. a superior ability of some parents to produce a high proportion of LARGE fish. This analysis was performed here under the three pedigrees QCEP, QCLODP and FP to assess whether this approach was robust to the type of pedigree. The counts L and S were actually integrated over the probabilistic nature of these pedigrees following the same approach that had been used for reproductive success (Section 3.3.2): the counts were the sum of the number of offspring x the probability for that parent. Statistical differences among parent fish for the proportion of large fish (L/L+S) was evaluated with conventional chi-square homogeneity tests when cells count number were above five.

3.3.6 Genetic Evaluation of Resistance to Superchill

In the winter of 1993-94, some rainbow trout originating from the 1991-92 spawning season (Year 4) were overwintering in one of SPA cages in the Bras d'Or lake (Herbinger, 1994). The majority of the overwintering fish died during a severe superchill episode but some fish survived. This specific study population consisted of 95 surviving fish and 100 mortality fish. DNA microsatellite genotypes at four microsatellite loci (Omy 2, Omy 38, Omy 77 and Omy 105) were made available for all fish. Genetic diversity between the two groups was estimated using CERVUS 3.0 (Marshall *et al*., 1998; Kalinowski *et al*., 2007) for the number of alleles, size in base pairs, allele frequency, observed (H_0) and expected (H_e) heterozygosity and null allele frequency. Reconstruction of sib-ships was again performed using the two available software programs, PEDIGREE Ver 2.2 (Herbinger, 2006) and COLONY Ver. 2.0 (Wang J. ,

2004; Wang and Santure, 2009), to determine if these fish could be partitioned into distinct full-sib and/or half-sib family groups based on available genetic information. The Chi-square Goodness-of-Fit test was used to test for differences in the proportion of survivor and mortality fish in each of the larger credible families with a significance level of 0.05 using Minitab 16.

CHAPTER 4: RESULTS AND DISCUSSION

Genetic diversity was estimated across each generation from this large data base. Three pedigrees (two exclusion -based[1] and one fractional) were generated from a real data set using microsatellite markers from three generations of rainbow trout from the SPA hatchery. Reproductive success was estimated for each parent spawned in generation 1 and generation 2 to see if some individuals were contributing to the next generation more than others. Individual inbreeding values were estimated from the average numerator relationship matrix ($\bar{A}$) created from each of the three pedigrees and average inbreeding values were estimated over each generation to evaluate the overall level of inbreeding in the closed population after 2 generations of selective breeding. In addition a hypothetical generation of individuals "P0" was created from the genetic data of the original broodstock population and added to the pedigree to see how genetic relatedness between "P1" founders might affect the individual inbreeding values and average inbreeding values. We finally conducted a retrospective analysis of whether the spawning recommendations supplied to the hatchery would have been successful in limiting the accumulation of inbreeding over time and whether it would have been possible to identify broodstock producing a higher proportion of large progeny and whether survival to a superchill event had in part a genetic basis.

[1] We will also in the text refer to the exclusion-based pedigree as quasi-categorical or categorical to contrast them with the fractional pedigree

4.1 Genetic Diversity

In total, 2234 individuals from three generations of Rainbow trout were genotyped at up to four microsatellite loci; Omy 2, Omy 38, Omy 77 and Omy 105. Summary statistics were estimated including: allele frequency analysis observed (H_0) and expected (H_e) heterozygosity, deviations from Hardy-Weinberg equilibrium and null allele frequency estimates. Summary statistics per generation are described in the Table 6. Included in this data set are all "P1" individuals from generation 1 (n=224), the subset of "P1" individuals actually used as broodstock (n=90), the AD, NC and RP groups of "P2" individuals (n=1370) from generation 2 and the Group 1Y5, Group 3Y5, Group 5Y5 and Parents 3Y8 "P3" individuals (n=640) that comprise generation 3 of the pedigree. In some instances, the number of individuals typed does not match the number of individuals in the group because data was missing at one or more loci for some individuals. Two other groups from generation 3 (Group 11Y5 and Parents 3Y9) were also originally looked at but were subsequently omitted as their genotypic information did not appear compatible with their purported parents (see details in section 4.2).

The number of alleles per locus per generation ranged from 8 at Omy 38 to 21 at Omy 105. Omy 2 and Omy 105 were the more variable loci with 14 and 18 alleles respectively observed in the founder generation making these loci potentially more discriminating for the reconstruction of parentage. Omy 38 and Omy 77 were less variable with only 8 and 9 alleles detected respectively in the 90 founders. For three out of four loci, the number of alleles seen in the 90 founders was slightly lower that what was observed in the larger Generation 1 fish (n=224). This is probably indicative of a small bottleneck in that generation, with a relatively small number of fish that were actually used as breeders out of the pool of potential "P1" broodfish. The effect of this initial bottleneck could be amplified if large difference in reproductive success was observed among the 90 founders (this is looked at in section 4.3.1) or if the unobserved "P0" parents of the founders were themselves related (this is analyzed in section 4.4). The number of alleles remained generally constant from the "P1" founders to the generation 2 fish except for locus Omy 105 which changed from 15 to 17. In contrast, for each locus,

the number of alleles increased in generation 3 (Table 5). In most cases, the newly detected alleles were 1 or 2 base pairs (a couple were 4 bp) apart from a more commonly occurring allele. These new alleles were most likely due to scoring errors due to changing technicians and genotyping platforms which occurred over the years (Appendices 1a, b, c and d). In three cases, the new alleles were quite distant from the rest of the allelic distribution and may have been transcribed incorrectly into the database. As a result, during the assessment of parentage, a 10% genotyping error rate was used to accommodate the possibility of genotyping error.

High levels of polymorphism were observed in each generation, particularly at Omy 2 and Omy 105 (Table 5). The level of heterozygosity observed (Ho) at all loci was greater than expected heterozygosity (H_e) and significant deviations from H-W expectations were observed in generation 2. In this case, it is not entirely unexpected to observe deviation from H-W expectation. The population of "P1" parents that created the generation 2 fish was not randomly mating and was of limited size, violating at least two of the basic H-W assumptions. The systematically higher level of observed heterozygosities compared to the expected heterozygosities in generation 2 is a good indication that the breeding scheme used at the hatchery with the "P1" parent fish (generation 1) to avoid mating fish from the same group was largely successful. As explained in the Materials section 2.1.1, the 90 "P1" fish were from two different origins (White Sulphur Spring and Loch Bras d'Or). The great majority of crosses that were performed with these fish involved mating fish from different origins, which probably resulted in the spike of H_o compared to H_e. In the other generations (1 and 3) there was no systematic trend, with H_o being sometimes higher, sometimes lower than H_e and with deviations from H-W expectations often being not significant. There was no overall evidence of the presence of null alleles in this large complex three generation population. All null allele frequency estimates were near zero. In addition, none of the loci exhibited a large excess of homozygotes in every generation, as would have been expected if there was a null allele present at a fairly high frequency.

Overall, a high level of genetic diversity appeared to have been maintained from generation to generation. The addition of several new alleles at minimal frequencies was

most likely an artifact due to scoring errors and it appeared to have had a negligible effect on the variability of the gene pool. As well, the increased level of observed heterozygosity in generation 2 is probably a consequence of the rudimentary mating design of not crossing "P1" fish from the same group at the inception of this program.

Table 5: Summary statistics by generation.

Locus	Summary Statistics	Gen. 1 (n=224)	"P 1" (n=90) (subset of Gen .1)	Gen. 2 (n=1370)	Gen. 3 (n=640)
Omy 2	Number of Individuals typed	224	90	1299	535
	Number of alleles	14	13	13	18
	Observed Heterozygosity (Ho)	0.8973 (+)	0.9111 (+)	0.8945 (+)	0.9084 (+)
	Expected Heterozygosity (He)	0.8835	0.8983	0.8732	0.8825
	Null allele frequency estimate	-0.0107	-0.0107	-0.0125	-0.0165
	Deviation from H-W proportions	NS	ND	***	*
Omy 38	Number of Individuals typed	224	90	1263	624
	Number of alleles	8	8	8	11
	Observed Heterozygosity (Ho)	0.6295 (-)	0.6111 (-)	0.7221 (+)	0.6704 (+)
	Expected Heterozygosity (He)	0.6481	0.6639	0.6968	0.6629
	Null allele frequency estimate	0.0139	0.0424	-0.0168	-0.0051
	Deviation from H-W proportions	NS	NS	***	NS
Omy 77	Number of Individuals typed	223	90	1359	635
	Number of alleles	10	9	9	14
	Observed Heterozygosity (Ho)	0.7265 (-)	0.7222 (-)	0.7204 (+)	0.7638 (-)
	Expected Heterozygosity (He)	0.7905	0.7543	0.7163	0.7781
	Null allele frequency estimate	0.0332	0.0096	-0.0015	0.0087
	Deviation from H-W proportions	***	NS	***	*
Omy 105	Number of Individuals typed	221	89	1293	587
	Number of alleles	18	15	17	21
	Observed Heterozygosity (Ho)	0.9502 (+)	0.9438 (+)	0.8515 (+)	0.8569 (-)
	Expected Heterozygosity (He)	0.8849	0.8784	0.8254	0.8648
	Null allele frequency estimate	-0.0395	-0.0428	-0.0157	0.0053
	Deviation from H-W proportions	**	ND	***	NS

(+) indicates that Ho is greater than He while (-) indicates the opposite. Significance of Hardy-Weinberg Equilibrium test (with Bonferroni correction) in evaluating deviation from Hardy-Weinberg proportions: NS = not significant, * = significant at 5%, ** = significant at 1%, *** = significant at 0.1%, ND = not done (CERVUS would not perform the Hardy-Weinberg test if there were too few individuals to allow the test to proceed).

4.2 Assignment of Parentage and Pedigree Reconstruction

Assignment of parentage was carried out to estimate the true pedigree structure of the three generation data set, using microsatellite data to assign progeny to parental pairs based first on exclusion (quasi-categorical allocation method) and later on non-exclusion (fractional allocation method). The overall pedigree was represented as a list of trios, which was created under the exclusion based allocation method and the fractional allocation method and became the structure of the two exclusion based pedigrees and the fractional pedigree evaluated in this study. The data base of 2172 progeny from generations 2 and 3 were evaluated.

4.2.1 Construction of the List of Trios Based on Exclusion

Trios were first determined using the program CERVUS 3.0 by comparing the multilocus genotype of each offspring, with the multilocus genotypes of its putative set of parental pairs. A 10% genotyping error rate was allowed in order to accommodate for the presence of genotyping errors in this large complex and incomplete data set. The proportion of individuals successfully assigned to at least one parental pair ranged from 99.9% for the AD PILOT group in generation 2 to ~ 66% for the RP group from generation 2 and Group 1 Y5 from generation 3 (Table 6). The AD PILOT group consisted of 874 juveniles that were assessed previously as part of the PILOT study (Herbinger *et al*., 1995). Only one individual was not assigned to a trio from this group. It is worth noting that the genotypic data of the AD PILOT fish had been previously completed (all fish were assayed at 4 loci) and corrected for any genotyping errors prior to this project, i.e. for the analysis reported in Herbinger *et al*., 1995. None of the genotypic data for the rest of the fish was corrected for possible genotyping errors and a substantial portion of the fish were scored at only 3 loci or less. Early on in the project it was decided to leave the raw data as is (uncorrected) to assess how successful (or not) pedigree reconstruction would be from data known to have scoring/transcription errors. In this manner, the large AD PILOT fish served as a control group for reconstruction of the

three generation pedigree. The importance of genotype data quality can clearly be seen here. The AD Pilot group was the largest of all, yet all fish except one belonged to at least a trio without any genotypic mismatch. In the other groups, the rate of successful parental allocation was generally lower and many trios involved one genotype mismatch (Table 6).

Table 6: Proportion of individuals assigned to a trio by group per generation.

Parentage assignment	Generation 2				Generation 3						Total
	AD PILOT	AD	NC	RP	G1Y5	G3Y5	G5Y5	G11Y5	P3Y8	P3Y9	
Number of Individuals typed	874	224	140	132	144	141	46	40	309	122	2172
• with genetic data at ≥ 3 loci	874	183	119	113	96	138	46	40	308	120	2037
• with genetic data at < 3 loci	0	41	21	19	48	3	0	0	1	2	135
Parental allocation - no mismatches	873	156	65	55	59	97	40	6	202	15	1724
Parental allocation - mismatch at 1 locus	0	19	31	20	5	30	3	3	58	0	166
No parental allocation	1	8	23	38	32	11	3	31	48	103	298
Proportion assigned to a trio *	0.999	0.956	0.807	0.664	0.667	0.920	0.935	0.225	0.844	0.125	0.854
Proportion not assigned to a trio	0.001	0.044	0.193	0.336	0.333	0.080	0.065	0.775	0.156	0.875	0.146

* includes parental allocations allowing a mismatch at one locus
** not assessed for this offspring group

The AD and NC broodstock, Group 3Y5, Group 5Y5 and Parents 3Y8 groups also had high success rates for assignment of parentage ranging from ~81% to ~96% (Table 6). Most of the assignments were observed with no mismatches at any locus for the trios for these groups. As with all of the other groups, additional trios were obtained by re-evaluating all of the data and allowing an acceptable mismatch at one locus according to the approach described in the Methods section 3.2.1.1. The rate of parental assignments involving one mismatch ranged from 6% (Group 3 Y5) to 26% (NC). In contrast, the RP and Group 1Y5 groups had lower success rates for assigning trios in the order of ~66%, even after allowing mismatches at one locus. These lower success rates are undoubtedly due in part to genotyping error but may have also been an indication that fish from these two groups could have been mixed with other fish at the hatchery. Yet all the RP fish were identified with a Right Pectoral clip (hence the RP designation) so inadvertent

mixing would have had to happen fairly early on before the young fish were marked. Another possible reason for a lower success rate (for the Group 1Y5 fish in particular) is that over half of these fish (57 of the 96) had genotyping profiles with information at only 3 loci (i.e. there was no data at Omy 2). As explained in the Methods section, a one locus mismatch was not allowed for fish with only three loci, and the higher stringency might have resulted in more fish not finding a perfect match by exclusion.

Lastly, two groups had a very large proportion of individuals that were not assigned to any parental pairs. Only 22.5% of the Group 11Y5 and 12.5% of the Parent 3Y9 fish were successfully assigned to a trio, even after allowing for one locus mismatch (Table 6). The genetic data for these two groups of fish did not to seem to fit well with their purported parental genetic information. The Group 11Y5 fish originated from spawning group 11 of the "CRP2Y5" crosses which consisted of seven candidate mothers (two RP and 5 AD fish) and one candidate father that was a sex reversed fish. It could be that these fish did not originate from this specific spawning, but were from another group from the hatchery. It may be as well that the genotype of the lone candidate father was in error for several alleles which would have impacted the overall rate of successful assignment. This group of juveniles was thus excluded from the quasi-categorical pedigrees, due to the small proportion of fish that were successfully assigned to a trio even after allowing for mismatches at one locus (Table 6). Similarly, the Parents 3Y9 individuals were thought to have originated from the "CRP2Y6" spawning that occurred in 1993/1994. After assessing this group, only a very small percentage of individuals were successfully assigned to a trio (Table 6). This group was also assessed against the earlier "CRP2Y5" spawning in case there had been a mistake in the hatchery records but again without success (data not shown). There was a prevalence of alleles 110, 138 and 166 at Omy 2 in this population (Appendix 1a). The high frequency of allele 110 was not remarkably out of place; however the large frequencies of alleles 138 and 166 did not fit with the genotypes of the candidate parents in either spawning group. Discrepancies in scoring at Omy 2 by a different lab technician were noted and corrected by the original technician in order to maintain continuity in scoring for this group. In contrast, the Parents 3Y8 parental allocation to the "CRP2Y5" spawning did fit well with a large

proportion of successful allocations even before having to evaluate the mismatches at one locus. A similarly high percentage of trio assignment for the Parents 3Y9 group without any mismatches was expected. This was not the case, so this group was also excluded from the pedigree. Again this is probably an indication that these fish were not the expected group of fish. It is notable that these broodstock fish had been sampled out of a cage site of one of the clients of SPA, and steelhead trout of another origin were also present on the site. It is quite probable that this group of fish were actually not from SPA which explains why they did not match their putative parents and why they had genetic information at Omy 2 that did not fit the parental profiles. Both groups (Group 11Y5 and Parents 3Y9) were excluded from further analysis with quasi-categorical pedigrees, but Group 11Y5 is later revisited under the fractional pedigree.

4.2.2 Degree of Resolution of Parental Allocation Based on Exclusion

DNA typing profiles at four microsatellite markers were sufficient to match most offspring to parents in most groups, following the strategy outlined in the Methods section 3.2.1. Ideally, most progeny would be allocated to only one pair of parents (i.e. would belong to only one trio). However, given the small number of loci and the fairly large and complex mating systems, it was not surprising that many offspring in many groups could not be unambiguously assigned to only one parental pair. As explained in the Methods section, these progeny that had multiple, equally plausible parental pairs (at least following our simple exclusion rules) were then fractionally allocated to these pairs, either with equi-probabilities or with probabilities reconstructed from the trio LOD scores calculated in the CERVUS runs.

The majority of the AD PILOT fish (~ 65%) were categorically assigned to only one parental pair out of 100 possible parental pairs and a further 26% were “nearly” categorically allocated to only two pairs of parents (Table 7). Only a small proportion of fish (~8%) were assigned to as many as three or four parental pairs with one fish assigned to 7. Only one fish was not assigned to any parental pairs. These findings are consistent

with those of the "PILOT" study in Herbinger *et al*., 1995. Similarly, the majority of the AD broodstock fish (55%) were categorically allocated to only one set of parents and an additional 33% were assigned to two sets of parents (Table 7). A few fish (~8%) were assigned to as many as three, four and even five parental pairs. Eight fish (~4%) were not assigned to any parents. These results were quite similar to those obtained for the AD PILOT group, with a slightly lower resolution. It is likely that the particularly good results for the AD PILOT are due to the fact that the genotypic information had been completed, checked and corrected for this group. Nonetheless, the majority of fish from both AD groups (>85%) were successfully allocated to one or two sets of parents.

Table 7: Distribution of offspring to trios by exclusion method

# trios*	Generation 2								Generation 3								Total	
	AD PILOT		AD		NC		RP		Group 1Y5		Group 3Y5		Group 5Y5		Parents 3 Y8			
	No	Prop	No	Prop	No	Prop	No	Prop	No	Prop	No	Prop	No	Prop	No	Prop	No.	Prop
0	1	0.00	8	0.04	23	0.19	38	0.34	32	0.33	10	0.07	3	0.07	48	0.16	163	0.09
1	57	0.65	10	0.55	32	0.27	61	0.54	42	0.44	99	0.72	42	0.91	12	0.41	107	0.57
2	22	0.26	60	0.33	38	0.32	11	0.10	17	0.18	20	0.14	1	0.02	69	0.22	441	0.23
3	38	0.04	10	0.05	7	0.06	3	0.03	3	0.03	7	0.05	0	0.00	24	0.08	92	0.05
4	35	0.04	3	0.02	12	0.10	0	0.00	1	0.01	1	0.01	0	0.00	13	0.04	65	0.03
5	4	0.00	1	0.01	5	0.04	0	0.00	0	0.00	1	0.01	0	0.00	9	0.03	20	0.01
6	0	0.00	0	0.00	0	0.00	0	0.00	1	0.01	0	0.00	0	0.00	7	0.02	8	0.00
7	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	5	0.02	6	0.00
8	0	0.00	0	0.00	1	0.01	0	0.00	0	0.00	0	0.00	0	0.00	2	0.01	3	0.00
≥ 9	0	0.00	0	0.00	1	0.01	0	0.00	0	0.00	0	0.00	0	0.00	4	0.01	5	0.00
Total	874	1.00	183	1.00	119	1.00	113	1.00	96	1.00	138	1.00	46	1.00	308	1.00	1877	1.00

* includes parental allocation allowing a mismatch at one locus

Assignment of parentage was less successful for the NC and RP broodstock. For the NC broodstock, only 59% of the fish were matched to one (27%) or two (32%) parental pairs (Table 7). A fairly large portion of fish (~19%) did not match to any parents. The remaining fish were assigned to three, four, five, eight and even nine non-excluded trios of which a few required allowing a mismatch at one locus to provide an acceptable match. It is likely that some of the multiple parentage assignments for this group could have occurred by chance as the exact identity of the candidate fathers was

unknown and every progeny was assessed against all possible "P1" males. As such, some of the unresolved relationships derived for this group could be the least credible in the pedigree. For the RP group, resolution was good for the fish that were assigned to any trios. Over half were allocated to one set of parents (54%), while 10% and 3% were allocated to two and three sets of parents respectively. However, ~ 1/3 of the RP group could not be allocated to any trio even after allowing for a mismatch at one locus. These fish were tagged and associated with the "CRP1Y2" spawning so the parental source was supposed to be known. Assuming these fish did originate from this spawning group, it is surprising that such a large proportion of fish did not allocate back to this large pool of candidate parents (40 mothers and 17 fathers) in the specific combinations designated by the mating design.

The Group 1Y5 fish were selected for size (large or small) and sampled as juveniles. Pedigree resolution was similar to that seen in the RP group. Most fish (~ 62%) matched to one or two parental pairs but 32 fish (33%) did not match to any parental pairs (Table 7). As explained previously, the most logical explanation is that these fish from both the RP and Group 1 Y5 that did not match to their parental groups might have been inadvertently mixed. Parental resolution was excellent for the Group 3Y5 and particularly for Group 5Y5 fish (Table 7). The great majority of fish in both of these groups were assigned to only one set of parents, 72% for Group 3 Y5 and 91% for Group 5Y5. Very few fish did not allocate to any parents. Such good result in Group 5Y5 is undoubtedly due to the simple mating design and limited number of parents (6 females and 4 males that had been crossed

For the most part, parental allocation was very successful for the Parents 3Y8 group (Table 7). Approximately 42% of the fish were unambiguously assigned to only one set of parents. An additional 22% were assigned to two sets of parents, 20% were assigned to as many as three to eleven non-excluded trios and ~16% did not allocate to any parents. This group was the most complex group to assess because it was associated to the "CRP2Y5" spawning, which involved 126 females and 24 males in a very complicated (and far from factorial) series of crosses. With such a large number of potential parental pairs, it is actually surprising to have observed such a good resolution.

As mentioned earlier, for each group, the parentage analysis started with CERVUS to determine the trios associated with each progeny. However CERVUS operates under the assumptions that every male parent had an equal chance of fertilization of each female parent (i.e. it makes the prior assumption of factorial mating). This was the case for the AD and NC groups. The AD group was entirely derived from a factorial mating of 10 males by 10 females. Such a design had probably not been used for the NC group, but since the actual males were unknown, this group had to be approached as well as a factorial mating where very known female could have mated with every male (unknown).These two groups could then be assessed using one CERVUS run. However the mating designs for all other groups were considerably more complicated and consisted of a series of disconnected partial diallel crosses (groups and pools). For example 20 females and 10 males might have been used, with females 1 to 4 factorially mated with males 1 and 2; females 5 to 8 to males 3 and 4 and so forth with all the eggs pooled at hatching. A single CERVUS run could not be used in this case because it might assign a progeny to female 1 and male 4 even though this specific cross did not happen. Several CERVUS runs would have to take place; one for each factorial cross subset (each partial diallel) and all of the results had to be meshed later as outlined in the methods. A further difficulty arose because the different partial diallels were not entirely disconnected from each other. Many males were reused in different spawning groups and pools and this even happened for a few females. As the number of spawning groups/pools increased, the number of CERVUS runs required also increased This resulted in a complicated situation where up to four scenarios could be seen for a specific progeny: 1) one parental pair was allocated from within a single spawning group/pool, 2) more than one parental pair was allocated from within a single spawning group/pool, 3) more than one parental pair was allocated from different spawning groups/pools and 4) the same parental pair was allocated from different spawning groups/pools. A summary of the distribution for each scenario type can be seen in Table 8.

The Parents 3Y8 offspring group was by far the most complex group to assess. They were associated to the “CRP2Y5” spawning (126 candidate mothers and 24 candidate fathers) comprised of 11 spawning groups further divided into 57 spawning

pools. Assessment of the trios was complicated for this group as some candidate parents were used in multiple spawning groups, and for some of the fish assigned to many parental pairs, all scenarios were observed (Table 8). Under the approach based on assignment by exclusion, each of the non-excluded parental pairs derived from any of the four different assignment scenarios were assumed to be potentially the true parent and were all included in the pedigree. For the list of trios corresponding to the quasi-categorical pedigree with equal probability, the fish assigned to multiple trios were allocated to each parental pair with equal probability. Each trio included in the list of trios corresponding to the quasi-categorical pedigree was also given a probability computed from trio LODs, as explained in the Methods section 3.2.2. Hence this approach required not only meshing the potential trios across the many runs but as well extracting and meshing all relevant trio LOD information from a very large number of runs.

Table 8: Distribution of the four different parental allocation scenarios

Scenarios	Generation 2				Generation 3			
	AD PILOT	AD	NC	RP	G1 Y5	G3 Y5	G5 Y5	P3 Y8
Number of spawning groups (pools)	1	1	1	9 (25)	1 (7)	1 (10)	1 (4)	11 (57)
• one parental pair from a single spawning group	570	101	32	61	42	99	42	126
• more than one parental pair from a single spawning group	303	377	64	3	11	11	0	2
• more than one parental pair from different spawning groups	0	0	0	11	11	18	1	130
• same parental pair from different spawning groups	0	0	0	0	0	0	0	45
• no allocation	1	8	23	38	32	10	3	48
Total:	874	183	119	113	96	138	46	>308

The large data base for this project started out consisting of 2172 fish from generation 2 and 3 that were to be evaluated and for which parental assignment would create the list of trios used in the two quasi-categorical pedigrees. It was determined that 135 fish had genetic data at less than 3 loci and these fish were immediately dismissed because using information at less than 3 loci would not be very discriminating . This left a pool of 2037 fish to be assessed. During the course of the evaluation, the Group 11 Y5 fish (n=40) and the Parents 3 Y9 fish (n=122) were excluded from the pedigree because very few fish from these groups successfully allocated to any parents, raising serious

doubts as to the real identity of the fish in these groups. This left 1877 fish in the pedigree. In the end, 1075 of the 1877 fish (57%) were uniquely assigned to one set of parents, 639 fish (34%) were assigned to multiple parental pairs and 163 fish (9%) had no parents assigned. The fish that ended up having no parents assigned based on the exclusion approach used here represented potential holes in the quasi-categorical pedigrees and were treated as founders. The gaps in relatedness left by these unconnected fish could affect inbreeding value estimates for these categorical pedigrees (see section 4.4). In total, the exclusion based "nearly" categorical pedigrees consisted of 2753 trios that were determined for 1714 fish from generations 2 and 3, with 1075 fish represented in one unique trio and the remaining 639 fish represented in average in ~2.6 trios.

4.2.2 Construction of the List of Trios by Fractional Allocation

As with the exclusion method, trios were determined by comparing the multilocus genotypes of all members of the trios, for those trios corresponding to the mating design of each spawning group. As for the CERVUS runs used to construct the two nearly categorical pedigrees, a 10% genotyping error rate was allowed in order to accommodate for the presence of genotyping error in the data. Parentage was determined for all fish from generations 2 and 3, regardless of the number of loci with genetic information, with the exception of the Parents 3Y9 group (n=122). Assessment of parentage by exclusion had convincingly determined that the Parents 3Y9 did not fit with either of the "P2" parent groups (CRP2Y5 or CRP2Y6) and were probably an unrelated group of fish so it was not included for evaluation by the fractional method. The Group 11 Y5 fish were included to see if fractional allocation might prove to be more successful than exclusion at resolving the most likely parents for this group. In total, 2050 fish from generations 2 and 3 were evaluated by the fractional method and 22994 trios were generated (a 10 fold increase in the number of potential parental allocations). Only trios with positive LODS (see Methods section 3.2.3) were retained in the pedigree. Proportional probabilities were estimated for each fish as described in the methods. For all groups, many more trios were

observed than by exclusion. A summary of the distribution of offspring to trios by the fractional approach can be seen in Table 9.

The AD PILOT group and the AD broodstock group showed similar results: most progeny were assigned between 4 and 17 parental pairs. Only 2 fish were attributed to either 1 or 2 trios in the AD broodstock and none in the AD PILOT fish. Interestingly, even with fractional allocation and with a fairly high error rate, some individuals still did not allocate to any parents: 6 AD broodstock had no trios, as compared to 8 in the quasi-categorical pedigrees. The one fish from the AD PILOT group that did not allocate to any parents by exclusion (see Table 7) was assigned to 3 trios by the fractional method with the most likely of the three having a probability of 0.69.

The distribution for the NC fish in trios was particularly flat and ranged greatly from 0 to >40 sets of parents (Table 9). The majority of NC fish (~61%) allocated to between 10 and 29 trios, but one fish allocated to 124 sets of parents and a second fish allocated to 144 sets of parents. These two fish had genetic information at only 1 locus so this is not a particularly surprising result. None of the NC fish allocated to only one set of parents as compared to 32 by exclusion and 4 fish did not allocate to any parents as compared to 23 by exclusion (see Table 8). The distribution of trios for the RP fish was tighter with the majority of RP fish (~62%) allocated to between 1 and 6 trios, as compared to 67% allocating to between 1 and 3 trios by exclusion. Many fish were allocated to parents by the fractional method that had been excluded with the categorical pedigrees: only 7 fish (5%) remaining unassigned vs. 38 previously unassigned by exclusion (38%). The distribution of trios for the RP fish is interesting because although they are not the same, the categorical and fractional allocation to trios are fairly similar where the majority of fish were allocated to relatively a small number of trios.

The pattern of trio distributions for the Group 1 and 5 Y5 progeny were very similar to what had been observed in the RP group (Table 9). The majority of the Group 1Y5 fish (~64%) allocated to 1 - 5 trios as compared to 62% allocating to 1 - 2 trios by exclusion (Table 7). Only two fish did not allocated to any parents as compared to 32 fish by exclusion. The remaining fish allocated to anywhere from six trios to as many as 30

trios. The majority of the Group 5Y5 progeny (~54%) allocated to only one set of parents as compared to 91% by exclusion (Table 9). The remaining fish allocated to only 2 - 4 sets of parents. Notably for this group, one fish still did not allocate to any parents even by the fractional method as compared to three that remained unassigned by exclusion. The Group 3 Y5 group showed larger variation in the distribution in trios under the fractional approach compared to the exclusion method. Most Group 3Y5 fish were allocated from two to over 10 sets of parents (Table 9). One fish was assigned to only one set of parents and one fish did not allocate to any parents. Twenty-one percent of the fish allocated to 11 to 20 trios and one fish allocated to 25 trios. This distribution was substantially more distant from the exclusion distribution where 72% of the fish were categorically assigned to only one set of parents, as compared to for example the RP fish or Group 1 Y5.

The distribution for the Parents 3Y8 group was also very different under the fractional compared to the exclusion distribution. The fractional distribution varied greatly from only 0 to 48 sets of parents (Table 9) and the majority (70%) allocated to between 10 and 29 trios as compared to the majority (64%) allocating to only one or two trios by exclusion. Some fish (5%) still did not allocate to any parents even by the fractional method compared to 16% by exclusion. These results are interesting because this particular group of fish originated from the most complicated set of crosses (CRP2Y5), and was the most time consuming group of fish to assess by the exclusion method due to the large number of CERVUS runs required to be interpreted. Hence the benefits of using the fractional approach (one single run integrating all the information over the pedigree) were the highest for this group, but the fractional pedigree for this group certainly appeared more diffuse than the categorical pedigrees had been.

Lastly, fractional allocation was attempted for Group 11Y5 fish. This group of fish had very poor success under the categorical methods and was not integrated into the pedigree. Most of these fish (45%) were uniquely allocated to only one set of parents and another 9 to 2 sets of parents (23%) one set of which had a very high probability (>0.99) of being the true parents. Twelve fish (30%) remained unassigned by the fractional method. This rate of unsuccessful assignment by the fractional approach was by far the

highest of any group. It would appear, as had been hypothesized earlier, that a relatively large proportion of the fish in this group did not really originate from the proposed crosses and were probably mixed in early on at the hatchery. Yet other fish in this group probably truly were progeny of the purported crosses and were recovered by the fractional allocation approach when they had not been with the quasi-categorical approach. The single sex reversed father used in the group 11 spawning of the "CRP2Y5" crosses only had information at 3 loci. Under the exclusion approach used here, this meant that no mismatches were allowed to call an acceptable match at the other three loci. Such a stringent rule probably resulted in several additional fish being falsely excluded, and based on exclusion, only five fish were allocated to trios with no mismatches. This was not sufficient to have confidence in the allocations so this group was not included in the exclusion based pedigree. It is interesting to see that many more fish were successfully allocated to trios by the fractional method

Table 9: Distribution of offspring to trios by fractional method

# trios	Generation 2								Generation 3										Total	
	ADPILOT		AD		NC		RP		G1Y5		G3Y5		G5Y5		P3Y8		G11Y5			
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
0	0	0.00	6	0.03	4	0.03	7	0.05	2	0.01	1	0.01	1	0.02	14	0.05	12	0.30	47	0.02
1	0	0.00	1	0.00	0	0.00	7	0.05	14	0.10	1	0.01	25	0.54	2	0.01	18	0.45	68	0.03
2	0	0.00	1	0.00	2	0.01	12	0.09	18	0.13	10	0.07	9	0.20	2	0.01	9	0.23	63	0.03
3	8	0.01	3	0.01	3	0.02	21	0.16	17	0.12	14	0.10	6	0.13	3	0.01	1	0.03	76	0.04
4	8	0.01	7	0.03	2	0.01	11	0.08	26	0.18	14	0.10	5	0.11	4	0.01	0	0.00	77	0.04
5	59	0.07	13	0.06	5	0.04	17	0.13	16	0.11	16	0.11	0	0.00	1	0.00	0	0.00	127	0.06
6	62	0.07	23	0.10	5	0.04	14	0.11	9	0.06	21	0.15	0	0.00	7	0.02	0	0.00	141	0.07
7	89	0.10	17	0.08	3	0.02	11	0.08	8	0.06	13	0.09	0	0.00	1	0.00	0	0.00	142	0.07
8	118	0.14	19	0.08	4	0.03	5	0.04	3	0.02	10	0.07	0	0.00	4	0.01	0	0.00	163	0.08
9	110	0.13	27	0.12	0	0.00	7	0.05	7	0.05	11	0.08	0	0.00	4	0.01	0	0.00	166	0.08
10	131	0.15	24	0.11	5	0.04	3	0.02	3	0.02	13	0.09	0	0.00	12	0.04	0	0.00	191	0.09
11	90	0.10	19	0.08	8	0.06	3	0.02	1	0.01	4	0.03	0	0.00	10	0.03	0	0.00	135	0.07
12	52	0.06	17	0.08	3	0.02	2	0.02	3	0.02	2	0.01	0	0.00	9	0.03	0	0.00	88	0.04
13	53	0.06	6	0.03	9	0.06	1	0.01	4	0.03	3	0.02	0	0.00	7	0.02	0	0.00	83	0.04
14	21	0.02	9	0.04	3	0.02	0	0.00	2	0.01	1	0.01	0	0.00	10	0.03	0	0.00	46	0.02
15	20	0.02	3	0.01	4	0.03	1	0.01	1	0.01	1	0.01	0	0.00	16	0.05	0	0.00	46	0.02
16	22	0.03	4	0.02	7	0.05	1	0.01	0	0.00	2	0.01	0	0.00	12	0.04	0	0.00	48	0.02
17	12	0.01	3	0.01	3	0.02	1	0.01	0	0.00	3	0.02	0	0.00	16	0.05	0	0.00	38	0.02
18	7	0.01	1	0.00	5	0.04	0	0.00	0	0.00	0	0.00	0	0.00	10	0.03	0	0.00	23	0.01
19	3	0.00	0	0.00	4	0.03	0	0.00	2	0.01	0	0.00	0	0.00	9	0.03	0	0.00	18	0.01
20	5	0.01	0	0.00	6	0.04	0	0.00	3	0.02	0	0.00	0	0.00	16	0.05	0	0.00	30	0.01
21	1	0.00	4	0.02	5	0.04	1	0.01	2	0.01	0	0.00	0	0.00	14	0.05	0	0.00	27	0.01
22	2	0.00	0	0.00	5	0.04	4	0.03	1	0.01	0	0.00	0	0.00	12	0.04	0	0.00	24	0.01
23	0	0.00	1	0.00	4	0.03	0	0.00	0	0.00	0	0.00	0	0.00	13	0.04	0	0.00	18	0.01
24	1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	13	0.04	0	0.00	14	0.01
25	0	0.00	0	0.00	2	0.01	0	0.00	0	0.00	1	0.01	0	0.00	12	0.04	0	0.00	15	0.01
26-30	0	0.00	2	0.01	16	0.11	1	0.01	2	0.01	0	0.00	0	0.00	34	0.11	0	0.00	55	0.03
31-35	0	0.00	4	0.02	5	0.04	0	0.00	0	0.00	0	0.00	0	0.00	22	0.07	0	0.00	31	0.02
36-40	0	0.00	0	0.00	2	0.01	0	0.00	0	0.00	0	0.00	0	0.00	10	0.03	0	0.00	12	0.01
>40	0	0.00	10	0.04	16	0.11	2	0.02	0	0.00	0	0.00	0	0.00	10	0.03	0	0.00	38	0.02
Total:	874	1.00	224	1.00	140	1.00	132	1.00	144	1.00	141	1.00	46	1.00	309	1.00	40	1.00	2050	1.00

As mentioned earlier, the large data base for this project started out consisting of 2172 fish from generation 2 and 3 that were to be evaluated for assignment of parentage. Unlike the exclusion method, those fish with molecular data at less than three loci were retained for evaluation by the fractional method. However, the Parents 3Y9 fish (n=122) were not included for evaluation because very few fish from this group had been successfully allocated to any parents by exclusion. In the end, 2050 were assessed. Only 3% of the fish were categorically assigned to one set of parents by the fractional method as compared to 57% by exclusion. However, still 2% remained with no parents assigned as compared to 9% by exclusion.

As expected, the fractional method was much more successful in allocating parent to offspring than the exclusion method and for the most part, the most likely trios observed with fractional allocation corresponded to those non-excluded trios with no mismatches observed. The few exceptions that were noted were mostly seen in the situation where a trio was accepted by the exclusion method after allowing a mismatch at one locus. Sometimes the trio was not the most likely one detected by the fractional method and corresponded to the 2^{nd} or 3^{rd} best choice. With the two quasi-categorical pedigrees, 163 fish (9%) were "holes" in the pedigree with no parental allocation. This was reduced to 47 fish (2%) under the fractional pedigree. For some groups (RP, Group 1Y5 and Group 5Y5) some small loss of resolution was observed under the fractional pedigree. For others, there was a higher loss of resolution (AD and Group 3Y5). Lastly, in the case of the NC and Parents 3 Y8, there was a considerable apparent loss of resolution with the fractional method which was most likely due to individuals having information at only two or even only one locus.

Genetic evaluation relies on accurate pedigree information. Ideally, the most accurate pedigree results when one offspring is categorically assigned to one set of parents. Genotyping errors and incomplete sampling of parents can cause false exclusions of true parents when trying to assign parents to offspring by the exclusion based method, which can lead to gaps or "holes" in the pedigree (O'Reilly *et al*., 1998; Bonin *et al*., 2004; Dodds *et al*., 2005). With this data set, however, not all offspring were successfully allocated to parents leaving "holes" which could affect estimates of reproductive success

and inbreeding values. Ford *et al*., 2010 observed that estimates of relative fitness were negatively biased when the true parents could not able to be unambiguously resolved and the downward bias was effectively eliminated by fractionally assigning progeny to parents. The "holes" in this data set were essentially resolved under fractional allocation (only 2% unresolved parentage) potentially eliminating any potential downward bias.

The large error rate (10%) that was used to create the categorical pedigrees by CERVUS and the fractional pedigree is a most likely source of some of the ambiguity in the three pedigree types especially in the situations where some individuals had many trios assigned after allowing a mismatch. A smaller acceptable error rate to would make trio assignment more accurate, potentially leading to a precise pedigree where at most one pair of parents is assigned to any offspring. With limited amount of genetic information (4 loci) this would still not necessarily resolve the issue of incomplete exclusion that was often observed with this data set. In addition, this would also increase the number of unallocated progeny due to false exclusion. This large data set had been typed at only four loci because of the limited number of loci available and the high cost of genotyping at the time. These limitations are much less acute nowadays. Many more microsatellite (and other types) loci are available for many aquaculture species which can often be amplified with PCR multiplexes. Multiplex-PCR consists of multiple primer sets within a single PCR mixture to produce amplicons of varying sizes that are specific to different DNA sequences. By targeting multiple loci at once, additional information may be gained from a single test run that otherwise would require several times the reagents and more time to perform (Fishback *et al*., 1999; Delghandi *et al*., 2003; Taris *et al*., 2005; Renshaw *et al*., 2006). Increasing the number of loci used and using a reasonably small error rate should help resolve incomplete exclusion and false exclusion, particularly under the fractional pedigree approach.

4.3 Reproductive Success

For the various generations, the reproductive success of each individual was estimated as the sum of the probabilities of any trio containing this individual as a parent (either a male or a female) (see Methods section 3.3.2). Different probabilities were associated with the three pedigrees (i.e. the three lists of trios). Differences among female and male reproductive success in the various groups were estimated with chi square tests, but these tests were performed only when the expected cell counts were above 5. Thirty three chi square tests were performed (Figures 5 to 10) and 0.0016 was used for the adjusted significance level following a Bonferroni approach. As had been noted earlier (section 4.2), there were a number of individuals which were not allocated a trio under the two quasi-categorical pedigrees but were under the fractional pedigree, and this depended upon the particular group of interest. For example, in the AD PILOT group, all individuals were allocated a trio under either quasi-categorical or fractional pedigrees (except one fish). In contrast, several other groups (i.e. the RP, G1Y5 or P3Y8 groups) had a substantial proportion of individuals not attributed to any trio under the quasi-categorical pedigrees. To allow a comparison of estimated reproductive success under the different types of pedigree for the groups which had large differences in the number of allocated progeny, the relative success of the different parents were presented in addition to the absolute success. In the following sections the three different types of pedigrees will be termed quasi-categorical-equal probability pedigree (QCEP), quasi-categorical-LOD based pedigree (QCLODP) or fractional pedigree (FP).

4.3.1 Generation 1

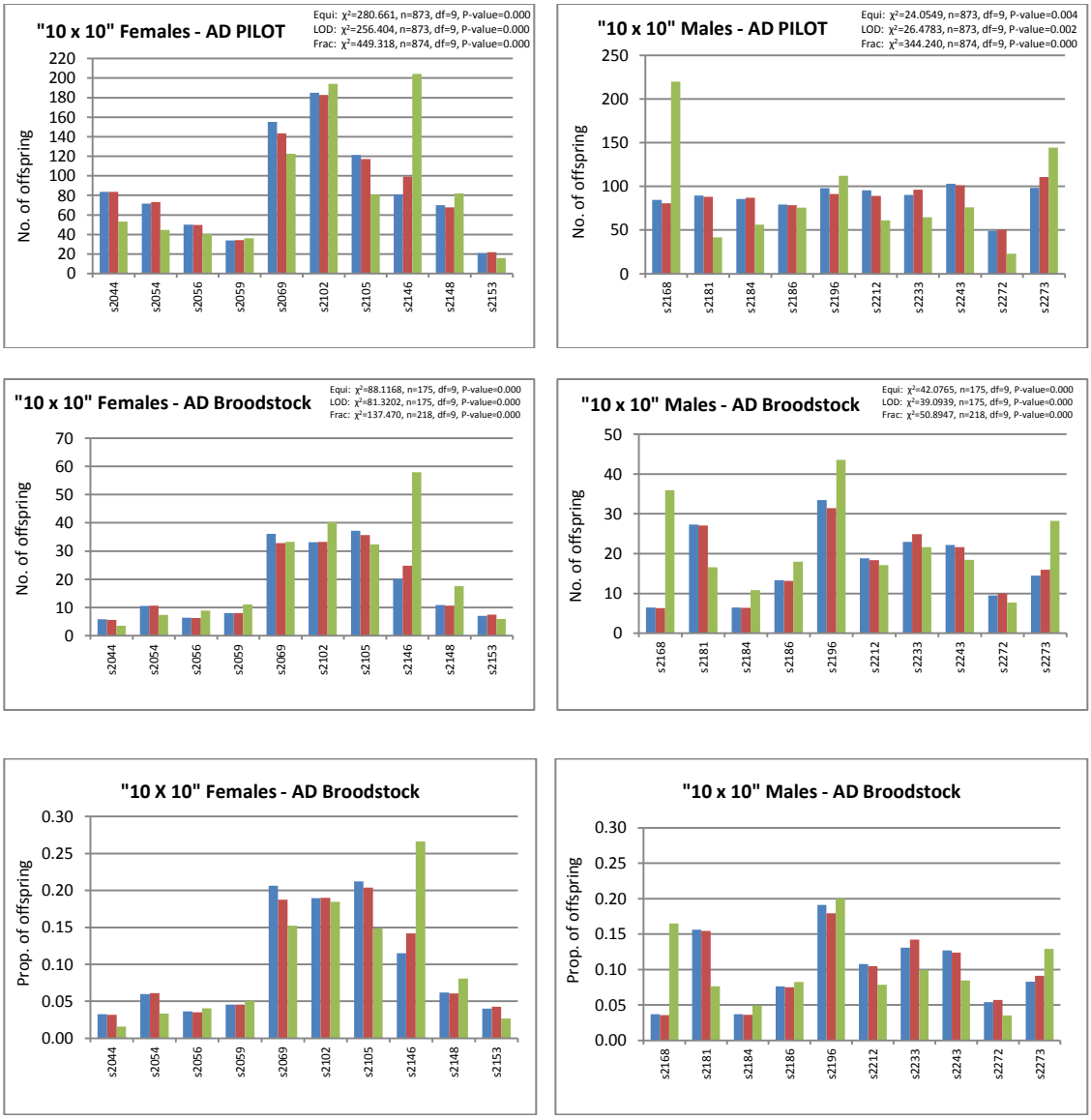
With the "P1" matings some of the males used in the "10 x 10" crosses (AD) were also used in the "CRP1Y2" crosses (RP) but none of the females were reused. For example, 6 of the 10 males used in the "10 X 10" factorial crosses were also used as candidate fathers in the "CRP1Y2" crosses and 7 were included as candidate fathers for

the NC group of offspring. As expected, some parents had much higher reproductive success than others as assessed both by the quasi-categorical and fractional pedigrees.

For the AD PILOT group, under both QCEP and QCLODP pedigrees, four of the mothers (s2069, s2102, s2105 and s2146) appeared to have produced most of the offspring (Figure 5, top left graph). Overall differences among females were highly significant even after Bonferroni correction for multiple tests (see chi-square values on Figure 5). The contribution of the fathers was more fairly evenly distributed (Figure 5, top right graph) with the exception of fish s2272. Differences were not strictly significant using the Bonferroni adjusted values but they were close to the threshold (see chi-square values on Figure 5). These findings are consistent with those obtained previously from this "PILOT" study (Herbinger *et al*., 1995) as expected since the offspring allocation scheme that was followed in that early study is similar to the QCEP scheme followed here. It is worth noting that most of the differences among females appeared to be linked with female fertility, as there was a significant correlation between the number of offspring produced and the number of eggs (Herbinger *et al*., 1995).

When looking at reproductive success of the 10 females and 10 male broodstock under the fractional pedigree (FP), fairly similar patterns could be seen with two notable exceptions: one mother (s2146) and one father (s2168) in particular appeared to have many more offspring allocated to them by the fractional method than by either quasi-categorical ones (Figure 5). Compared to either QCEP or QCLODP, the fractional pedigree reproductive success of that specific female was twice as large. Reproductive successes under the quasi-categorical or fractional pedigrees were in good agreement for the other dams and despite the difference for that female s2146, there was a significant Spearman rank correlation ($r = 0.782$, p-value $=0.008$) between the FP reproductive success and the one evaluated under QCEP. Among the males, more differences could be seen when comparing the reproductive success calculated from these two pedigrees. In addition to male s2168 already mentioned, substantial differences could also be seen in male s2181 and s2273 for example. Overall Spearman rank correlation of reproductive success under FP and QCEP was lower ($r = 0.328$) and not significant.

Figure 5: Reproductive success of the parents of the AD PILOT fish (top graphs), the AD broodstock (middle graphs) and the relative reproductive success for the AD broodstock (bottom graphs) Pedigree legend: blue-QCEP, red-QCLODP, green-FP. Chi-square test results under each pedigree are also presented for the absolute reproductive successes.



The AD broodstock fish came from the same "10x10" crosses than the AD PILOT, but were sampled as broodstock of second generation at a later age, while the PILOT fish had been sampled as one year old juveniles. The distribution of reproductive

success of the mothers of the AD broodstock (Figure 5, middle graphs) under either QCEP, QCLODP or FP was very similar to that observed for the AD PILOT fish. The same four mothers appeared to have produced the most offspring, while some slight variation was noted for the other mothers, which is probably due to higher sample variance due to smaller sample size. Within the AD broodstock fish, there was a significant rank correlation of reproductive success obtained under QCEP and FP ($r = 0.818$ $p = 0.004$). Across the two progeny groups, there were significant rank correlations of female reproductive success evaluated in the AD PILOT progeny and in the AD broodstock under QCEP ($r = 0.648$ p-value $= 0.043$) or FP ($r = 0.818$ p-value $= 0.004$). This indicates overall that ranking of females in terms of number of offspring produced in the next generation appeared to be fairly robust to the use of a quasi-categorical pedigree versus a fractional pedigree (despite the large difference seen for female s2146) and robust as well to the use of younger progeny (AD PILOT fish) versus older ones (AD broodstock).

The trend for reproductive success of the fathers of the AD broodstock (Figure 5, right middle graph) was quite similar to what had been seen in the AD pilot progeny. Spearman rank correlation of male reproductive success under QCEP and FP was again lower ($r = 0.316$) and not significant. In contrast, there were a nearly significant rank correlation of male reproductive success evaluated in the AD PILOT progeny and in the AD broodstock under QCEP ($r = 0.628$ p-value $= 0.052$) and a significant one under FP ($r = 0.903$ p-value < 0.001). This indicates overall that ranking of males in terms of number of offspring produced in the next generation appeared to be fairly robust to the use of younger progeny (PILOT fish) versus older ones (AD broodstock), but was more sensitive to the choice of estimation of pedigree.

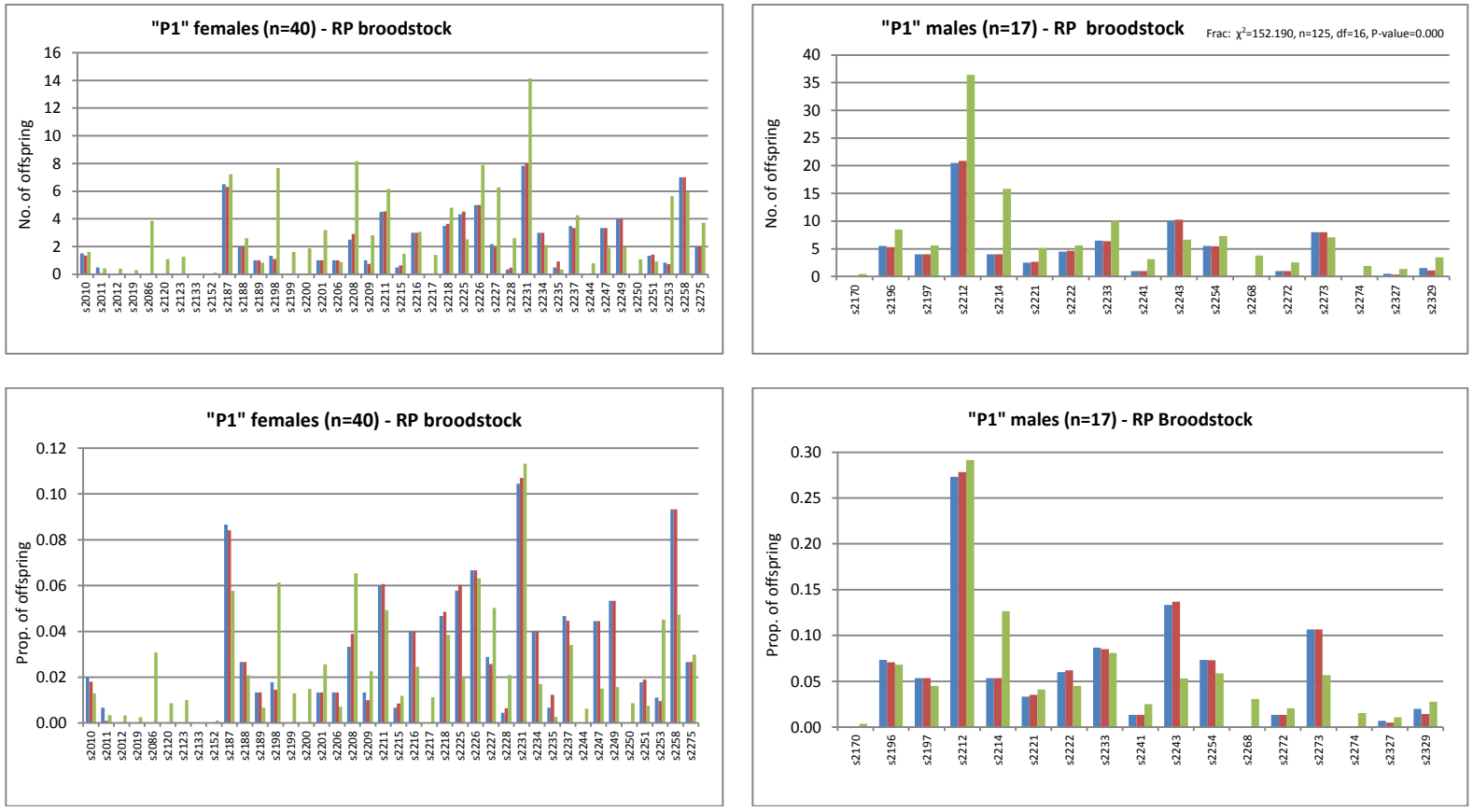
The distribution of male and female reproductive success was estimated in the AD broodstock for QCEP over 175 offspring while the distribution for FP was estimated over 218 offspring. The relative distribution of the number of offspring under the three pedigrees for the mothers and fathers of the AD broodstock (Figure 5, bottom graphs) shows that large differences between the quasi-categorical and fractional reproductive success were observed for mother s2146 and father s2168 and to a smaller extent for male

s2181 and s2273, just like they had been observed for the AD PILOT fish. Such difference appears therefore to be an effect of categorical (or quasi-categorical) vs. fractional assignment of parentage. The offspring allocated to mother s2146 and father s2168 under FP and QCEP were re-examined to try to understand where such differences could have originated. It appears that under the fractional pedigree, a large number of "fractions of offspring" (i.e. probabilities) were assigned to these two parents where the fractions were not the highest possible ones for these offspring. In other words, an offspring "o" could be assigned to a trio including that female or that male (or both) with a probability “p” but “o” was also attributed to at least another trio not involving this male or female with a probability higher than “p”. The sum of the fractions for a large number of such offspring under the fractional pedigree resulted in a large reproductive success for these two individuals. In contrast, under either QCEP or QCLODP, these offspring were generally not assigned to a trio involving that male or that female and thus they did not contribute to their reproductive success. Equivalently, a large number of partial offspring were assigned to these two fish but they were not necessarily the “first” choice of parents. Upon examination of these two fish genotypes, there were no obvious explanations as to why these two specific fish should so often be included as the second or third best trio. They had fairly common alleles at all loci but they were not the only parent fish in this situation. It is also possible that the large error rate that was used to generate the probabilistic (i.e. fractional) pedigree might have somehow provided them an "edge" in being allocated to such a large number of offspring as second or third best trio.

In order to estimate reproductive success accurately the pedigree must be as accurate as possible (i.e. most progeny must be allocated to exactly one parental pair). “Holes” in the pedigree due to genotyping errors have been shown to produce a downward bias of estimated reproductive success which can be eliminated with fractional allocation because fractional allocation does not attempt to assign only one pair of parents to an offspring but instead divides each offspring among multiple parents in proportion to their likelihood (Jones *et al*., 2003; Ford *et al*., 2010). Large differences were also observed in the number of offspring that could be assigned to the various mothers and fathers of the RP broodstock (Figure 6) by both the quasi-categorical and fractional

methods. Three females in particular (s2187, s2231 and s2258) one male (s2212) appear to have produced the most RP offspring based on QCEP and QCLODP and were also top contributors based on the fractional pedigree. The distribution of female and male reproductive success based on the fractional method appeared quite different for both. Seemingly a disproportionate number of offspring were assigned to many of the parents. As was the case with the AD broodstock group above, it is important to remember that there were more individuals assigned at least one trio under the fractional pedigree (126) compared to the quasi-categorical ones (76), so a relative distribution of the number of offspring was estimated for the mothers and fathers of the RP broodstock (Figure 6, bottom graphs) to correct for this effect. Most large differences previously observed disappeared indicating that these differences were mostly due to the effect of sample size and not to the type of pedigree. However, there were still some females and males which had fairly different estimated reproductive success under the fractional versus quasi-categorical pedigrees. For example male s2214 had a much larger relative reproductive success under the fractional pedigree and the converse could be seen for male s2243. This is probably indicative again of differences in partial allocations to trios and their associated probabilities. Male s2214 may have experienced the same process that has been proposed for female s2246 and male s2168 in the AD group. In contrast male s2243 might have been allocated to the same collection of offspring under both the fractional and quasi-categorical pedigrees but with probability much higher (mostly 1.0) under the quasi-categorical pedigree. Interestingly, quite a few parents did not appear to produce any offspring or a very small number (≤ 1) as assigned by the quasi-categorical methods but were assigned offspring by the fractional method. Even though these parent fish also showed generally low reproductive success under the fractional pedigree, this was not always the case. Female s2086 had a reproductive success of ~ 4 (the twelfth highest of the forty females) under the fractional pedigree but had no offspring assigned to her under either quasi-categorical pedigrees. The fact that under the quasi-categorical pedigrees, some progeny could not be assigned to a parental pair and therefore had to be treated as founders unconnected to the rest of the pedigree, and conversely the fact that some

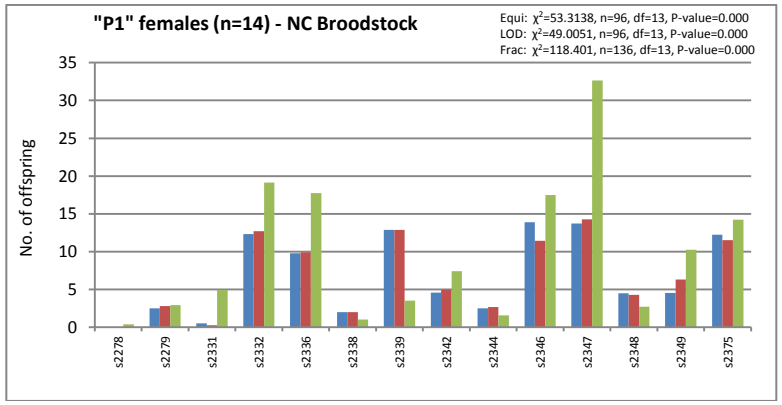
Figure 6: Reproductive success of the parents of the RP broodstock (top graphs) and relative reproductive success (bottom graphs). Pedigree legend: blue-QCEP, red-QCLODP, green-FP.

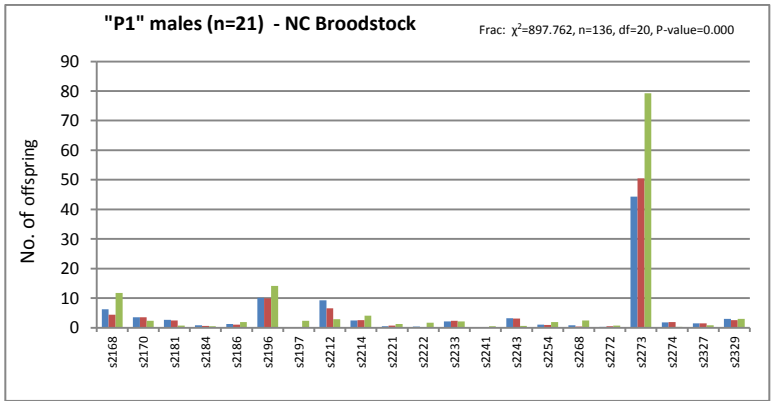


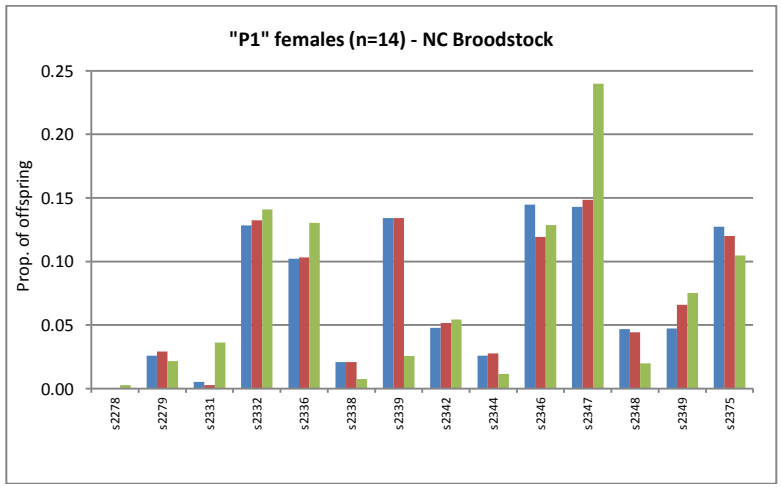
parents may not be assigned any progeny appeared to be a nuisance. These "holes" left by the categorical method could potentially affect the estimation of individual and average inbreeding levels in the pedigree (see section 4.4).

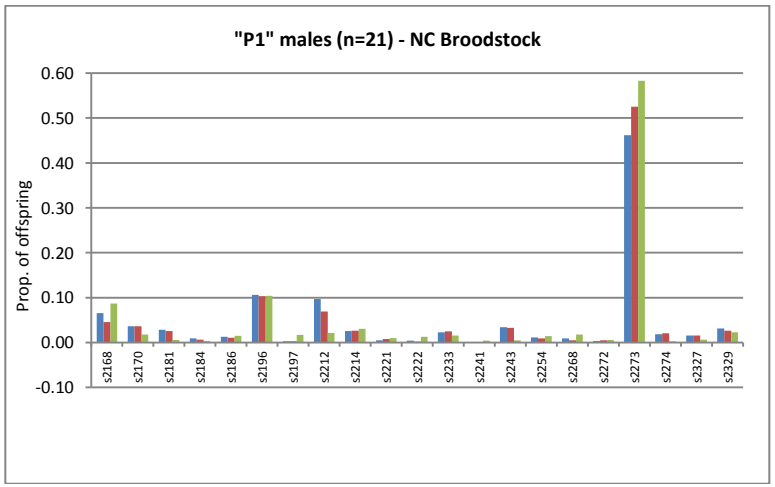
Similarly, large differences in reproductive success were observed for the various mothers and fathers of the NC broodstock (Figure 7) by both the quasi-categorical and fractional methods. The characteristic differences between these quasi-categorical and fractional methods were also observed, particularly with the females. Interestingly, even after normalizing the distribution for the total number of offspring, the reproductive success for one female (s2339) was observed to be much lower for the fractional method than the quasi-categorical and for another fish (s2347) the converse was seen. These two fish have very similar genotypes, differing by only one allele at Omy 2. By the exclusion method, 26 fish were allocated to both s2339 and s2347. However, with the fractional method, where all genotypic information is considered and all trios are considered, a higher proportion of offspring were allocated to s2347. The NC broodstock was assessed against all possible "P1" males as the true males were not known, and neither was the number of males that had been used. Under all three pedigrees, the top male contributor to the NC broodstock (s2273) appears to have produced as many progeny (~50%) as all the other males combined (Figure 7). Most males had no offspring assigned at all (quasi-categorical methods) or very few (fractional method). This is probably reflecting the fact that in reality very few males (probably no more than 6) had been used in this spawning, and furthermore, for reasons unknown, one male managed to monopolize most of the reproductive success.

Figure 7: Reproductive success of the parents of the NC broodstock (top graphs) and relative reproductive success (bottom graphs). Pedigree legend: blue-QCEP, red-QCLODP, green-FP.









It is notable that a lower number of males were used in the first generation for the both the RP (n=17) and NC (~6??) groups as compared to the females (n=40 for the RP and n=14 for the NC). This is a fairly common practice in fish hatcheries (Dube and Mason, 1995). One male in particular, s2273, appears to be a significant contributor to the pedigree overall, being the most significant contributor to the NC broodstock and also contributing to the RP and AD broodstock.

For the parents of the AD, RP and NC offspring groups differences in both male and female reproductive success was observed. Indeed nearly all the relevant chi-square tests in Figures 5 to 7 were significant even after Bonferroni corrections. In the last two groups in particular, many parents did not appear to be assigned any offspring or were assigned to very few. However, reproductive success was more evenly distributed among the parents of the AD PILOT group than any of the parents of the three other groups. This is not entirely unexpected. The AD PILOT progeny were mostly collected at a young age (after ~ 1 year) and their group was run as an experimental group with presumably fairly good control. In contrast, the AD, RP and NC broodstock groups of progeny analyzed here were collected as adults, as parents of the next generation (P2). The three broodstock groups and particularly the last two are therefore more representative of a typical mating design and reproductive success that could be seen in a real hatchery like SPA. The distribution of reproductive success for the fathers of the NC group is perhaps indicative of what actually happens in a hatchery where most offspring may be attributed to only one or a very small number of fathers (Perez-Enriquez *et al*., 1999; Sekino *et al*., 2003; Frost *et al*., 2006; Trippel *et al*., 2009). The distribution observed for the RP group with a larger number of males and a slightly better evenness of their reproductive success may be an indication that the geneticists, which were collaborating with SPA, were beginning to have an influence at the hatchery by encouraging them to the use more males to try and maintain a large genetic pool. None the less, it is noteworthy that a very large proportion of the females and males did not contribute, or contributed very little, to the next generation and that most of the next generation broodfish (the "P2" broodstock) selected out of RP and NC groups actually appear to have originated from quite a limited number of successful females (~10-14) and males (~4-6). In the following generations, crossing

the descendants of the males and females characterized by high reproductive success could lead to quick inbreeding increase over several generations.

4.3.2 Generation 2

Generation 2 females and males used in the "CRP2Y5" crosses could be evaluated as two different groups of generation 3 fish in a situation similar to that in generation 1 with the AD PILOT fish and the AD, RP and NC broodstock. The Group 1, 3 and 5 Y5 progeny were selected for size and sampled at a young age (after ~ 1 year) to identify parents producing the larger offspring so that these parents could be preferentially respawned in subsequent years. The Parents 3Y8 fish were sampled as adults three years later to be parents of the next generation (P3). The Parents 3 Y8 fish were chosen as the best fish (based on size and other phenotypic traits) among all the fish that had been produced in the "CRP2Y5". This group consisted of 11 spawning groups (Group 1 Y5, Group 2 Y5 and so on until Group 11 Y5) which were eventually mixed in the hatchery and raised together.

Large differences in reproductive success were observed for the "P2" females and males used in spawning group 1 of the "CRP2Y5" crosses associated with the Group 1Y5 progeny (Figure 8, top graphs). Three females and one male did not produce any offspring when assessed by the quasi-categorical pedigrees. The distributions appeared to be very different between the QCEP/QCLODP and FP for this group but this reflected to a large part in differences in sample size with 65 offspring allocated in QCEP and QCLODP compared to 143 under FP. When a normalized distribution was estimated to eliminate the effect of sample size (Figure 8, bottom graphs), the fractional distribution of reproductive success seemed to fall more in line with that seen under the quasi-categorical pedigrees. Under QCEP/QCLODP, the top 3 females among the 17 females that were used in the crosses (7F7E502837, 7F7F395A30 and 7F7F420272) produced ~ 42% of the offspring but produced only ~28% under FP. Two of the 8 males (7F7F345B43 and 7F7F47197C) produced ~ 52% of the offspring under QCEP/QCLODP

and ~ 36% under FP. Nonetheless as with the previous groups, obvious differences in reproductive success could be seen for some individuals. This was particularly striking for a few individuals without assigned progenies ("holes") under the quasi-categorical methods (three females and one male). This male without any assigned progeny under QCEP or QCLODP actually had the second highest (out of 8) reproductive success under the fractional pedigree and similarly two of these unassigned females were characterized by the 3rd and 6th highest reproductive success (out of 17).

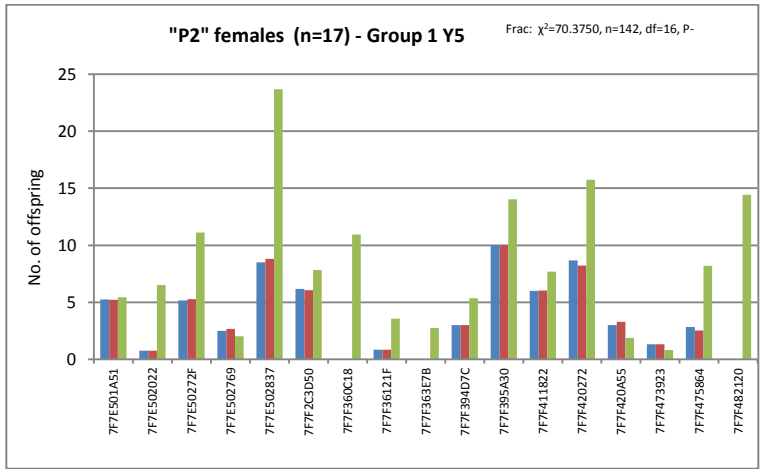
A larger pool of female (n=39) and male (n=11) parents had been used for spawning group 3 of the "CRP2Y5" crosses that were associated to the Group 3Y5 offspring (Figure 9, top graphs). Again, large, significant differences in reproductive success were observed for the females and males used in this spawning group, and some differences could also be seen between the quasi-categorical pedigrees and the fractional one. Nine females did not have any offspring allocated under QCEP/QCLODP but seven were subsequently assigned offspring with the fractional method (as expected). However, with one exception (female 7F7E502767) most of these had very small reproductive success and two remained without any offspring assigned in the fractional pedigree. All males were assigned some success by both methods, with fairly strong differences between the quasi-categorical and fractional pedigrees seen for three males. Although distribution of reproductive success appears to be more widely distributed across many females, only 6 of the 39 female fish are actually producing ~45 % of the offspring under the quasi-categorical pedigrees and 28% under the fractional pedigree. As well, only 4 of the 11 males are producing ~ 61% of the offspring under the quasi-categorical pedigrees and ~ 58% under the fractional pedigree. Even though this group used a larger number of parents for spawning compared to the other groups (i.e. spawning groups 1 and 5) still only very few individuals were are actually producing the majority of the offspring.

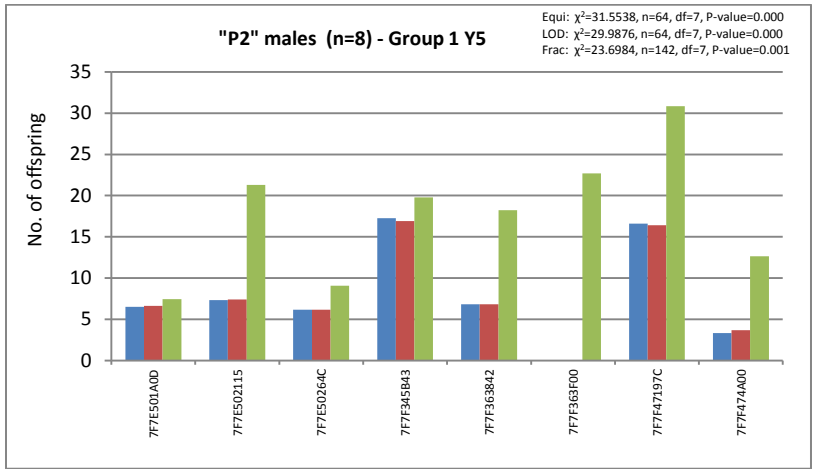
Spawning group 5 used a very small pool of parents (6 females and 4 males) and large, significant differences in reproductive success were observed for this group (Figure 9, bottom graphs). Interestingly, reproductive successes were nearly identical under the three pedigrees in this group. Offspring were assigned to all but one female parent by either allocation method. Clearly, one female (7F7F411663) and one male (7F7E640563)

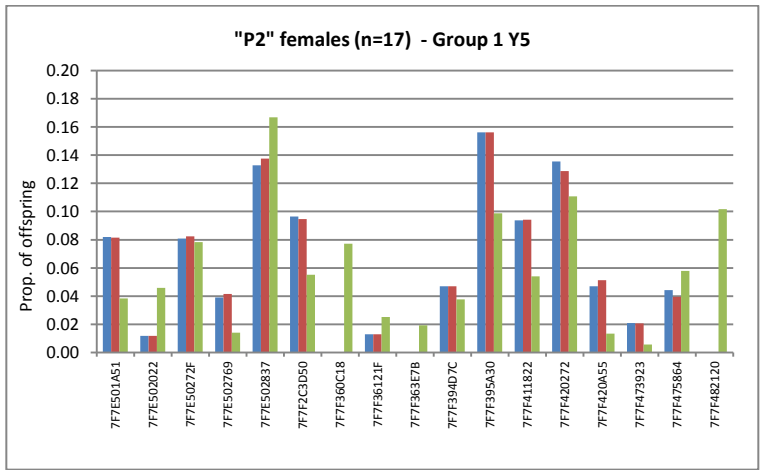
were the most successful parents from this group producing ~ 59% and ~ 63% of the offspring respectively.

Differences in reproductive success were observed for the females and males from the "CRP2Y8" crosses associated to the "Parents 3Y8" offspring, similar to what had been seen in the previous generation (Figure 10). The proportional distributions of reproductive success for the males and females were very similar and were not included for this group. Of the 126 females used in the "CRP2Y8" crosses, approximately 20% of the females (n=25) did not have any offspring assigned by the quasi-categorical pedigrees but offspring were assigned to 17 of them by the fractional pedigree, generally with very small reproductive success, and 8 females still had no reproductive success under the fractional pedigree. A large number of the females (n=53) contributed very little to the next generation being assigned to very few offspring (< 2). In the case of the males, very large differences in reproductive success were also observed. Again, few "P2" males (n=24) were used for this set of crosses as compared to the number of females, of which just over half (n=13) appear to have contributed at least two offspring to the next generation.

Figure 8: Reproductive success of the parents of Group 1Y5 (top graphs) and relative reproductive success (bottom graphs). Pedigree legend: blue-QCEP, red-QCLODP, green-FP.







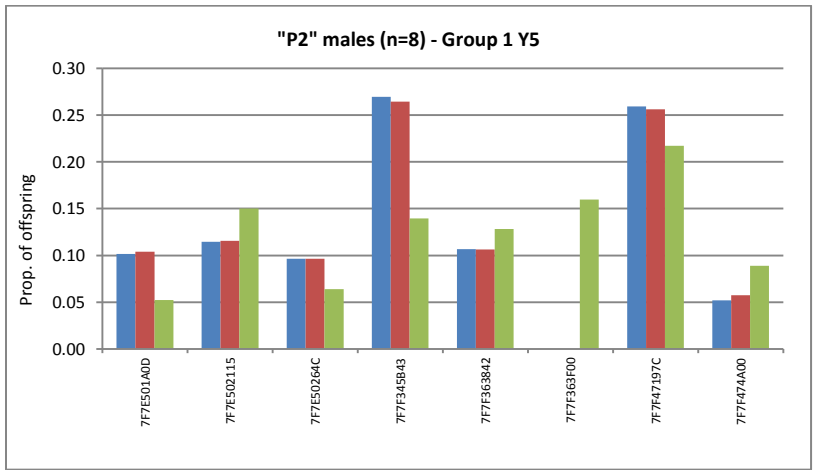
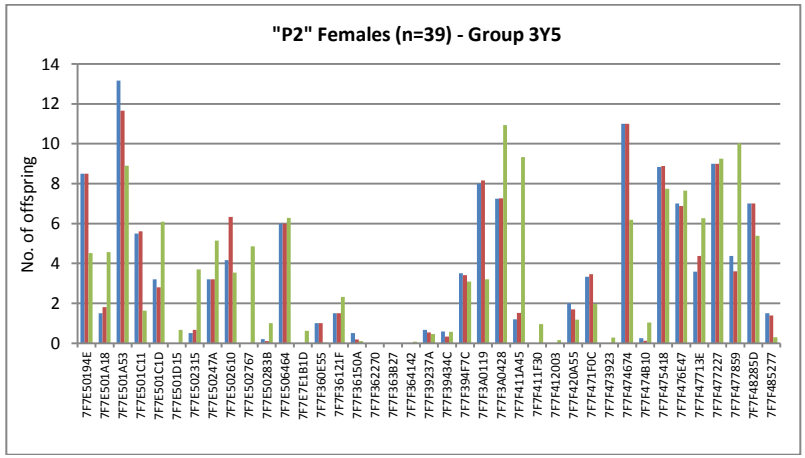
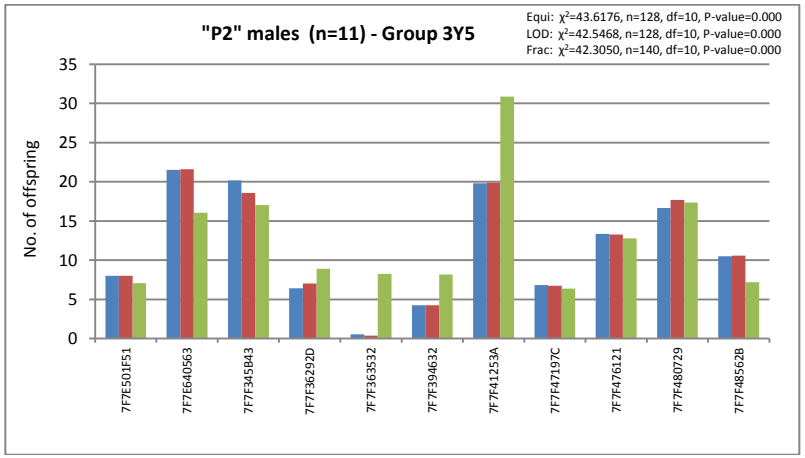
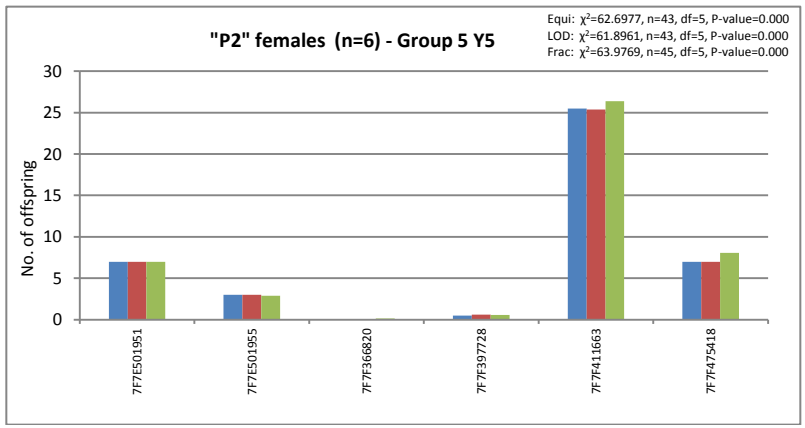
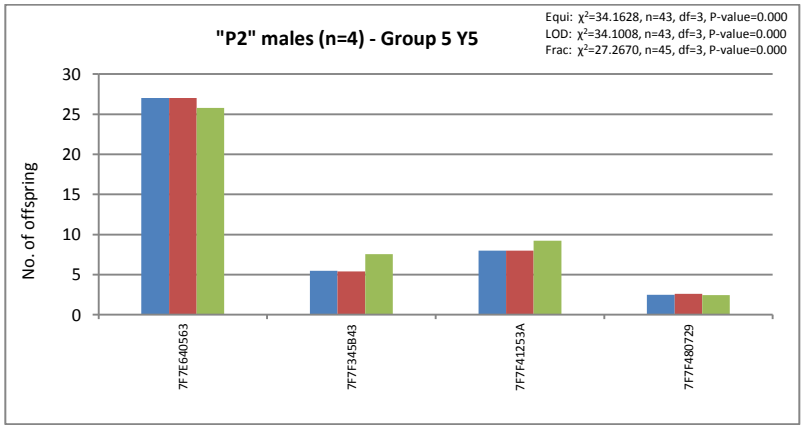


Figure 9: Reproductive success of the parents of Group 3Y5 (top graphs) and Group 5Y5 (bottom graphs). Pedigree legend: blue-QCEP, red-QCLODP, green-FP.









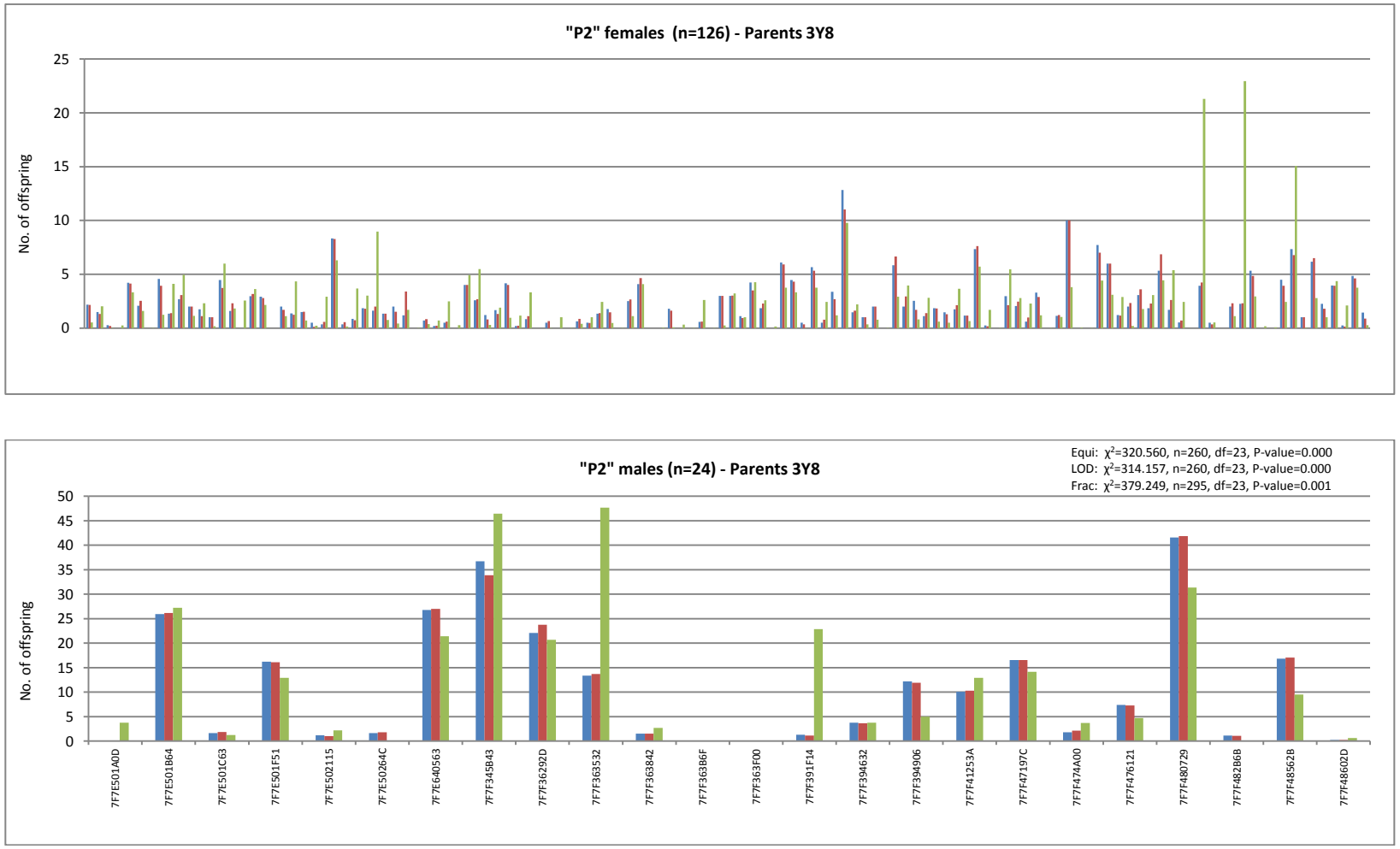
Two males (7F7F36352 and 7F7F391F14) and three females (7F7F477D12, 7F7F481354 and 7F7F482A3D) in particular had very different distributions with the fractional pedigree as compared to the categorical pedigree (Figure 10). For each of these fish the situation is similar to what had been seen earlier in that many more offspring were allocated with much lower probabilities again demonstrating the effect of quasi-categorical vs. fractional allocation. From a biological perspective, many females were used in the "CRP2Y8" crosses and quite a few males (n=24) were also selected as broodstock to try and maintain a good level of diversity. Although female reproductive success was more evenly distributed across a larger number of individuals compared to the previous generation, male reproductive success was still narrowly distributed with a few individuals contributing to a large portion of next generation

To sum up, in all groups, substantial differences in reproductive success were observed among sires and dams. The two quasi-categorical pedigrees gave very similar figures, a result not overtly surprising since the majority of offspring were categorically assigned to the same unique parental pair under both categorical pedigrees. Reproductive success under the fractional pedigree was often in good agreement with the categorical one for the majority of the dams and sires, but there were a number of individuals where both types of success were substantially different (in both directions). Many parents were not assigned any offspring under the quasi-categorical pedigrees and similarly many offspring were not successfully assigned parents. This created disconnection (holes) in the quasi-categorical pedigrees with the potential to underestimate inbreeding. This problem was largely resolved under the fractional pedigree with most offspring assigned to at least a parental pair.

Exclusion is very effective at determining reproductive success when one offspring can be categorically assigned to one set of parents or when there are few candidate parents and sufficient polymorphic markers available to allow a single mismatch to exclude a candidate parent. Exclusion is not as effective when very few markers are available and the data has a high degree of uncertainty due to scoring errors (Jones and Arden, 2003) because false exclusions due to scoring errors create disconnects

in a pedigree which can cause downward biases in over all estimates of reproductive success. From a biological point of view the fractional allocation approach is almost guaranteed to be false, since a progeny can only have one mother and one father. Yet the fractional method is considered by many to be the most accurate method for determining reproductive success in a population because, by assigning some fraction, between 0 and 1, of each progeny to all non-excluded candidate parents (Devlin *et al*., 1988; Jones and Arden, 2003) it averages reproductive success of parents to a potentially larger set of allocated offspring. Therefore, the fractional method is not, or is less, subject to biases of false exclusions or potential "holes" that can result from scoring errors that might influence a strictly exclusion based or categorical approach (Devlin *et al*., 1988; Jones and Arden, 2003). The fractional approach generates a more resolved estimate of reproductive success because it incorporates all levels of genetic relatedness (Smouse and Meagher, 1994). Nevertheless, it is important to note that the results of the reproductive success estimation will depend on the exact parameters used for pedigree estimation, both for the quasi-categorical and fractional methods. Even under the fractional method, using a different error rate or allowing different rules for considering what constitutes a fully matching locus between a progeny and a putative parent will result in a different list of probabilized trios, and therefore will yield potentially different estimated reproductive success.

Figure 10: Reproductive success of the parents of the Parents 3Y8 fish. Pedigree legend: blue: QCEP, red: QCLODP, green: FP.



These reproductive success findings illustrate as well the potential difficulty to manage inbreeding on a typical fish farm. In the very first generation, it was apparent that very few males might be used and could dominate the production of the following generation. Even after the start of this breeding program, and the discussion and modification of the mating procedures that were implemented to try to minimize the erosion of genetic diversity on the farm, it was apparent that there could still be a low number of males used, with some skew in reproductive success. The situation appeared more even on the female side. These findings are consistent with a study by Sekino *et al*. (2003) on a hatchery raised Japanese flounder (*Parlichthys olivaceus*). They observed that the contribution of candidate broodstock to the next generation was skewed by almost all of the offspring originating from one single male and half of the females did not produce any offspring. Frost *et al*., (2006) attributed a loss of genetic diversity at a barramundi (*Lates calcarifer*) hatchery to a limited number of effective breeders where 55% of the offspring from one cohort were found to be sired by a single male individual. Trippell *et al*. (2009) observed that of seven male haddock broodstock, one single dominant male was responsible for fertilizing the communally spawned eggs, which could lead to a high degree of inbreeding in the absence of additional broodstock. Interestingly, there was no apparent loss of genetic diversity over three generations with the population of rainbow trout as evaluated in this study by the number of alleles or heterozygosities (see section 4.1), despite an apparently overall reduced number of contributing parents (this section). This indicates that such metrics (allelic richness, heterozogosity) might be fairly coarse and would only allow detecting strong bottlenecks.

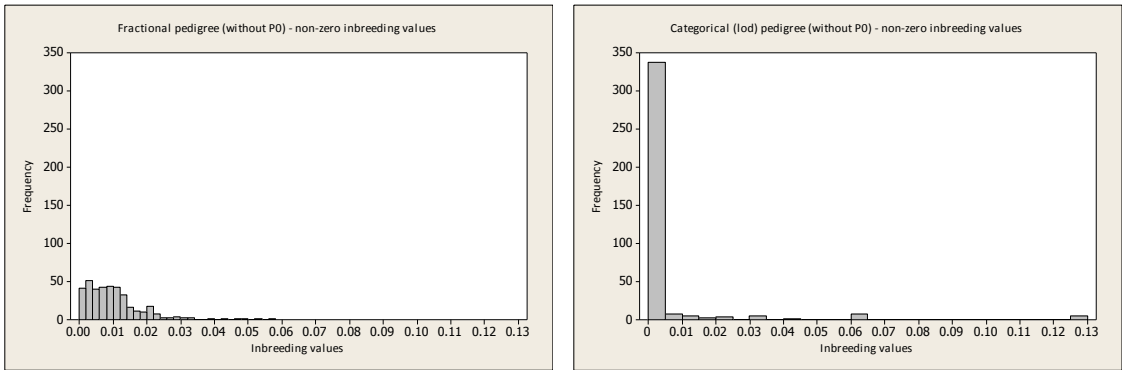
4.4 Inbreeding

Inbreeding values were estimated from the average numerator relationship matrix derived for each of the three different pedigrees for all individuals in generation 2 and generation 3 and compared (see Methods section 3.3.3). A hypothetical generation of phantom founders ,"P0", relating individuals of the "P1" generation by full-sib relationships, was also created using the pedigree reconstruction software COLONY Version 2.0 (Wang and Santure, 2009) and added to the pedigree. New inbreeding values were determined for the three types of pedigrees and compared to see how (or if) the addition of a fourth level to the pedigrees would influence the inbreeding values.

4.4.1 Estimation of Inbreeding Based on the Categorical Pedigrees.

No inbreeding was observed for the majority of the 1714 fish that were categorically (or nearly categorically) assigned to trios in the three generation pedigree. Under the two quasi-categorical pedigrees, only 37 individuals from generation 3 (~2%) were observed to have any level of inbreeding (Figure 11), and as expected, the estimated inbreeding values were very similar under the QCEP method (ranging from 0.003 to 0.125) and under the QCLODP method (0.005 to 0.125). Twenty-five fish were from the Parents 3Y8 group and 12 were from the Group 3Y5 group. The distribution of inbreeding values between the two quasi-categorical pedigrees were very similar as would be expected because they were based on the same number of trios (n=2753). Interestingly, the inclusion of genotypic information in the form of weighted averaged likelihoods (exponentials of trio LODs) under the QCLODP pedigree caused the inbreeding coefficient for some individuals to shift substantially compared to the value under QCEP. For example, one individual from the Parents 3Y8 group had an inbreeding value of 0.018 under the QCEP pedigree which shifted to 0.0625 under the QCLODP pedigree. This individual demonstrated the largest shift between the two categorical methods. Nonetheless, the average observed inbreeding values were very close to one another and very low under the two quasi-categorical methods.

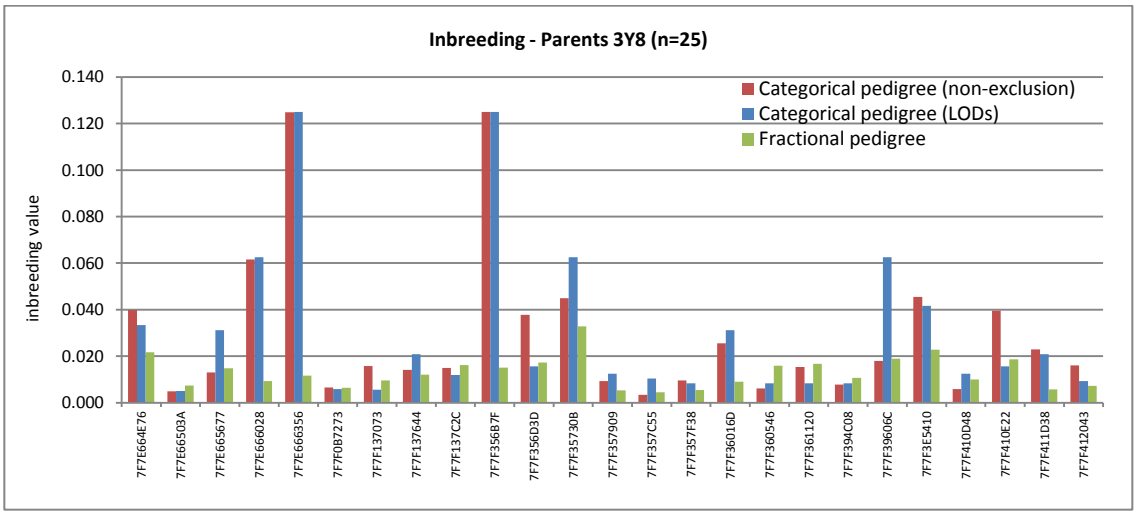
Figure 11: Individuals with non-zero inbreeding values (n=374) under the fractional pedigree (left graph) and corresponding values observed under QCLODP (right graph).



On the other hand, many more individuals (374) with some level of inbreeding were observed under the fractional pedigree, including the 37 observed under the categorical pedigrees (Figure 11). Inbreeding values estimated for the fractional pedigree range from 0.0000175 to 0.057.

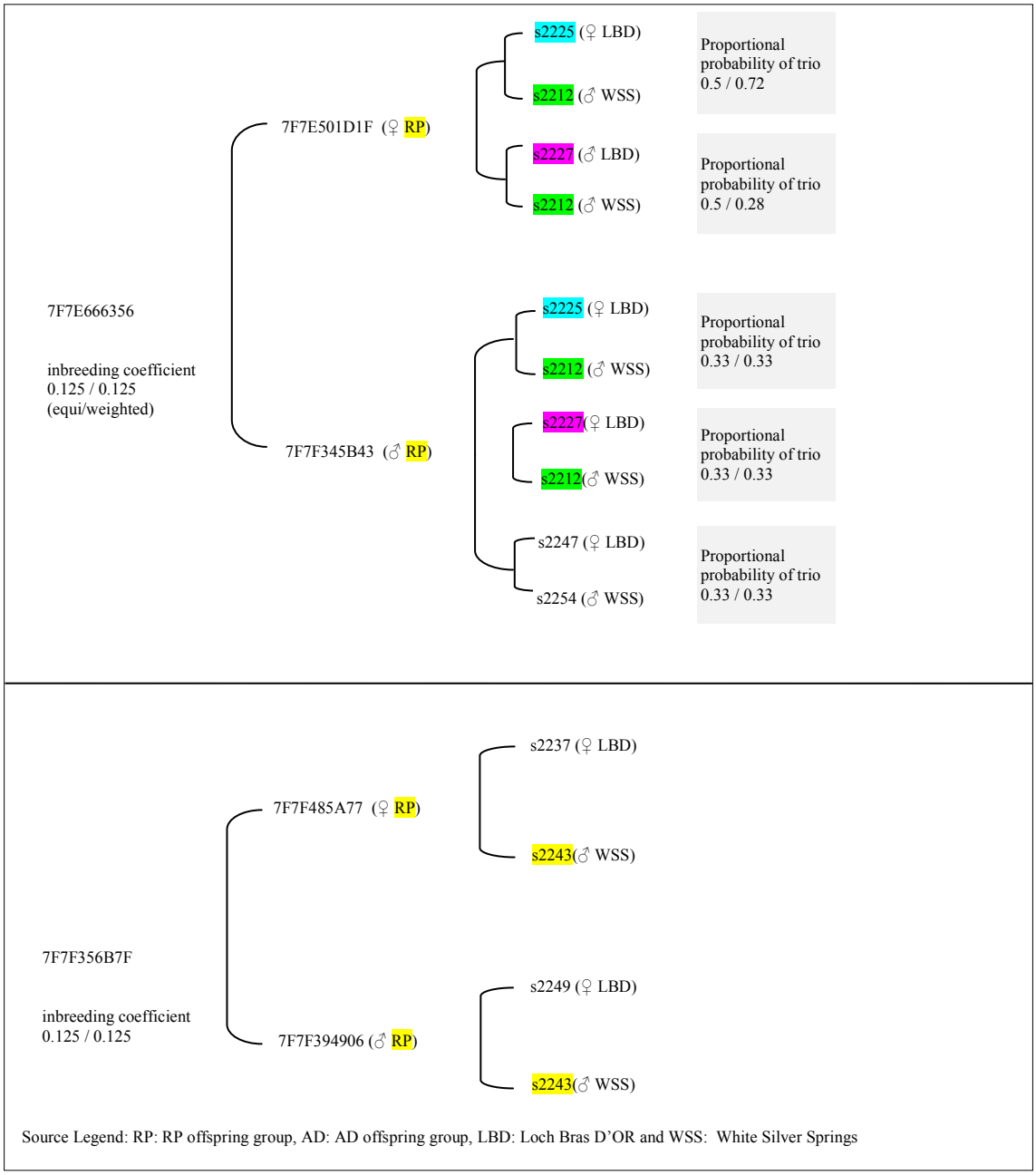
The 25 individuals from the Parents 3Y8 group observed to have non-zero levels of inbreeding under the quasi-categorical pedigrees were the most interesting for this study as some of these fish were potentially selected as next generation broodstock (Figure 12). Only two individuals (7F7E666356 and 7F7F356B7F) had inbreeding levels as high as 0.125, a level that could be seen as a result of half-sib matings. Inbreeding values for these same twenty-five individuals by the fractional pedigree were much lower ranging from 0.004 to 0.033 with inbreeding values for individuals 7F7E666356 and 7F7F356B7F of 0.012 and 0.015 respectively (~10 fold lower, Figure 12).

Figure 12: Comparison of non-zero inbreeding values obtained for the Parents 3Y8 group under the two quasi-categorical and fractional pedigrees.



The genealogies of the two parent fish (7F7E666356 and 7F7F356B7F) from the CRP3Y8 group with the highest inbreeding coefficients (0.125) by the quasi-categorical method are illustrated in Figure 13. Fish 7F7E666356 (top box) allocated to one set of parents with both male and female from the RP offspring group.. Both the mother and the father allocated back to two of the same sets of parents from "CRP1Y2" spawning group (group 4). The mother (7F7E501D1F) allocated to two sets of parents (s2225 x s2212 and s2227 x s2212) while the father (7F7F345B43) could not be excluded as having originated from three sets of parents (s2225 x s2212, s2227 x s2212 and s2247 x s2254). Hence these two parents could be unrelated (if the male actually came from the third set) or could be half-sibs or could be full sibs. When averaging over the 2 generations the proportional contributions from all of the possible sets of grandparents and parents, the expected inbreeding value of the offspring 7F7E666356 was 0.125. Individual 7F7F356B7F (bottom box in Figure 13) also had an expected inbreeding level of 0.125 under both QCEP and QCLODP, however in this case, this was not due to averaging over several possible pedigree loops. This fish was uniquely assigned to 2 parents, themselves uniquely assigned to a set of grand-parents including a common sire, i.e. the two parents were half-sibs. This mother/father combination had been given a "Black" rating in the

Figure 13: Categorical genealogy of two fish (7F7E666356 (top box) and 7F7F356B7F (bottom box) with proportional probabilities under QCEP (equally weighted) and under QCLODP (proportional probabilities weighted by likelihood).



spawning recommendations for the "CRP2Y5" crosses provided to the hatchery (see section 2) suggesting that the combination should have been avoided. Although the hatchery had been advised not to spawn fish from the same group (i.e. AD, NC or RP),

depending on the status of the breeders at the time of spawning, it was not always possible to avoid such mating.

Twelve individuals from the offspring group "Group 3Y5" also had non-zero observed inbreeding levels ranging from 0.006 to 0.125 under the QCEP and QCLODP pedigrees. Inbreeding values for these same individuals under the FP were much lower ranging from 0.005 to 0.031. Only three fish from this group had inbreeding coefficients as high as 0.125 under the two quasi-categorical pedigrees. Their parents were determined to be half-sibs. These three fish had inbreeding values of 0.01, 0.025 and 0.019 under the fractional pedigree.

These observations illustrate some of the consequence of using a fractional pedigree compared to a categorical (or quasi-categorical) pedigree and it can be seen as well in Figures 11 and 12. Under QCEP and QCLODP, the great majority of fish have zero inbreeding level, but a few fish have fairly high levels (up to 0.125). These latter fish have much lower inbreeding under the FP, but conversely a large number of fish that had zero inbreeding under quasi-categorical pedigrees now exhibit low but non-zero inbreeding. By averaging over many possible pedigrees, the fractional pedigree decreases the occurrence of extreme low (i.e. zero) or extreme high inbreeding levels that were seen under the quasi-categorical pedigrees. Overall, the average level of inbreeding observed in generation 3 was estimated to be 0.0028 under each of the quasi-categorical pedigrees and approximately two fold higher (0.0069) under the fractional pedigree. This illustrates another consequence of using a fractional pedigree approach. The increased average inbreeding under the FP is most likely a consequence of the fractional method reconnecting a large number of individuals to the rest of the pedigree, when these fish were not allocated to any trio under the quasi-categorical pedigree approaches (see section 4.2). As explained earlier, the unallocated fish ("holes") had to be treated as unrelated founders in the QCEP or QCLODP pedigrees. If these individuals were actually part of the pedigree and simply were not allocated into trios because of genotype issues, the presence of these holes would artificially decrease the overall inbreeding level since all pedigree loops including these individuals would be severed. Under the fractional approach, most of these individuals were reconnected and the generation 3 average

inbreeding more than doubled. In other words, the lower average inbreeding level observed under the quasi-categorical pedigrees was probably a consequence of genotyping error, and the higher average inbreeding level observed under the fractional pedigree was probably closer to the real level.

Tolerable levels of inbreeding are not well defined for fish species. Perez-Enriquez *et al*. (1999) observed low levels of estimated inbreeding values of < 0.8% in a hatchery reared stock of red sea bream after one generation of breeding from a relatively small number of contributing parents. Pante *et al*., (2001) observed rates of inbreeding of 1.3% per generation or 0.4% per year that were calculated from pedigree information in three populations of rainbow trout under selection for six generations. Even under the more credible FP, the average inbreeding values determined for the 3 generations of rainbow trout examined for this study were well beneath these levels and would not appear to be cause for concerns.

4.4.2 Reconstruction of the "P0" Founders.

The previous analysis assumed that the "P1" founders were all unrelated and non-inbred. This is a standard assumption in these sorts of calculation. In reality, this was clearly not the case as the SPA population was a relatively small closed population. The reproductive success analyses (section 4.3) showed in several instances that the number of parents, particularly the sires, that were used before the start of the SPA breeding program was probably quite limited. Hence, a substantial portion of the "P1" parents could well be related if they originated from a limited pool of "P0" parents, and this could impact the estimation of inbreeding levels. This possibility is explored in the following section.

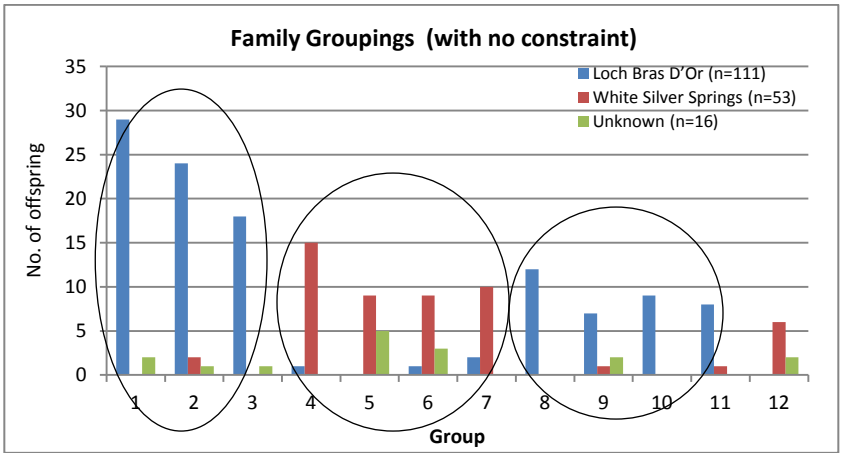
4.4.2.1 Partitioning of "P1" Individuals

Genotypic data from four microsatellite loci were used in the reconstruction of the unobserved "P0" founders from the 224 "P1" individuals in the rainbow trout data set. From the hatchery documentation, 124 fish were thought to have originated a few generations ago from the White Silver Springs (WSS) strain, 75 from the Loch Bras D'Or (LBD) strain and 25 were of unknown origin. Initial mating plans developed for the" CRP1Y2" crosses and the "10 x 10" factorial crosses were designed to avoid crossing parents originating from the same strain or source, to avoiding mating between potentially related fish in order to minimize the accumulation of inbreeding. The assumption was that the two different groups of fish in the SPA hatchery originated from different sources, were not closely related, and that the spawning in previous generations have been performed within strain, with little cross-contamination. However, documentation was not available from the hatchery to verify these claims or that these fish did not originate from the same stock several generations ago.

Initially, the "P1" data set was assessed using the software PEDIGREE Ver. 2.2 (Herbinger, 2006) without family constraint, to see if they would partition into groups of genetically related individuals (kins). The 224 "P1" fish were partitioned into ~24 kin

groups, ranging in size from 1 to 32 individuals. The majority of the fish were incorporated into 12 larger family or kin groups (Figure 14). Most of the Loch Bras D'Or (LBD) fish (111 of 124 represented as blue) tended to cluster together in seven family groups with only 4 WSS fish mixed in these kin groups. The majority of the White Silver Spring (WSS) fish (53 of 75 represented as red) partitioned into five different family groups, again with only 4 LBD fish mixed in. The remaining fish grouped into smaller less credible groups of only 1 to 3 individuals that could have grouped together by chance. The fish of "unknown" origin (represented by green) were presumably from these same two strains but had lost their identification. This seemed consistent with the observation that they seemed to be randomly distributed across almost all family groups (Figure 14). The LBD and WSS family clusters were large and showed very little overlap. This would support the idea that the "P1" fish did indeed originate from a limited number of crosses within each of the two groups.

Figure 14: Partitioning of "P1" fish (n=224) into family groups by source (Loch Bras D'Or, White Silver Springs or unknown) using PEDIGREE.



4.4.2.2 Reconstruction of “P0” using COLONY

Since it was known that in the SPA hatchery crosses, both male and female broodfish were typically used across several partners, the “P1” data set was also evaluated using the software COLONY Version 2.0 (Wang and Santure, 2009) allowing for polygamy of both sexes. This is something that cannot be done easily with PEDIGREE. The “P1’s” clustered into large and small full-sib and half-sib families as illustrated in Figure 15. Nineteen “Parent 1’s” and 23 “Parent 2’s” were reconstructed (these designations are representing the two parents but the actual gender is unknown, i.e. Parents 1 could be the sires and Parents 2 the dams or the converse). The reconstruction of “P0” resulted in family clusters ranging in size from 1 to 11 offspring per family. However, 71 of these family groups had only one offspring assigned which is not very informative so these very small groups were essentially disregarded. The majority of the Loch Bras D’Or fish (~90%) appeared to cluster together, originating from mating of 6 “Parent 1’s” and 8 “Parent 2’s” (Figure 15, upper left circle). Similarly, the majority of the White Silver Springs fish (~70%) also appeared to cluster together (Figure 15, middle circle) and originated from mating among 6 different “Parent 1’s” and 10 different “Parent 2’s”.

The majority of the 224 “P1” fish appear to have originated from a limited number of parents (~30). The “hypothetical” reconstructed “P0” generated by COLONY for each of the 224 “P1” individuals was then added as a layer on top of the QCEP, QCLODP and FP pedigrees and three new pedigrees were created and evaluated in order to compare inbreeding values with and without the addition of “P0” individuals. It is important to note that this extra pedigree layer linking the observed “P1” fish to their hypothesized “P0” parents was strictly categorical. Each “P1” parent was assigned to exactly one trio involving a pair of ghost "P0" parents, as indicated in the COLONY analysis. Even though there is clearly a considerable amount of uncertainty and imprecision around the sibship reconstruction of these “P1” fish, particularly considering the limited amount of genetic information available, there is at present no method available that could generate a fractional pedigree for these fish in the absence of DNA samples from the “P0” parents.

Figure 15: Distribution of sibships of "P1" fish (n=224) to reconstructed parents using COLONY, where P1= reconstructed Parent 1, P2=reconstructed Parent 2, pink row=fish from White Silver Springs, yellow row=fish from Loch Bras D'Or and white row=origin of fish is unknown (e.g. 11 fish from Loch Bras D'Or were assigned to reconstructed P1-4 and reconstructed P2- 9).

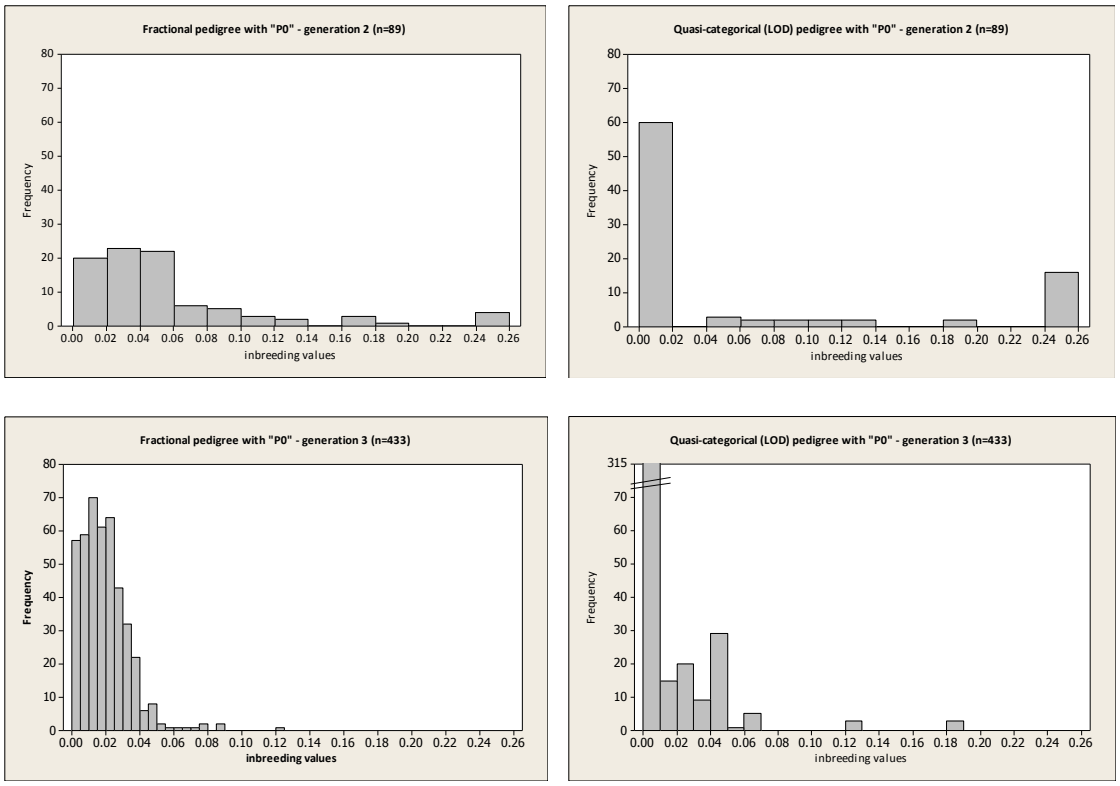
P1	P2																							#
	9	14	4	3	5	2	7	12	8	10	1	15	11	17	13	19	6	16	22	18	21	20	23	
	0	0	0	0	0	0	0					0										0		0
4	11	4	10	1	5	4	4					1										1		41
	0	2	0	0	0	0	0					0										0		2
	2	0	0	0	0		0	0								0								2
5	10	4	3	3	1		2	1								1								25
	0	0	0	0	0		0	0								0								0
	0	0	0	0	0	0							1	0										1
2	1	4	4	2	1	4							0	1										17
	1	0	0	0	0	0							0	0										1
		0	0	0			0											0			0			0
3		6	1	2			1											1			2			13
		0	0	0			0											0			0			0
	0					0		0								1						1	0	2
10	0					2		6								0						0	0	8
	1					0		0								0						0	1	2
		0		0	0		0																	0
8		3		1	2		2																	8
		0		0	0		1																	1
									3		1	3		2	1	2	2							14
11									0		0	0		0	0	0	1							1
									0		0	0		0	0	0	0							0
				0					1	3	0	2	3		1	1	0	1			1			13
9				0					0	0	0	0	0		0	0	0	0			0			0
				1					0	1	1	1	1		0	0	1	0			0			6
										3	2	0	1						2					8
7										0	0	0	0						0					0
										0	0	1	0						0					1
									2		2			0	2				1	0				7
15									1		0			1	0				0	1				3
									0		0			0	0				0	0				0
			1						1		1		0	2	1		0	0						6
6			0						0		0		0	0	0		1	1						2
			0						0		0		1	1	1		0	0						3
				1						2	1				1		0							5
1				0						0	0				0		0							0
				0						0	0				0		1							1
				0	0			1	1					1				0		1	0			4
16				1	1			0	0					0				0		0	0			2
				0	0			0	0					0				1		0	1			2
			0						0				1	1			1		1					4
12			1						0				0	0			0		0					1
			0						1				0	0			1		0					2
					1							1				0		1					1	4
14					0							0				1		0					1	1
					0							0				0		0					0	0
			0															1	0					1
13			0															0	0					0
			1															0	1					2
								0				1				1	1							3
17								0				0				0	0							0
								1				0				0	0							1
		0	1																					1
18		0	0																					0
		1	0																					1
	0							0																0
19	1							1																2
	0							0																0
	2	0	2	1	1	0	0	1	8	8	7	7	6	6	6	5	4	3	4	1	1	1	1	75
#	23	21	19	10	10	10	9	8	1	0	0	1	0	2	0	2	2	2	0	1	2	1	0	124
	2	3	1	1	0	0	1	1	1	1	1	2	2	1	1	0	3	1	1	0	1	0	1	25

4.4.3 Inbreeding Values with the Addition of "P0".

With the addition of "P0" founders, inbreeding was observed for 522 fish under the fractional pedigree including 89 fish from generation 2 (~6%) and 433 from generation 3 (~67%) as compared to 0 and 374 such fish respectively under FP without the "P0" layer added. The distribution of the "non-zero" inbreeding values observed for generations 2 and 3 from the fractional pedigree are illustrated in Figure 16 (left graphs).The majority of the 89 generation 2 fish and 433 generation 3 fish exhibited very low levels of inbreeding ranging from 0.00009 to 0.06. Only four fish from generation 2 were observed to have inbreeding values as high as 0.25.

With the addition of the "P0" layer, non-zero inbreeding was now observed in a substantially larger number of fish under the quasi-categorical pedigrees (29 from generation 2 and 118 from generation 3), when this had been the case for only 37 generation 3 fish under the pedigree without the "P0" layer. The distribution of the "non-zero" inbreeding values observed for generations 2 and 3 under QCLODP are illustrated in Figure 16 (right graphs). Of the 89 fish in generation 2 that had "non-zero" inbreeding values by the fractional pedigree, 60 had inbreeding values of 0, but 16 individuals had individual levels of 0.25 (i.e. the level resulting from full-sib mating) and another 6 fish also had very substantial inbreeding ranging from ~0.1 to ~0.2 (Figure 16, top right graph). This is the same sort of observations that was made earlier (section 4.4.1). Many individuals with very low (i.e. zero) or very high inbreeding (i.e. inbreeding of ~0.25) under quasi-categorical pedigrees have less extreme values under the fractional pedigree. The same sort of trend can be seen in generation 3 (Figure 16, bottom right graph). Of the 433 fish observed that had "non-zero" inbreeding values from the fractional pedigree, 315 had inbreeding values of 0, but a fair number of individuals exhibited pretty high inbreeding level under the quasi-categorical pedigrees, higher than was seen under the fractional pedigree.

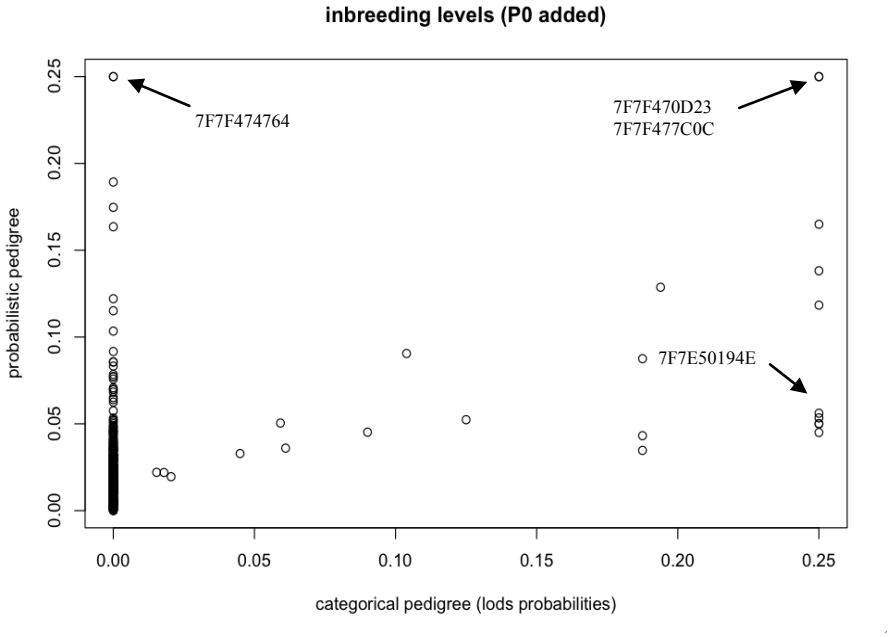
Figure 16: Distribution of the "non-zero" inbreeding values observed in generations 2 (n=89) and 3 (n=433) under the fractional pedigree with the addition of "P0" (left graphs) and corresponding distributions under the quasi-categorical (LOD) pedigree (right graphs).



Interestingly, in some instances individual inbreeding values were higher under the fractional method than under the quasi-categorical ones and in some instances the reverse was seen (Figure 17). For example, one individual (7F7F474764) had inbreeding values of 0.25 under FP and 0 under either QCEP or QCLODP. This individual was a "P2" fish that did not allocate to any "P1" parents by the categorical method, hence the inbreeding value of zero, but it was associated to 21 offspring. This fish presented a disconnect between the "P1" and "P3" fish under the quasi-categorical pedigrees which would lead to an underestimation of the inbreeding values of its offspring. However, this fish allocated to 22 sets of parents under the fractional pedigree essentially repairing the "hole" and providing a more accurate level of inbreeding for this fish and its offspring. In contrast fish 7F7E50194E had inbreeding level of 0.25 under either QCEP or QCLODP

but only about 0.05 under FP. Under the two first pedigrees, this individual was categorically assigned to a pair of "P1" parents that were full sibs. Under FP, this fish was allocated to many trios and its inbreeding level was considerably diluted when averaged over many trios.

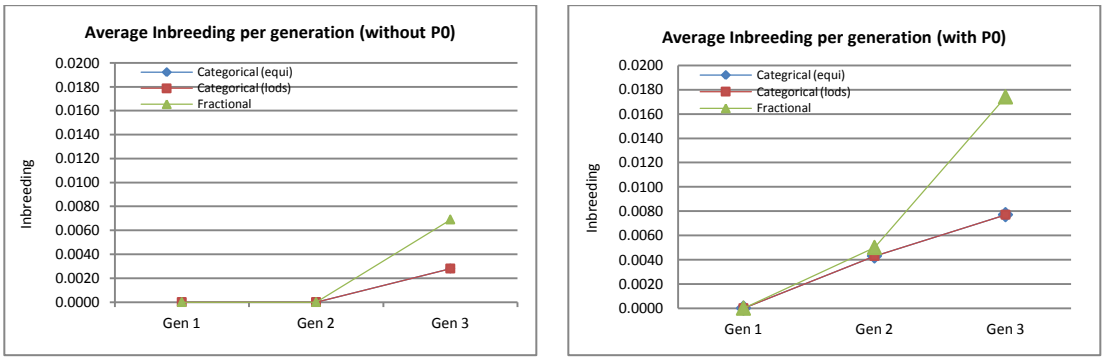
Figure 17: The distribution of inbreeding values (with P0 added) between QCLODP and FP.



Average population inbreeding levels observed under QCEP and QCLODP (without "P0") were very low and essentially the same (~0.3%) and were found to be approximately twice as high under the fractional pedigree (~0.7%). As expected, with the addition of "P0", the inbreeding level increased for both types of pedigrees. Some inbreeding was now observed in generation 2 with inbreeding values of ~ 0.4% under both quasi-categorical pedigrees and ~ 0.5% under the fractional pedigree (Figure 18). For generation 3, compared to what had been observed without the "P0" layer, the average inbreeding values increased more than two fold to ~ 0.8% under QCEP and QCLODP and to ~ 1.7% under the fractional pedigree. This is not unexpected as the "P1"

parents appeared to have come from as few as 30 “P0” parents. With the "P0" layer added, the inbreeding level was substantially higher under FP than under either QCEP or QCLODP, just like it was in the pedigrees without the "P0" layer. Undoubtedly this was due to the large number of individuals that became reconnected under the fractional pedigree, compounded by the fact that with the “P0” layer, there were now more potential pedigree loops in the upstream generations that could contribute to individual inbreeding. A number of individuals in generation 2 (n=29) were allocated one or more “P3” offspring, but did not have their own “P1” parents recognized though both QCEP and QCLODP. For example, 7F7F480729 (an RP male) was associated to 144 offspring but did not allocate to any parents by exclusion and represented the largest potential hole in the two quasi-categorical pedigrees. Under the fractional method, this fish allocated to 5 sets of "P1" parent trios thus reconnecting the pedigree loops with the result that the inbreeding of its offspring increasing. The lower estimates of inbreeding observed with the categorical pedigrees were thus most probably due to the disconnect (holes) between generation 1 and 2 resulting in an underestimation of inbreeding when the “P0” layer was added. This problem was largely resolved under the fractional pedigree with most offspring assigned to at least one parental pair.

Figure 18: Average inbreeding per generation without “P0” (left graph) and with “P0” (right graph).



In summary, initial levels of inbreeding observed in generation 3 (without the addition of "P0") were minimal in all three pedigrees. As expected, with the addition of "P0", some fairly low levels of inbreeding became evident in generation 2 (under all pedigrees). The average level of inbreeding detected in generation 3 under the quasi-categorical pedigrees increased to ~0.8% and resulted from very few individuals having moderate levels of inbreeding. The average inbreeding levels under the fractional pedigree increased to ~ 1.7% and resulted from many individuals having fairly low levels of inbreeding. Even though such an average inbreeding level is still tolerable from a hatchery genetic management point of view, it is nonetheless getting closer to the level where concerns for inbreeding depression could be justifiable. Additionally, many fish were showing high individual inbreeding levels in generation 2 and 3 under QCEP/QCLODP and even under FP. These specific fish could demonstrate inbreeding depression. Adding a single credible reconstructed layer of "P0" parents resulted in a significant jump in the inbreeding estimated under FP. This is a consequence of the fairly small effective number of broodfish that seems to be a common feature of many hatchery, including SPA. By extension, then, it is probable that the same type of situation took place in the generations preceeding "P0". As compared to an hypothetical population(s) that founded the SPA population a few generation ago, the inbreeding rate increase achieved by the "P3" generation might be substantial.

Although dangerous values for inbreeding levels are not well defined for fish, inbreeding levels of 11% were observed in three populations of rainbow trout under selection for six generations without significant loss of genetic diversity (Pante *et al*., 2001). The average inbreeding values estimated for the SPA population are well below the values observed by Pante *et al*. (2001) and the spawning strategy designed for the SPA hatchery was successful in containing rapid inbreeding increase (see next section). However, a slow and constant increase over several generations due to a limited number of contributing parents could result in substantial inbreeding levels leading to inbreeding depression if not properly monitored. For most species, an increase in the rate of inbreeding will result in inbreeding depression (i.e. a reduced fitness of important

performance traits such as growth, viability and survival) and may cause potentially deleterious abnormalities (Falconer and Mackay, 1996; Dube and Mason, 1995).

Gallardo *et al*. (2004) observed fairly high levels of inbreeding (9.3% and 4.3%) after four generations of selection for weight at harvest in two populations of Coho salmon (*Oncorhynchus kisutch*). Inbreeding depression was noted for two female reproductive traits. Inbreeding was significantly associated with reduced the gonadosomatic index in one population and body length at spawning for the second population; however other traits such as body weight at spawning and relative fecundity were not affected. Significant inbreeding depression in performance traits of yield, individual growth rate and survival was observed after the first growing season in families of Pacific oysters (*Crassostrea gigas*) with expected inbreeding coefficients (F) of 0.203 but not in families with expected inbreeding coefficients of 0.0625 (Evans *et al*., 2004). After two growing seasons, significant inbreeding depression of yield and individual growth rate were observed in some families with expected F as low as 0.0625 but depressed survival at harvest was observed only in families with F=0.203. These results indicate the importance of maintaining pedigree records in breeding programs to help avoid the deleterious effects of inbreeding depression.

Overall, the average inbreeding values (and many individual inbreeding values as well) estimated from the quasi-categorical pedigrees were probably underestimated due to potential disconnects (holes) resulting from genotyping/scoring errors. By using all the information and integrating over the uncertainty, the fractional pedigree approach mostly resolved that problem and would appear to be less biased. Yet it was obvious as well that for some individuals, inbreeding levels could be much higher under quasi-categorical assignment than under fractional allocation, because the effect of possible inbreeding loops was diluted over many possible trios.

4.5 Retrospective Analysis of Two Components of the SPA Breeding Program.

Another component of the present thesis was to evaluate retrospectively two aspects of the farm-based genetic improvement project that had taken place but had not been previously analyzed. The two questions to be addressed with this retrospective analysis were: 1) to see if the spawning recommendations provided to the hatchery at the onset of the 92/93 spawning season were efficient in minimizing the accumulation of inbreeding, and 2) to see if the sampling and parentage determination of the large and small juvenile fish from the AD PILOT group and offspring Groups 1, 3 and 5 Y5 would have been successful in identifying parents producing larger offspring, so that these potentially superior parents could be used again in the next spawning season.

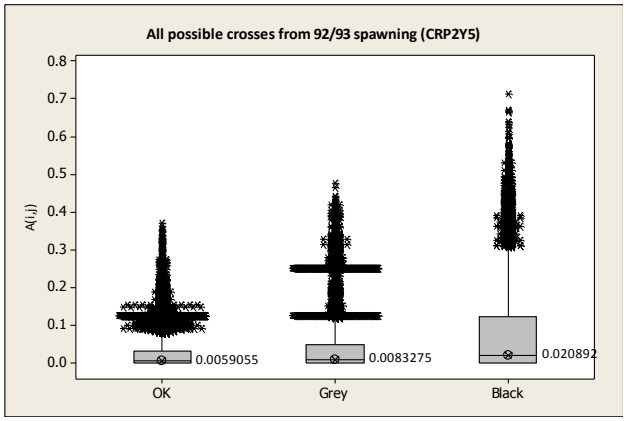
4.5.1 Evaluation of Spawning Recommendations.

In the initial stages of the SPA breeding program, the genotypes of putative pairs of parents were assessed to see if specific pairs of fish might be related according to a very rudimentary scoring scheme as outlined in the Methods section 3.3.4. The different putative pairs were evaluated as “OK” “Grey” or “Black” with “OK” crosses being preferred whenever possible, “Grey” crosses being recommended only when no “OK” pairing was available for a specific broodfish, and “Black” crosses to be avoided. Over the course of the spawning season, as fish were maturing and ready for spawning, recommendations from the biologists were returned to the hatchery and, depending on who was ready at the time, fish were spawned in small pools in a way that would hopefully avoid crossing possibly related fish. This exercise was very cumbersome and time consuming at the time and the actual value of this exercise was never fully evaluated. This section will assess, in hind sight, if the spawning recommendations sent to SPA were successful in limiting inbreeding increase. To answer this question, a retrospective analysis was done to compare 1) the co-ancestry of every possible pair of “P2” parents with the “OK”, “Grey” and “Black” classification of these pairs of fish and 2) the co-ancestry of the actual pairs of fish that were spawned in 92/93 with their “OK”,

"Grey" and "Black" classification. This was done by estimating the coefficient of relationship ($A_{(i,j)}$) of; 1) every possible pair of parents and 2) all realized mating based on the average numerator relationship matrix ($\bar{A}$) derived from the fractional pedigree with the "P0" layer added so as to maximize the amount of information about genetic relatedness in the pedigree. The coefficient of relationship ($A_{(i,j)}$) is two times the co-ancestry of a pair of individuals and can be interpreted as twice the expected inbreeding value of the progenies of this pair of individuals, if they have progeny (Falconer and Mackay, 1996) and so the inbreeding values of any progeny of a pair of parents would be ½ their coefficient of relationship.

The coefficient of relationship for all possible crosses that could have been performed in 92/93 according to the scoring scheme "OK", "Grey" or "Black" is represented in Figure 19, estimated over the 110215 putative pairs that could result from crossings between all "P2" individuals with appropriate information (n=470). The scoring scheme was based on two parameters; 1) the group identity of the parent fish (AD, NC or RP) and 2) the amount of sharing of alleles detected at Omy 77 (and Pupupy if available), in the same manner as the recommendations that had been provided to SPA as described in the Methods section. Essentially, crossing within groups was discouraged as there was a greater chance of common parentage and crossing fish with different alleles was encourages as there was a less likely chance that they were related.

Figure 19: Box plots of the coefficient of relationship ($A_{(i,j)}$) for all possible crosses that could have been performed in 92/93.

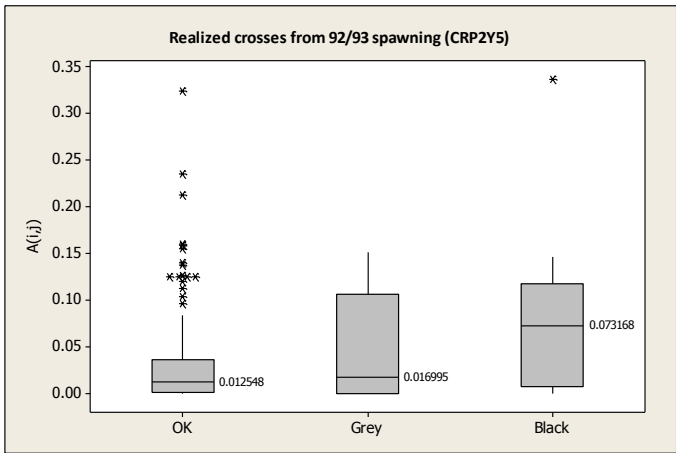


The median values of $A_{(i,j)}$ were very low for the pairs that received "OK" or "Grey" scores, 0.006 and 0.008 respectively (Figure 18). Because the inbreeding values are ½ the value of $A_{(i,j)}$, the expected average inbreeding of any offspring obtained from any of the pairs of parents that had received an "OK" score would be estimated to be very small (< 0.003). Likewise, progeny resulting from a "Grey" couple would on average have an inbreeding value of ~ 0.004. On the other hand, progeny from the couples that scored "Black" would on average produce offspring with average inbreeding values of ~0.01, more than two or three times those of the two other groups. The ratings of crosses in three categories thus appeared to capture some information about risk of inbreeding increase. Both "OK" and "Grey" crosses would, on average, be acceptable choices, while "Black" crosses would have been on average slightly more risky. This last group represented mostly within group crosses and individuals with a lot of allele sharing, while the "OK" crosses should have been representative of crosses of fish from different groups (AD, NC or RP). It is thus not unexpected that "Black" pairs should have a higher probability of being related than "OK" pairs and the simple scheme applied in 1992/1993 appeared to have had some value.

However, it is notable that the tails for each of the distributions are very long (Figure 19). Even with an "OK" or "Grey" score it was possible to have offspring with substantial inbreeding values in the order of 0.05 to 0.2. A possible explanation for this is that the "P1" parents were found to be related through the inclusion of a reconstructed "P0" layer. In addition, quite a few males were used in more than one spawning group in generation 1, as there were a limited number of males available. For example, six of the males used in the "10 x 10" crosses that produced the AD broodstock were also used in the "CRP1Y2" crosses that produced the RP broodstock. Hence a "P2" parent from the RP group could be a paternal half-sib of a "P2" Parent from the AD group. Yet if these individuals did not share any or shared only one allele at locus Omy 77, they would still be classified as an "OK" cross. This clearly shows that the common hatchery practice of using (and re-using) a relatively limited number of males (in the inferred "P0" layer and in the "P1") could easily lead to an inbred cross even when efforts were exerted to try to minimize such crosses.

The second phase involved looking at the estimated average inbreeding values for offspring of crosses that were actually performed by SPA in order to assess the efficiency of the spawning recommendations that had been provided. Approximately 260 crosses were performed in 92/93 however screening scores were only available for 206 crosses because genotype information was missing for some of the parents. The scores for 206 crosses were recorded of which 169 received and "OK" rating, 20 a "Grey" rating and 17 a "Black" rating. The fairly small number of "Black" and "Grey" crosses that had been done compared to the "OK" crosses is an indication that the hatchery team tried to use the recommendation to the greatest extent possible. However they had to resort in some cases to less favorable pairing because of the specific fish that may have been mature and available on a given spawning date. As in the previous figure, $A(_{i,j})$ was generated from the average numerator relationship matrix of the fractional pedigree with "P0", and plotted against the recommended spawning score ("OK", "Grey" or "Black") that was generated for each pair (Figure 20).

Figure 20: Box plots of coefficient of relationship, $A(_{i,j})$ of the spawning recommendations for the crosses that were performed in 92/93.



The average inbreeding values realized for the crosses that were performed in 92/93 were estimated to be 0.005 (estimated from half of the median value) for the "OK"

group, 0.008 for the “Grey” group and 0.036 for the “Black”. This indicates again that the few “Black” crosses that were performed were indeed carrying a substantial risk of inbreeding increase. A few outliers were noted from the “OK” group with values observed to be almost 10 fold higher than the median. One pair in particular was identified whose offspring would have an expected inbreeding value as high as 0.1625. These few higher values were observed for fish whose parents were actually half-sibs because of the use of some of the “P1” males across different spawning groups. The median for the “Grey” crosses was again a little higher than that of the “OK” crosses but these pairs were not so related that their mating would greatly affect their offspring inbreeding levels. In contrast, although the hatchery only performed 17 "Black" crosses, these carried a substantial level of inbreeding, ~ 4 times higher than that of the “OK” group.

It is notable that the median values observed for the realized crosses of the three types are at least twice as high as those observed from all possible crosses (Figures 19 and 20). Although these estimates are based on a much smaller number of realized crosses, it would appear that the realized crosses were not a random sample of all possible crosses. A possible explanation is that fish that were mature and ready to spawn on a given date might have been more related than randomly drawn fish. Indeed there is strong evidence that for rainbow trout, timing of maturation within a spawning season has a genetic basis (Sakamoto *et al*., 1999; Fishback *et al*., 2000). If fish that were mature and ready to spawn on a given date had indeed a higher probability of being related, this would compound the potential rate of inbreeding increase due to the relatively limited number of males used in this hatchery and indeed in most typical hatcheries. This certainly supports the idea that any mechanism to minimize crossing of related individuals would have value in typical hatchery where the exact pedigree of fish is unknown. Overall, it appears that the low levels of expected inbreeding observed for the pairs of parents that scored “OK” or "Grey" from the realized crosses of the “CRP2Y5” crosses is a good indication that perhaps the simple rule used here to provide recommendations to the hatchery was relatively effective.

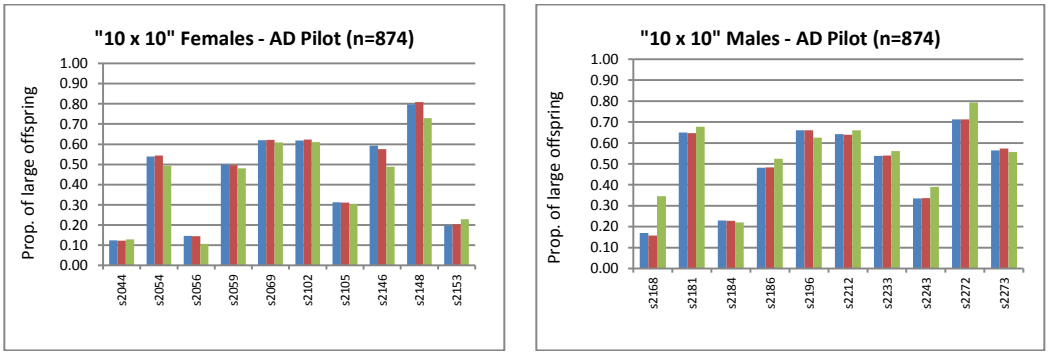
4.5.2 Assessment of "P2" Fish Producing the Largest Offspring.

The second of the objectives of this retrospective was to assess whether some broodstock were producing larger progeny than others and whether they could be identified to be preferentially spawned again in subsequent years (see Methods section 3.3.5). This analysis was performed on the AD PILOT and Groups 1, 3 and 5Y5 fish that were sampled as juveniles, either as large (belonging to the 10% largest quantile) or small (belonging to the 10% smallest quantile).

Large and significant differences were observed among the females under each of the three pedigrees QCEP, QCLODP and FP (χ^2=146.55, 147.751 and 104.363 respectively, each with df=9 and P-values=0.000) and similarly for the males (χ^2=110.489, 108.687 and 70.387 respectively, each with df=9 and P-values=0.000) from the "10 x 10" crosses, for the size of offspring produced as estimated by the relative proportion of large offspring among all offspring (large + small). The proportion of large sized offspring (≥ 15.7 cm) attributable to each female and male was estimated as the number of large offspring (sum of the number of large offspring x the probability for that parent / reproductive success (sum of the number of offspring x the probability for that parent, as explained in section 4.3). This was done based on the three different lists of probabilized trios (Figure 21). Six females and seven males appear to have produced larger offspring than the rest. These findings are similar to those previously published (Herbinger *et al*., 1995).

Interestingly, despite fairly large differences in total reproductive success estimated from the fractional pedigree compared to the two nearly categorical pedigrees (see figure 5 in section 4.3), the relative proportions of large offspring based on the three pedigrees were extremely close for the females and for most of the males (Figure 20). There was a significant Spearman rank correlation for the proportion of large offspring under the QCEP and FP pedigrees for the females (r = 0.976, p-value < 0.001) and for the males (r=0.939, p-value < 0.001). This indicates that both the ranking of females and males in terms of proportion of large offspring produced in the next generation was quite robust and was independent of the type of pedigree used.

Figure 21: Proportion of large offspring (AD PILOT) attributable to each female (left graph) and male (right graph) from the "10 x 10" crosses. Pedigree legend: blue-QCEP, red-QCLODP and green-FP.



As with the AD PILOT group, differences in relative proportion of large and small offspring size were observed among the females and males parents from spawning groups 1, 3 and 5 of the "CRP2Y5" crosses (Figures 21 and 22). In the case of the male parents for the three groups and as well the female parents for Group 5 Y5, the proportion of large individuals attributed to each parent was fairly close under the three pedigrees. In these four cases, the number of parents was fairly small, ranging from 4 to 11. Total reproductive success was 10 or more progeny for about half of these parents and thus the proportion of large offspring was estimated on a reasonable number, but for the other half of these parents the total number of allocated progeny was too low to have much confidence in the estimated proportion of large offspring. This problem was exacerbated in the case of the female parents of the Group 1 Y5 and Group 3 Y5. In the first group only 1 female had more than 10 offspring under the categorical pedigree and 6 under the fractional one, while only two females had more than 10 progeny under either pedigree for the latter group. In addition, many females had progeny allocated under the fractional pedigree but not under the categorical one, which rendered difficult the comparison of both type of pedigrees.

Overall, the proportion of large offspring produced by the various males and females appeared fairly similar, at least superficially, under the quasi-categorical and

fractional pedigrees. Most differences were observed for sires or dams with a small number of offspring. Nevertheless, it would appear that the number of offspring evaluated here was probably too low to have much confidence in the ranking of parental broodstock. The only exception was found in spawning group 5 which used a very small pool of parents (6 females and 4 males) of which two produced a large number of offspring (Figure 21, bottom graphs). Female (7F7F411663) and male (7F7E640563) not only had the highest reproductive success (Figure 9, bottom graphs) but were also the top producers of large offspring. These two fish would have been top candidates for re-spawning.

The screening technique of sample juveniles for size to identify the parents producing the larger offspring appeared to have some value in the case of the AD PILOT group (874 fish). The total number of offspring allocated to each of the females and males was large enough, the proportion of large offspring per parent was very similar under the quasi-categorical and the fractional pedigrees and these proportions were significantly different among the various parents. Identifying superior parents producing larger offspring and re-using them in subsequent years would have been possible. This exercise would have had some value as well in the case of Group 5 Y5 (46 fish). In contrast, Group 1 Y5 (144 fish) and Group 3 Y5 (146 fish) were of a more limited value, mostly because the sample sizes were too low given the number of parents to assess the proportion of large offspring per parent with confidence. In one of the few studies looking at pedigree uncertainty and breeding value estimation, Famula (1993) showed with simulated data that including progeny with uncertain paternity in addition to the ones with known paternity could lead to dramatic improvement in breeding value accuracy. Although this situation is not exactly similar to the one here, since none of the fish here have certain parentage, it is nonetheless pointing out that the use of the fractional pedigree could have an important advantage in terms of improving the accuracy of estimated breeding values, if only through the intuitive mechanism that more progenies would be allocated to any tested parent when using a fractional pedigree than when using a categorical or quasi-categorical one. Finally, the procedure of sampling in the tails of the size distribution had been used in the hope that this would reduce sample size

necessary to rank parental values, but at the potential cost of not being able to estimate classical breeding values. However, through statistical analyses and simulations, Li *et al*. (2003) showed that appropriate estimation of breeding values could be conducted on data resulting from such tail-subsampling procedure.

Figure 22: Proportion of large offspring attributable to each female (left graphs) and male (right graphs) from spawning groups 1 (top graphs) and 5 (bottom graphs) of the “CRP2Y5” crosses. Pedigree legend: blue-QCEP, red-QCLODP, green-FP.

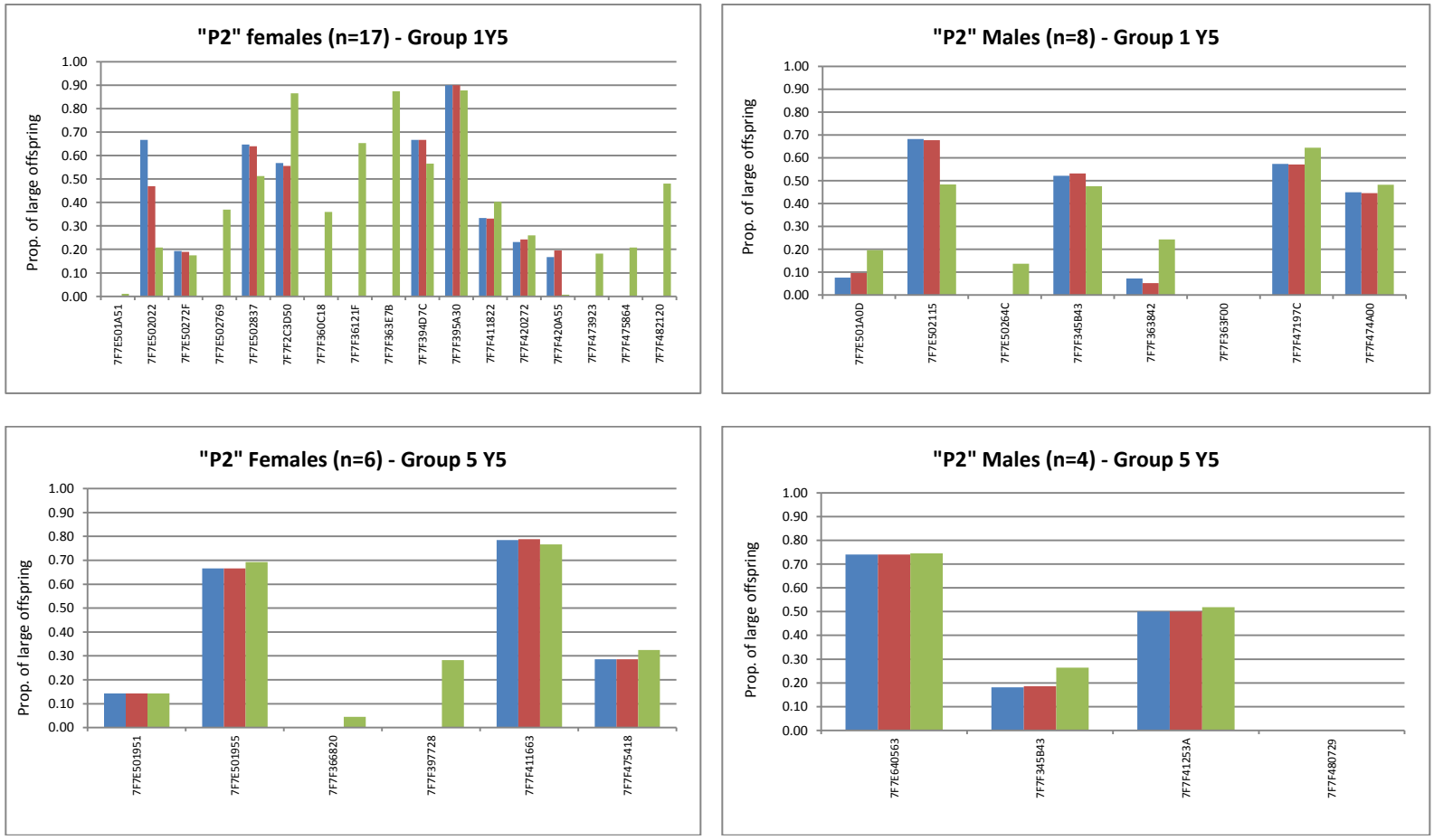
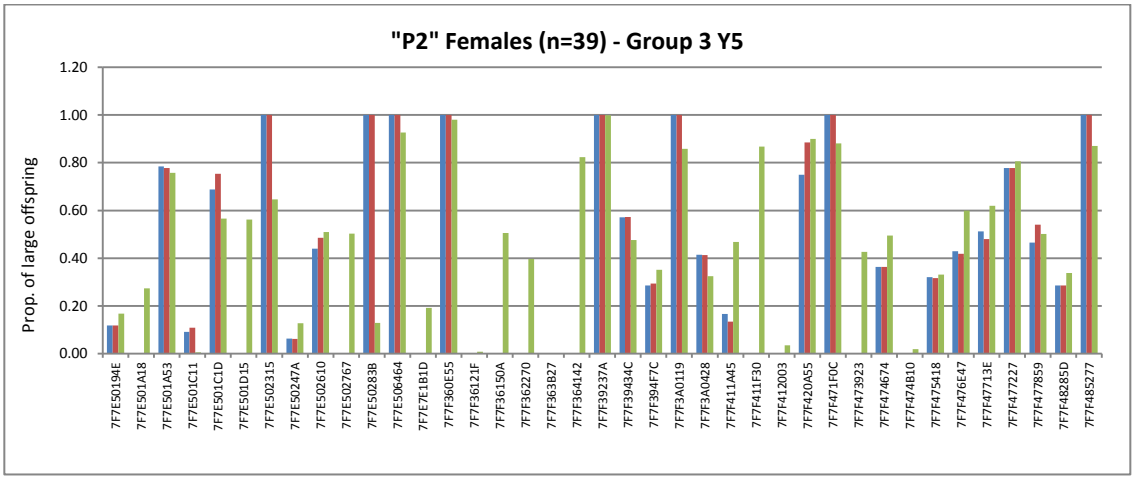
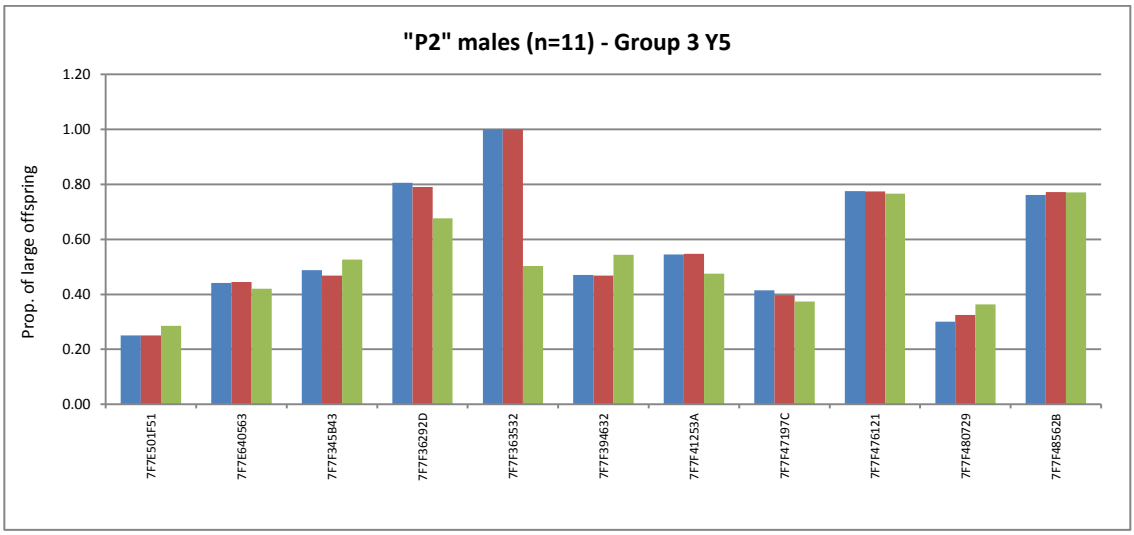


Figure 23: Proportion of large offspring attributable to each female (top graph) and male (bottom graph) from spawning group 3 of the “CRP2Y5” crosses. Pedigree legend: blue-QCEP, red-QCLODP, green-FP.





4.6 Superchill Event

In the winter of 1993-94, some rainbow trout originating from the 1991-92 spawning season (Year 4) were overwintering in one of SPA cages in the Bras d'Or lake (Herbinger, 1994). There was no direct information about the crossings that had been performed in year 4 and produced these individuals. Indeed many broodstock fish used that year were not tagged as the fish produced that year were destined for sale and were not part of the breeding program. However, hatchery records indicated indirectly that these fish originated from very few parents and might have been produced in 8-10 crosses. The majority of the overwintering fish died during a severe superchill episode but some fish survived. Samples were obtained from 95 surviving fish and 100 mortality fish. DNA microsatellite genotypes of all fish were generated at four microsatellite loci: Omy 2, Omy 38, Omy 77 and Omy 105. Subsequently, it was determined that 24 of the mortality fish were actually salmon that had been overwintering in a nearby cage and also suffered massive mortality. These fish were removed from the data set leaving 76 mortality fish. It was thought at the time that such an accidental mortality episode could be perhaps used in the sense that there could be genetic variation for resistance to extreme cold conditions and that using the surviving fish as broodstock might lead to increased robustness and cold resistance. Some of the surviving fish were therefore incorporated into the mating design for the 94/95 crosses (CRP2Y7) hoping to achieve this. However, the hypothesis that superchill survival might have a genetic basis was never properly tested.

The objectives of this side project are:

1. To evaluate the level of genetic diversity and the allelic frequency distribution in the survivors and the mortalities to assess potential genetic distinctness.
2. To assess whether both the superchill survivors and mortalities could be partitioned into distinct family groupings.
3. If #2 was successful, to see if the different families had different rates of survival in the superchill event.

4.6.1 Genetic Diversity

Genetic diversity between the two groups was determined using CERVUS 3.0 (Marshall *et al*., 1998) and includes: the number of alleles, size in base pairs, allele frequency, observed (H_0) and expected (H_e) heterozygosity and null allele frequency. At all loci, number of allele, and observed and expected heterozygosities were similar in both groups and there was no evidence of a null allele in either population based on the limited amount of genetic information available (Table 10). Allelic frequency distributions in the two populations were very similar with one exception: allele 134 at Omy 2 was very frequent in the survivor population but had very low occurrence in the mortalities (Figure 24).

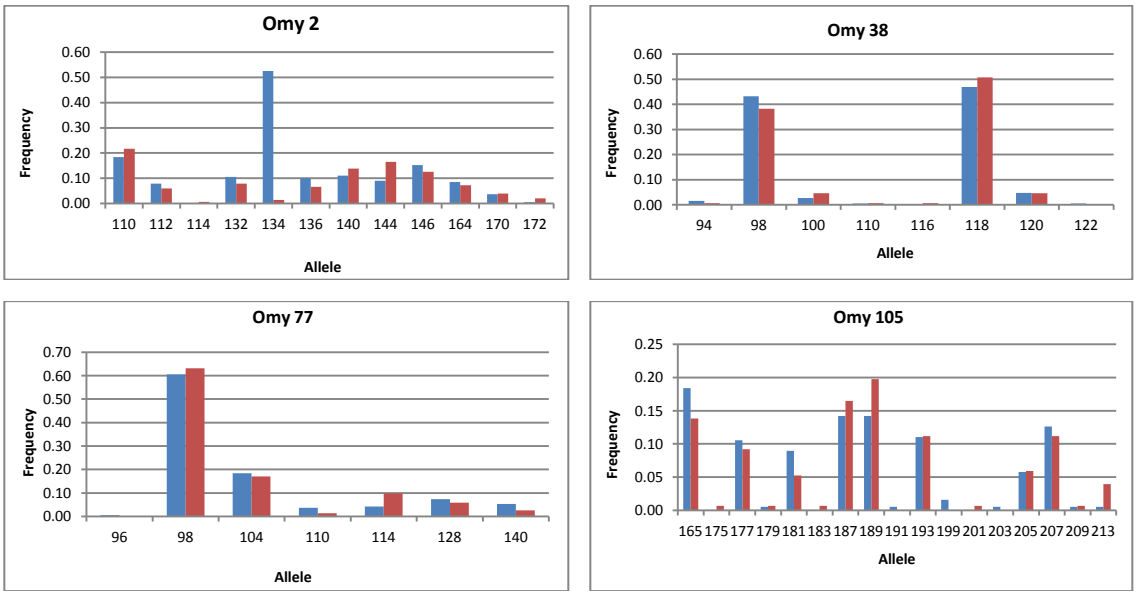
Table 10: Heterozygosities and number of alleles in the survivor and mortality groups.

	Omy 2		Omy 38		Omy 77		Omy 105	
	Survivors	Mortalities	Survivors	Mortalities	Survivors	Mortalities	Survivors	Mortalities
Number of alleles	11	12	7	7	7	6	14	14
H_O	0.9263	0.9211	0.5648	0.5789	0.5158	0.6053	0.9158	0.9079
H_E	0.8887	0.8754	0.5942	0.5973	0.5915	0.5614	0.8793	0.8791
Null allele frequency	-0.0248	-0.0289	0.0229	0.0170	0.0646	-0.05	-0.0221	-0.022

H_O = Observed Heterozygosity

H_E = Expected Heterozygosity

Figure 24: Allele frequency distribution at four loci (Omy 2, Omy 38, Omy 77 and Omy 105) for the survivors (blue) (n=95) and mortalities (red) (n=76).

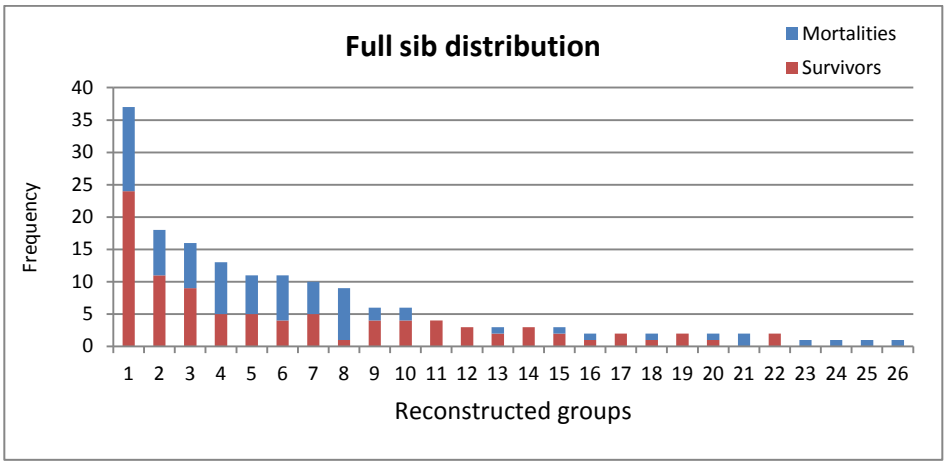


4.6.2 Sib-ship Inference.

Reconstruction of sibships was performed using two available software programs, PEDIGREE Ver. 2.2 (Herbinger, 2006) and COLONY Ver. 2.0 (Wang and Santure, 2009) to determine if the two groups could be partitioned into distinct full-sib and/or half-sib family groups based on available genetic information. Full-sib family structuring was observed with the 171 survivor/mortality fish using PEDIGREE with large differences observed in family size ranging from 1 to 37 fish per family (Figure 24). Significance testing was performed on the full-sib partition with 200 randomization trials as explained in Herbinger (2006). The overall partition was significant ($p < 0.005$) and the 4 largest reconstructed full-sib families, ranging in size from 13 to 37 individuals were significant as well and comprised 84 fish (49%). The remaining smaller groups could have been observed by chance and could be artifacts. Such a result is in reasonable agreement with the expected number of families given hatchery records. The proportion of survivors/mortalities in either the four or the eight largest PEDIGREE families (Figure

25) were not significantly different ($\chi^2 = 2.847$, df = 3, NS; $\chi^2 = 11.390$, df = 7, NS, respectively).

Figure 25: Partitioning of survivor and mortality fish (n=171) into full-sib families from PEDIGREE.



The survivor/mortality data set was also evaluated using COLONY Version 2.0 to try and infer sibship allowing for polygamy of both sexes. The 171 survivor/mortality fish clustered into full and half-sib family groups in size from 1 to 15 offspring per family (Figure 26). Fourteen hypothetical "Parents 1" and sixteen hypothetical "Parents 2" were reconstructed based on the available genetic data. The "Parent" designations are independent of the sex. Both COLONY and PEDIGREE were in good general agreement and grouped the same individuals together, although COLONY tended to generate a finer structure with smaller family size. The largest credible full-sib family (n=37) observed with PEDIGREE was partitioned into four half-sib families (5x4, 5x1, 7x4 and 7x1) highlighted in turquoise in Figure 28. The second largest full-sib family (n=18) also partitioned into four half-sib families (8x2, 8x7, 8x15 and 3x2) highlighted in purple. The third largest full-sib family (n=16) partitioned into two half-sib families (4x3 and 4x12) highlighted in green as did the fourth largest family (n=13) highlighted in red (6x5 and

6x2). The remaining widely dispersed very small groups of only 1 or 2 fish are likely artifacts.

Figure 26: Distribution of survivors (yellow rows) and mortalities (grey rows) to reconstructed sibships using COLONY where P1= reconstructed Parent 1, P2=reconstructed Parent 2 (e.g. 1 mortality fish and 2 surviving fish were assigned to reconstructed P1- 6 x reconstructed P2-2 highlighted in red).

P1	P 2																#
	4	2	1	7	5	11	3	12	15	13	10	14	6	8	9	16	
6		1			7	3	1	1		1	3	0	1				18
		2			3	0	0	1		0	2	3	0				11
4							4	2	1	1		1					9
							5	5	3	4		2					19
12						2	0	0									2
						3	1	1									5
7	2		3							0				1			6
	6		5							1				1			13
5	5	0	3					0	0				1				9
	10	1	3					1	1				0				16
8		3		0					2						1	0	6
		5		4					0						0	1	10
2	1	2		3		1			1						1	0	9
	0	2		1		0			0						0	1	4
1		2	2	2							1		1		1		9
		0	0	1							0		0		0		1
3		1		0							1			0	0		2
		1		1							0			1	1		4
13						0			0				1	0			1
						1			1				1	2			5
9		0	1							1						0	2
		1	0							0						1	2
10					1		1					0					2
					0		0					1					1
11					1	0				0					0		1
					0	1				1					1		3
14			0														0
			1														1
#	8	9	9	5	9	6	6	3	4	3	5	1	4	1	3	0	76
	16	12	9	7	3	5	6	8	5	6	2	6	1	4	2	3	95

4.6.3 Conclusion

The results of this side experiment indicate that the survivor/mortality fish that were sampled after the superchill event in the winter of 93/94 could be partitioned into distinct family groupings. The four largest reconstructed full-sib families were significant

and comprised 49% of the population and two half-sib family clusters. Survivor and mortality fish were found in each of the major reconstructed full-sib and half-sib families. However, the different families did not appear to have different rates of survival in the colder water.

Using microsatellite DNA markers and quantitative trait locus (QTL) mapping, studies by Cnaani *et al*., (2003) have shown that there is a genetic link between cold tolerance and fish size in tilapia hybrids. Unfortunately, there is no evidence that the surviving fish from the superchill event at the SPA hatchery were genetically predisposed to be more resistant to extreme cold conditions than those fish that died.

CHAPTER 5: CONCLUSIONS

5.1 Summary of Results

Genotypic and phenotypic data from three successive generations of rainbow trout from the SPA hatchery in Cape Breton, Nova Scotia was provided for this retrospective assessment of whether the SPA selective breeding program was successful in limiting inbreeding accumulation and potentially identifying superior broodstock. Spawning recommendations based on a simple scoring scheme related to origin of the broodstock and the number of shared alleles was provided to the hatchery to minimize the probability of crossing possibly related fish from within the same group. After one generation a semi-selective pressure was applied and the parents producing the best (largest) progeny were re-spawned. These genotypic and phenotypic data were never analyzed to assess reproductive success and change in inbreeding over the generations and to evaluate whether the spawning scheme and identification of putatively superior parents were successful in creating a non-inbred line of improved fish.

There was no apparent loss of genetic diversity from generations 1 to 3 with this population of rainbow trout as evaluated by the number of alleles or heterozygosities, despite an apparently overall reduced number of contributing parents. Differences in reproductive success were observed for the AD, RP, NC and Parents 3Y8 broodstock from the SPA hatchery and were typical of what actually happens in a real hatchery: many offspring were attributed to a limited number of mothers and a small number of fathers. The hatchery started out using very few “P1” males compared to the number of “P1” females. On the advice of the geneticists working with the SPA hatchery, with each subsequent spawning, more males and females were added to the pool and incorporated as broodstock so as not to limit the level of genetic diversity in the broodstock pool. By generation 3, a much larger pool of females were being used and quite a few more males were being used (although they were not all successful in producing offspring). From that aspect, the biologists were largely successful in encouraging the hatchery to keep as large a broodstock pool as possible to reduce the chance of breeding within family groups.

Differences in reproductive success were observed among sires and dams from all groups as estimated based on the three different pedigrees (QCEP, QCLODP and FP). The two quasi-categorical pedigrees gave very similar estimates of reproductive success since the majority of offspring were assigned to the same sets of trios. Reproductive success under the fractional pedigree was often in good agreement with the categorical one for the majority of the females and males, but there were a number of individuals where estimates of reproductive success under the different types of pedigrees were substantially different. A number of individuals did not allocate to any parents under the two quasi-categorical pedigrees which seemed to be directly related to false exclusions due to the quality of the genotypic data (i.e. scoring error). This problem was largely resolved under the fractional pedigree with most offspring assigned to at least a parental pair. The fractional approach appeared to generate a more accurate estimate of reproductive success because it incorporated all levels of genetic relatedness and was not affected to the biases observed with the quasi-categorical approach.

Inbreeding values for each fish were estimated from the average numerator relationship matrix ($\bar{A}$) derived for each of the three different pedigrees for all individuals in generation 2 and 3. Little or no inbreeding was observed for the majority of the fish using either the quasi-categorical and fractional approach (~0.3% and 0.7% inbreeding observed in generation 3 respectively). With the addition of a layer of reconstructed "P0" parents, inbreeding was observed in generation 2 as expected, and the average level of inbreeding increased from generation 2 to 3 for both types of pedigrees (quasi-categorical and fractional). Even so, the levels observed were still very low. An average inbreeding level of ~ 0.8% was observed under the quasi-categorical pedigrees by generation 3 and ~ 1.7% under the fractional pedigree. The lower estimate of inbreeding observed under the quasi-categorical pedigrees was most probably an underestimate of inbreeding attributable to the disconnect (holes) observed between generation 1 and 2, a problem that was largely resolved under the fractional pedigree. The overall low level of inbreeding observed in generation 3 (by either method of pedigree reconstruction), after two generations of semi-selective breeding, also seems to indicate that the simple breeding

scheme devised for the SPA hatchery was successful in limiting inbreeding increase over three generations.

The spawning recommendations provided to the hatchery at the onset of the 92/93 spawning season were reasonably efficient in minimizing the accumulation of inbreeding. The low levels if inbreeding observed from the group of crosses that were give an "OK" rating from the realized crosses performed in 92/93 show that the simple screening rule used at the hatchery to avoid crossing fish from the same offspring group and assessing the number of common alleles in a putative pair was relatively effective. However it was apparent that this screening rule could still assess poorly some putative parental pairs. High levels of co-ancestry were observed in several pairs that were rated as "OK", mostly as a consequence of the use of the same sires across different spawning groups. Reconstructing the pedigree, particularly with a fractional approach, and estimating the co-ancestry among pairs of putative broodfish as was done in this thesis would have certainly been more efficient and accurate than the simple pair-wise estimation that was done at the time. Based on empirical estimates of inbreeding depression in salmonids (Wang *et al*., 2002), inbreeding depression associated with an overall inbreeding level of 1.7% would have been negligible for the great majority of the 3rd generation fish. However offspring of the few crosses that were produced with much higher co-ancestry levels (e.g. 7% and higher, Figure 20) could be expected to show substantial depression for many traits of commercial interest. The screening technique of sampling juveniles for size to identify parents producing larger offspring appeared to have some value when the offspring sample size was large as was the case with the AD PILOT group; however when the sample size was small, the evaluation was less precise and that screening technique would have been of less value. Lastly, there did not seem to be a genetic basis as to why some fish survived and some fish did not survive a superchill event that occurred at the SPA hatchery in the winter to 93/93.

Despite a very limited amount of genetic information available at the time (data at only 4 loci or less), it was possible to answer questions about the status of the 3 generations of rainbow trout from the SPA hatchery such as the level of genetic variability in the population, the number and identification of the most successful

breeders in each generation and the individual and average levels of inbreeding for the population. The fractional parentage approach that was used here for the first time on a real data set seemed to be very useful and avoided several problems that were seen with more classical categorical approaches. Although this selective breeding program did not continue past the third generation, it would appear that many of the basic ideas that were tried in that pioneering program worked to some degree. In essence, the simple spawning and selection scheme used at the hatchery was relatively successful in creating a non-inbred line of potentially mildly improved fish.

5.2 Implications for Aquaculture Operations

Inbreeding control is generally regarded as a highly desirable goal for fish and shellfish farmers and breeders. However, recommendations regarding methods of limitation of inbreeding often concentrate on the population level estimation and management of the effective population size N_e (e.g. Tave, 1999). In the simplest case, this parameter is linked to the number of males and number of females. The average inbreeding level can then be estimated as half the inverse of the effective population size. The recommended policy is then to keep this number within reasonable bounds in hatcheries. This is a simple approach but it can present difficulties. The formula suggested above for population size will just give a gross estimation of an order of magnitude. Real populations deviate indeed from theoretical population models which use assumptions such as random mating and equal reproductive success. Unequal sex ratio, variance in reproductive success among males and females and non random mating, potentially lead to effective population sizes being an order of magnitude smaller than the more easily calculated breeder census size. Yet the real extent of unequal sex ratio and unequal family size can only be determined accurately by looking at the family distribution of the next generation of breeders. The present study showed for example that there was a considerable variance in reproductive success with many males and females actually not producing any offspring in the next generation. The average inbreeding level

is more precisely estimated by averaging individual inbreeding levels than can be derived from pedigree reconstruction as was done here and furthermore individual inbreeding levels are also potentially of interest. A probabilistic or fractional approach to pedigree reconstruction was used here for the first time on a real data set to integrate over the pedigree uncertainty after fully taking into account the genetic (DNA marker) and mating design information, and it appeared to allow a reasonably robust pedigree reconstruction with a modest number of markers. The fact that genetic information had been limited to four loci, and furthermore that this information had not been exploited to the fullest extent possible at the time (i.e. not reconstructing the pedigree as was done in this thesis) simply reflected technical limitations in the time and expenses associated with microsatellite markers two decades ago. These limitations are considerably less stringent now and genotyping a few hundred putative broodstock at a much larger number of loci allowing finer pedigree resolution would be feasible in a short time and for a reasonable price tag. The basic ideas that were proposed in that early DNA-marker based breeding program could therefore be built upon and improved with the approaches that were used here. In particular, generating in real time the DNA marker genotypes, assembling the fractional pedigree and calculating the average relationship matrix, would allow an instantaneous evaluation of the individual and average inbreeding levels, would allow estimating the co-ancestry of putative pairs to avoid mating related individuals, and in conjunction with phenotypic data would even allow estimating breeding values following the animal model approach.

REFERENCES

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C. and Taberlet, P. (2004). Invited Review: How to track and assess gentoyping erroros in population genetics studies. *Molecular Ecology , 13*, 3261-3273.

Borrell, Y. J., Alvarez, J., Vasquex, E., Pato, C. F., Tapia, C. M., Sanchez, J. A. and Blando, G. (2004). Applying microsatellites to the managment of farmed turbot stocks (*Scophthalmus maximus* L.) in hatcheries. *Aquaculture , 241*, 133-150.

Brown, A. H., Barrett, S. C. and Morgan, G. F. (1985). Mating system estimation in forest trees: models, methods and meanings. In H. R. Gregorius (Ed.), *Population Genetics in Forestry.* (pp. 32-49). Berlin Heidelbert New York: Springer.

Castro, J., Bouza, C., Presa, P., Pino-Querido, A., Riaza, A., Ferreiro, I., Sanchez, L. and Martinez, P. (2004). Potential sources of error in parentage assessment of turbot (*Scophthalmus maximus*) using microsatellite loci. *Aquaculture , 242*, 119-135.

Cnaani, A., Hallerman, E. M., Ron, M., Weller, J., Indelman, M., Kashi, Y., Gall, G. A. and Hulala, G. (2003). Detection of a chromosomal region wtih two quantitative trait loci, affecting cold tolerance and fish size, in an F2 tilapia hybrid. *Aquaculture , 223* (1-4), 117-128.

Delghandi, M., Mortensen, A. and Westgaard, J.-I. (2003). Simultaneous Analysis of Six Microsatellite Markers in Atlantic Cod (*Gadus Morhua*): A Novel Multiplex Assay System for Use in Selective Breeding Studies. *Marine Biotechnology , 5*, 141-148.

Devlin, B., Roeder, K. and Ellstrand, N. C. (1988). Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical Applied Genetics , 76*, 369-380.

Dodds, K. G., Tate, M. L. and Sise, J. A. (2005). Genetic evaluation using parentage information from genetic markers. *Journal of Animal Science , 83*, 2271-2279.

Doyle, R. W. and Herbinger, C. M. (1994). The use of DNA fingerprinting for high-intensity within-family selection for fish breeding., *19*, pp. 364-371. Guelph, Ontario.

Doyle, R. W., Herbinger, C. M., Taggart, C. T. and Lochmann, S. (1995). Use of DNA microsatellite polymorphism to analyze genetic correlations between hatchery and natural fitness. *American Fisheries Society Symposium , 15*, 205-211.

Dube, P. and Mason, E. (1995). Trout Culture in Atlantic Canada. In A. D. Baghan (Ed.), *Cold-Water Aquaculture in Atlantic Canada.* The Canadian Institute for Research on Regional Developement.

Easton, A. A., Moghadam, H. K., Danzmann, R. G. and Ferguson, M. M. (2011). The genetic architecture of embryonic developmental rate and genetic covariation with age at maturation in rainbow trout *Oncorhynchus mykiss*. *Journal of Fish Biology , 78*, 602-623.

Ellestrand, N. C. (1984). Multiple paternity within the fruits of the wild radish, Raphanus sativus. *American Nature , 123*, 819-828.

Ellstrand, N., Devlin, B. and Roeder, K. (1988). Factional paternity assignment: theoretical development and comparison to other methods. *Theoretical Applied Genetics , 76*, 369-380.

Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P. and Guyomard, R. (1998). Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. *Canadian Journal of Fisheries and Aquatic Science , 55*, 715-725.

Evans, F., Matson, S., Brake, J. and Langdon, C. (2004). The effects of inbreeding on performance traits of adult Pacific oysters (*Crassostrea gigas*). *Aquaculture , 230* (1-4), 89-98.

Falconer, D. S. and T.F.C Mackay. (1996). *Introduction ot Quantitative Genetics* (4 ed.). Harlow, UK: Longman Scientific and Technical.

Famula, T. R. (1993). The contribution of progeny of uncertain paternity to the accuracy of sire evaluation. *Journal of Animal Science , 71*, 1136-1141.

Ferguson, M. M. and Danzmann, R. G. (1998). Role of genetic markers in fisheries and aquaculture: useful tools or stamp collecting? *Canadian Journal of Fisheries and Aquatic Science , 55*, 1553-1563.

Fishback, A. G., Danzmann, R. G. and Ferguson, M. M. (2000). Microsatellite allelic heterogeneity among hatchery rainbow trout in different seasons. *Journal of Fish Biology , 57*, 1367-1380.

Fishback, A. G., Danzmann, R. G., Ferguson, M. M. and Gibson, J. P. (2002). Estimates of genetic parameters and genotype by environment interactions for growth traits of the rainbow trout (*Oncorhynchus mykiss*) as inferred using molecular pedigrees. *Aquaculture , 206*, 137-150.

Fishback, A. G., Danzmann, R. G., Sakamoto, T. and Ferguson, M. M. (1999). Optimization of semi-automated microsatellite multiplex polymerase chain reaction systems for rainbow trout (*Oncorhynchus mykiss*). *Aquaculture , 172*, 247-254.

Ford, M. J. and Williamson, K. S. (2010). The Aunt and Uncle Effect Revisited - The Effect of Biased Parentage Assignment on Fitness Estimation in a supplemented Salmon Population. *Journal of Heredity , 101(1)*, 33-41.

Frost, L. A., Evans, B. S. and Jerry, D. R. (2006). Loss of genetic diversity due to hatchery culture practices in barramundi (*Lates calcarifer*). *Aquaculture , 261* (3), 1056-1064.

Garcia de Leon, F. J., Dallas, J. F., Chatain, B., Cannonne, M., Versini, J. J. and Bonhomme, F. (1995). Development and use of microsatellite markers in sea bass (*Dicentrarchus labrax*). *Molecular Biology and Biotechnology , 4*, 62-68.

Gjedrem, T. (2000). Genetic improvement of cold-water fish species. *Aquaculture Research , 31* (1), 25-33.

Hamerick, J. L. and Schnabel, A. (1985). Understanding the genetic structure of plant populations: some old problems and a new approach. In H. R. Gregorius (Ed.), *Population Genetics in Forestry* (pp. 50-70). Berlin Heidelberg New York: Springer.

Henderson, C. R. (1988). Use of an average numberator relationship matrix for multiple-sire joining. *Journal of Animal Science , 66*, 1614-1621.

Herbinger, C. M. (1993b). *Development of a farm-based genetic improvement program for the SPA hatchery using DNA fingerpring technology.*

Herbinger, C. M. (1997b). *Final Report. An overview of the project "Development of a farm-based genetic improvement program for the SPA hatchery using DNA fingerprinting technology" (Oct 93 - Dec 96).*

Herbinger, C. M. (1993a). *High-yield rainbow trout for aquaculture in eastern Canada using recombinant DNA.*

Herbinger, C. M. (2006). *Pedigree 2.2*. Retrieved from Dalhousie Univeristy Welcome to Pedigree: http:/herbinger.biology.dal.ca:5080/Pedigree/

Herbinger, C. M. (1994). *Report of the MGPL activities in Year 1 (Oct 93 - Oct 94) of the project "Development of a farm-based genetic improvement program for the SPA hatchery using DNA fingerprinting technology".*

Herbinger, C. M. (1995). *Report of the MGPL activities in Year 2 (Oct 94 - Oct 95) of the project "Development of a farm-based genetic improvement program for the SPA hatchery using DNA fingerprinting technology.".*

Herbinger, C. M. (1997a). *Report of the MGPL activities in Year 3 (Oct 95 - dec 96) of the project "Development of a farm-based genetic improvement program for the SPA hatchery using DNA fingerprinting technology".*

Herbinger, C. M., Doyle, R. W., Pitman, E. R., Paquet, D., Mesa, K., Morris, D. B. and Wright, J. M. (1995). DNA fingerpring based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture* , 245-256.

Herbinger, C. M., O'Reilly, P. T., Doyle, R. W., Wright, J. M., and O'Flynn, F. (1999). Early growth performance of Atlantic salmon full-sib families reared in single family tanks versus in mixed family tanks. *Aquaculture , 173*, 105-116.

Herbinger, C. M., Reith, M. E. and Jackson, T. R. (2003). DNA markers and Aquaculture Genetics. In *Recent Advances in Marine Biotechnology.* Enfield: Science Publisher, Inc.

Jones, A. G. and Arden, W. R. (2003). Invited Review: Methods of parentage analysis in natureal populations. *Molecular Ecology , 12*, 2511-2523.

Jones, B., Grossman, G. D., Walsh, D. C., Porter, B. A., Avise, J. C. and Fiumera, A. C. (2007). Estimating Differential Reproductive Succcess From Nests of Related Individiuals, With Application to a Study of the Mottled Sculpin, *Cottus bairdi*. *Genetics , 176*, 2427-2439.

Kalinowski, S. T., Taper, M. L. and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology , 16*, 1099-1006.

Lemay, M. A. and Boulding, E. G. (2009). Microsatellite pedigree analysis reveals high variance of reproductive success and reduced genetic diversity in hatchery-spawned northern abalone. *Aquaculture , 295*, 22-29.

Li, K., Field, C. A. and Doyle, R. W. (2003). Estimation of additive genetic variance components in aquaculture populations selectively pedigreed by DNA fingerprinting. *Biometrical Journal , 45*, 61-72.

Liu, Y., Chen, S. and Li, B. (2005). Assessing the genetic structure of three Japanese flounder (*Paraclichthys olivaceius*) stocks by microsatellite markers. *Aquaculture , 243*, 103-111.

Liu, Z. J. and Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture , 238*, 1-37.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits.* Sunderland, Massechusetts: Sinauer Associates Inc.

Marshall, C. T., Slate, J., Kruuk, L. B. and Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology , 7*, 639-655.

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

McKay, L. R., McMillan, I., Sadler, S. E. and Moccia, R. D. (1992). Effects of mating system on inbreeding levels and selection response in Salmonid aquaculture. *Aquaculture , 100*, 100-101.

McLean, J. E., Bentzen, P. and Quinn, T. P. (2005). Nonrandom, Size- and Timing-Biased Breeding in a Hatchery Population of Steelhead Trout. *Conservation Biology , 19* (2), 446-454.

Meagher, T. R. and Thompson, E. A. (1987). Analysis of parentage for naturally established seedlings within a population of *Chamaelirium luteum* (Liliaceae). *Ecology , 68*, 803-812.

Minitab 16 Statistical Software. (2010). *[Computer software] State College, PA: Minitab, Inc. (www.minitab.com)* .

Morris, D. B. (1993). The isolation and characterization of microsatellites from rainbow trout (*Oncorhynkiss mykiss*). *M.Sc. Thesis, Dalhousie University, Halifax* .

Morris, D. B. and Richard, K. R. (1995). Polymorphic microsatellites from rainbow trout (*Oncorhynchus mykiss*) are conserved in salmonids. *Canadian Journal of Fisheries and Aquatic Science* .

Morris, D. B., Richard, K. R. and Wright, J. M. (1996). Microsatellites from rainbow trout (*Oncorhynchus mykiss*) and their use for the genetic study of salmonids. *Canadian Journal of Fisheries and Aquatic Science , 53*, 120-126.

Neilson, J. L. (1998). Population genetics and the conservation and management of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Science , 55* (Suppl. 1), 145-152.

Norris, A. T., Bradley, D. G. and Cunningham, E. P. (2000). Parentage and relatedness determination in farmed Atlantic Salmon (*Salmo salar*) using microsatellite markers. *Aquaculture , 182*, 73-83.

O'Connell, M. and Wright, J. M. (1997). Microsatellite DNA in fishes. *Review of Fish Biology and Fisheries , 7*, 1-33.

O'Reilly, P. and Wright, J. M. (1995). The evolving technology of DNA fingerprinting and its application to fisheries and aquaculture. *Journal of Fish Biology , 47* (Suppl. A), 29-55.

O'Reilly, P., Herbinger, C. and Wright, J. M. (1998). Analysis of parentage determination in Atlantic salmon (*Salmo salar*) using microsatellites. *Animal Genetics , 29*, 363-370.

Pante, M. G., Gjerde, P. and McMillan, B. (2001). Inbreeding levels in selected populations of rainbow trout, *Oncorhynchus mykiss*. *Aquaculture , 192*, 213-244.

Perez-Enciso, M. and Fernando, R. L. (1992). Genetic evaluation with uncertain parentage: a comparison of methods. *Theoretical Applied Genetics , 84*, 173-179.

Perez-Enriquez, R., Takagi, M. and Taniguchi, N. (1999). Genetic variability and pedigree tracing of a hatchery-reared stock of read sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. *Aquaculture , 173*, 413-423.

Renshaw, M. A., Saillant, E., Bradfield, S. C. and Gold, J. R. (2006). Microsatellite multiplex panels for genetic studies of three species of marine fishes: red drum (*Sciaenops ocellatus*), red snapper (*Lutjanus campechanus*) and cobia (*Rachycentron canadum*). *Aquaculture , 253* (1-4), 731-735.

Sakamoto, T., Danzmann, R. G., Okamoto, N., Ferguson, M. M. and Ihssen, P. E. (1999). Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture , 173*, 33-43.

Sekino, M., Saitoh, K., Yamada, T., Kumagai, A. and Hara, M. (2003). Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: implications for hatchery management related to stock enhancement program. *Aquaculture , 221* (1-4), 255-263.

Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E. and McEwan, J. C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity , 93*, 255-265.

Smith, B. R., Herbinger, C. M. and Merry, H. R. (2001). Accurate partition of individuals into full sib families from genetic data without parental information. *Genetics , 158*, 1329-1338.

Smouse, P. E. and Chakraborty, R. (1986). The use of restriction fragment length polymorphism in paternity analysis. *American Journal of Human Genetics , 38*, 918-939.

Smouse, P. E. and Meagher, T. R. (1994). Genetic Analysis of Male Reproductive Contributions in *Chamaelirium luterum* (L.) Gray (Liliaceae). *Genetics , 136*, 313-322.

Taris, N., Baron, S., Sharbel, T., Sauvage, C. and Boudry, P. (2005). A combined microsatellite multiplexing and boiling DNA extraction method for high-throughput parentage analyses in the Pacifid oyster (*Crassostrea gigas*). *Aquaculture Research , 36(5)*, 516-518.

Tave, D. (1999). Inbreeding and broodstock management. *Fisheries Technical Paper , No. 392*, 122p.

Tessier, N., Bernatches, L., Persa, P. and Angers, B. (1995). Gene diversity analysis of mitochondrial DNA, microsatellites and allozymes in landlocked Atlantic salmon. *Journal of Fish Biology , 158*, 156-163.

Trippel, E. A., Rideout, R. M., O'Reilly, P. T., Herbinger, C. M., Neil, S. R. and Hamilton, L. (2009). Communal spawniing leads to high potential for inbreeding in gadoid aquaculture. *Aquaculture , 296*, 27-35.

Vandeputte, M., Kocour, M., Mauger, S., Dupont-Nivet, M., De Guerry, D., Rodina, M., Gela, D., Vallod, D, Chevassus, B. and Linhart, O. (2004). Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (*Cyprinus carpio* L.). *Aquaculture , 235*, 223-236.

Wang, J. (2004). Sibship reconstruction from genetic data with typing errors. *Genetics , 166*, 1963-1979.

Wang, J. and Santure, A. W. (2009). Parentage and sibship inference from multilocus genotype data under polygomy. *Genetics , 181*, 1579-1594.

Wang, S., Hard, J. J. and Utter, F. (2002). Salmonid inbreeding: a review. *Reviews in Fish Biology and Fisheries , 11* (4), 301-319.

Was, A. and Wenne, R. (2002). Genetic differentiation in hatchery and wild sea trout (*Salmo trutta*) in the Southern Baltic at microsatellite loci. *Aquaculture , 204*, 493-506.

Weaver, R. F. and Hedrick, P. W. (1989). *Genetics.* Dubuque, IA: Wm. C. Brown Publishers.

Wilson, A. J., McDonald, G., Moghadam, H. K., Herbinger, C. M. and Ferguson, M. M. (2003). Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*. *Genetic Research , 81*, 145-156.

Wright, J. M. and Bentzen, P. (1994). Microsatellites: genetic markers for the future. *Review of Fish Biology and Fisheries , 4*, 384-388.

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Nature , 56*, 330-339.

Appendix 1a: Omy 2 Summary Table (yellow=alleles not observed in the "P1" founders).

		Generation 1	Generation 2			Generation 3					
Locus	Allele	Founders (n=224)	AD group (n=1098)	NC group (n=140)	RP group (n=132)	Group 1 Y5 (n=144)	Group 3 Y5 (n=141)	Group 5 Y5 (n=46)	Group 11 Y5 (n=40)	Parents 3 Y8 (n=309)	Parents 3 Y9 (n=122)
Omy 2	98	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0000
	108	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0065	0.0000
	110	0.1362	0.1915	0.0992	0.1314	0.0476	0.1187	0.1304	0.2917	0.1266	0.1598
	112	0.0714	0.0604	0.0744	0.0975	0.0952	0.0504	0.0652	0.0694	0.0568	0.0533
	114	0.0179	0.0189	0.0000	0.0127	0.0833	0.0000	0.0000	0.0417	0.0049	0.0000
	132	0.0201	0.0000	0.0000	0.0466	0.0238	0.0324	0.0217	0.0000	0.0519	0.0000
	134	0.0335	0.0538	0.0041	0.0339	0.0119	0.0000	0.0217	0.0000	0.0373	0.0082
	136	0.1384	0.1099	0.0496	0.1144	0.1190	0.1259	0.1848	0.0278	0.1494	0.0861
	138	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0097	0.1393
	140	0.1295	0.0976	0.1033	0.1314	0.0952	0.0791	0.0217	0.1111	0.0942	0.0082
	142	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0820
	144	0.1942	0.1741	0.0744	0.0932	0.1905	0.1475	0.0652	0.0000	0.1380	0.0123
	146	0.1027	0.0778	0.0992	0.1017	0.1071	0.1619	0.2609	0.0000	0.0828	0.0656
	147	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0109	0.0000	0.0000	0.0000
	148	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0656
	150	0.0022	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0000
	162	0.0536	0.0571	0.0207	0.0169	0.1548	0.0576	0.0000	0.0000	0.0373	0.0000
	164	0.0737	0.1580	0.3926	0.1949	0.0714	0.1978	0.2174	0.0000	0.1899	0.0410
	166	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2254
	170	0.0223	0.0005	0.0372	0.0169	0.0000	0.0252	0.0000	0.1111	0.0114	0.0000
	172	0.0045	0.0005	0.0455	0.0085	0.0000	0.0036	0.0000	0.3472	0.0000	0.0533
Number of individuals typed		224	1060	121	118	42	139	46	36	308	122
Number of alleles		14	12	11	13	11	11	10	7	16	13
Observed Heterozygosity		0.8973	0.8896	0.9504	0.8814	0.9524	0.8993	0.9565	0.6944	0.8994	0.8689
Expected Heterozygosity		0.8835	0.8706	0.8014	0.8855	0.8916	0.8723	0.8325	0.7731	0.8832	0.8775
Null Allele Frequency		-0.0107	-0.0109	-0.1044	0.0000	-0.0398	-0.0181	-0.0782	0.0405	-0.0116	0.0035

Appendix 1b: Omy 38 Summary Table (yellow=alleles not observed in the "P1" founders).

		Generation 1	Generation 2			Generation 3					
Locus	Allele	Founders (n=224)	AD group (n=1098)	NC group (n=140)	RP group (n=132)	Group 1 Y5 (n=144)	Group 3 Y5 (n=141)	Group 5 Y5 (n=46)	Group 11 Y5 (n=40)	Parents 3 Y8 (n=309)	Parents 3 Y9 (n=122)
Omy 38	92	0.0201	0.0000	0.0088	0.0048	0.0000	0.0000	0.0000	0.0000	0.0049	0.0000
	94	0.0625	0.0225	0.0442	0.0381	0.0321	0.0303	0.0000	0.0000	0.0147	0.0041
	98	0.2567	0.3502	0.3451	0.2714	0.2929	0.3333	0.2609	0.2125	0.3567	0.3934
	100	0.0112	0.0000	0.0000	0.0048	0.0286	0.0000	0.0000	0.0000	0.0033	0.0000
	104	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0000
	110	0.0379	0.0517	0.0398	0.0857	0.0571	0.0644	0.0652	0.1250	0.1026	0.0205
	112	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0114	0.0000
	118	0.5268	0.3914	0.4956	0.5000	0.4964	0.5152	0.5543	0.3875	0.4137	0.4877
	120	0.0580	0.1043	0.0664	0.0667	0.0929	0.0530	0.1196	0.2750	0.0733	0.0943
	130	0.0268	0.0799	0.0000	0.0286	0.0000	0.0000	0.0000	0.0000	0.0179	0.0000
	132	0.0000	0.0000	0.0000	0.0000	0.0000	0.0038	0.0000	0.0000	0.0000	0.0000
Number of individuals typed		224	1045	113	105	140	132	46	40	307	122
Number of alleles		8	6	6	8	6	6	4	4	10	5
Observed Heterozygosity		0.6295	0.7129	0.8407	0.6857	0.5786	0.6894	0.6739	0.8250	0.7036	0.5984
Expected Heterozygosity		0.6481	0.7040	0.6300	0.6654	0.6564	0.6180	0.6128	0.7225	0.6862	0.6005
Null Allele Frequency		0.0139	-0.0031	-0.1660	-0.0175	0.0652	-0.0631	-0.0431	-0.0698	-0.0118	0.0045

Appendix 1c: Omy 77 Summary Table (yellow=alleles not observed in the "P1" founders).

		Generation 1	Generation 2			Generation 3					
Locus	Allele	Founders (n=224)	AD group (n=1098)	NC group (n=140)	RP group (n=132)	Group 1 Y5 (n=144)	Group 3 Y5 (n=141)	Group 5 Y5 (n=46)	Group 11 Y5 (n=40)	Parents 3 Y8 (n=309)	Parents 3 Y9 (n=122)
	96	0.0561	0.0000	0.0146	0.0388	0.0035	0.0217	0.0000	0.0000	0.0340	0.0458
Omy 77	98	0.3812	0.4703	0.4562	0.4612	0.4507	0.3333	0.0761	0.8750	0.4159	0.3042
	99	0.0000	0.0000	0.0000	0.0000	0.0141	0.0000	0.0000	0.0000	0.0000	0.0000
	102	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0032	0.0000
	104	0.1211	0.1615	0.0803	0.0814	0.1268	0.1051	0.2174	0.1000	0.1230	0.1375
	108	0.0045	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0000
	110	0.0291	0.0206	0.0036	0.0310	0.0211	0.0000	0.0000	0.0000	0.0016	0.0000
	114	0.1143	0.0819	0.0328	0.1357	0.1056	0.0507	0.1304	0.0000	0.0906	0.0542
	115	0.0000	0.0000	0.0000	0.0000	0.0035	0.0000	0.0000	0.0000	0.0000	0.0000
	116	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0042
	120	0.0247	0.0059	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	124	0.0157	0.0220	0.0036	0.0310	0.0317	0.0435	0.0109	0.0000	0.0615	0.0000
	128	0.1525	0.0869	0.0365	0.0930	0.1479	0.1413	0.2935	0.0000	0.1036	0.1167
	130	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0146	0.0125
	140	0.1009	0.0000	0.0000	0.0000	0.0915	0.3043	0.2717	0.0250	0.1505	0.3250
	141	0.0000	0.1510	0.3723	0.1279	0.0035	0.0000	0.0000	0.0000	0.0000	0.0000
Number of individuals typed		223	1093	137	129	142	138	46	40	309	120
Number of alleles		10	8	8	8	11	7	6	3	11	8
Observed Heterozygosity		0.7265	0.7264	0.6569	0.7364	0.7183	0.8261	0.8696	0.2500	0.7411	0.7167
Expected Heterozygosity		0.7905	0.7151	0.6466	0.7366	0.7403	0.7631	0.7783	0.2266	0.7664	0.7673
Null Allele Frequency		0.0332	-0.0055	-0.0082	-0.0025	0.0102	-0.0480	-0.0616	-0.0603	0.0175	0.0329

Appendix 1d: Omy 105 Summary Table (yellow=alleles not observed in the "P1" founders).

		Generation 1	Generation 2			Generation 3					
Locus	Allele	Founders (n=244)	AD group (n=1098)	NC group (n=140)	RP group (n=132)	Group 1 Y5 (n=144)	Group 3 Y5 (n=141)	Group 5 Y5 (n=46)	Group 11 Y5 (n=40)	Parents 3 Y8 (n=309)	Parents 3 Y9 (n=122)
Omy 105	119	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0000
	161	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0049	0.0000
	165	0.1855	0.1879	0.1583	0.2193	0.2158	0.1898	0.2717	0.1282	0.1408	0.2607
	171	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0081	0.0000
	173	0.0023	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0000
	175	0.0317	0.0005	0.0083	0.0395	0.0368	0.0146	0.0000	0.0128	0.0113	0.0214
	177	0.1923	0.3187	0.3292	0.2281	0.2842	0.2482	0.0543	0.3205	0.2654	0.1154
	179	0.0407	0.0807	0.0500	0.0439	0.0158	0.0620	0.0217	0.0000	0.0518	0.0897
	181	0.0181	0.0000	0.0125	0.0175	0.0053	0.0474	0.1522	0.0000	0.0307	0.0000
	183	0.0136	0.0000	0.0083	0.0000	0.0000	0.0000	0.0000	0.0000	0.0049	0.0000
	187	0.0679	0.0401	0.0500	0.1404	0.1053	0.1095	0.0652	0.2051	0.1327	0.0855
	189	0.1109	0.1176	0.0958	0.0702	0.0842	0.0511	0.1957	0.0256	0.0793	0.0855
	191	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0043
	193	0.0543	0.0099	0.0500	0.0175	0.0263	0.0182	0.0000	0.2179	0.0065	0.0342
	195	0.0023	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	199	0.0701	0.1185	0.1958	0.0921	0.0632	0.1058	0.1848	0.0897	0.1100	0.1624
	201	0.0045	0.0000	0.0000	0.0088	0.0000	0.0036	0.0000	0.0000	0.0146	0.0043
	203	0.0158	0.0000	0.0125	0.0000	0.0000	0.0036	0.0000	0.0000	0.0146	0.0043
	205	0.1222	0.0765	0.0292	0.0965	0.0632	0.1168	0.0543	0.0000	0.0744	0.1282
	207	0.0158	0.0000	0.0000	0.0088	0.0368	0.0000	0.0000	0.0000	0.0065	0.0000
	209	0.0475	0.0378	0.0000	0.0132	0.0632	0.0255	0.0000	0.0000	0.0243	0.0043
	211	0.0000	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	213	0.0045	0.0113	0.0000	0.0044	0.0000	0.0036	0.0000	0.0000	0.0049	0.0000
	215	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0113	0.0000
Number of individuals typed		221	1059	120	114	95	137	46	39	309	117
Number of alleles		18	12	12	14	12	14	8	7	21	13
Observed Heterozygosity		0.9502	0.8536	0.8667	0.8158	0.8421	0.8248	0.8913	0.9231	0.8706	0.8803
Expected Heterozygosity		0.8849	0.8200	0.8136	0.8568	0.8433	0.8587	0.8290	0.7925	0.8645	0.8552
Null Allele Frequency		-0.0395	-0.0197	-0.0358	0.0229	-0.0049	0.0195	-0.0394	-0.0828	-0.0026	-0.0169