Use of random amplified microsatellites (RAMS) to discern genotypes of *Saprolegnia parasitica* isolates on the west coast of British Columbia

by

Cayla Naumann

B.Sc., University of Victoria, 2011

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

Supervisory Committee

Use of random amplified microsatellites (RAMS) to discern genotypes of *Saprolegnia parasitica* isolates on the west coast of British Columbia

by

Cayla Naumann

B.Sc., University of Victoria, 2011

Supervisory Committee

Dr. William Hintz, Supervisor
Department of Biology

Dr. Juergen Ehlting, Departmental Member
Department of Biology

Dr. John Taylor, Departmental Member
Department of Biology

Dr. Paul de la Bastide, Departmental Member
Department of Biology

Abstract

Supervisory Committee
Dr. William Hintz, Supervisor
Department of Biology
Dr. Juergen Ehlting, Departmental Member
Department of Biology
Dr. John Taylor, Departmental Member
Department of Biology
Dr. Paul de la Bastide, Departmental Member
Department of Biology

Several oomycete species of the genus *Saprolegnia* are recognized as devastating fish pathogens and are responsible for the loss of millions of fish annually for the aquaculture industry. Until recently, these pathogens were kept in check using malachite green; however, due to its toxicity, this chemical has now been banned from use. *Saprolegnia parasitica* is recognized as the major pathogen of aquaculture fish species. The industry is struggling to predict and control *S. parasitica* outbreaks in fish hatcheries and there is a need for new knowledge regarding the population genetic structure of this pathogen. Random amplified microsatellites were used to compare isolates of *S. parasitica* collected from a variety of hatchery locations during the period of November 2009 - August 2011, in order to determine the level of genetic variability and determine changes in genetic diversity over time. Allele frequencies of scored characters were graphically compared. Population genetic diversity was measured using Nei's genetic distance, Shannon's Information Index, number of polymorphic loci and phylogenetic trees. Due to the presence of *Saprolegnia parasitica* in the facilities tested, it appears to be ubiquitous in aquaculture facilities and treatment and prevention will be an ongoing concern in aquaculture management. Overall, genetic diversity of *S. parasitica* isolates

was determined to be low with at least some sexual recombination occurring over time. There was a diversity of genotypes collected from the same hatchery on a single day, indicating there was not a single genotype present at a given time point. Genetic profiling, such as used here, could provide facility managers with a new approach to develop a series of best practices to control sporadic outbreaks of disease. Use of these genetic markers and close monitoring of *S. parasitica* genotypes will permit early detection and sanitation protocols.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

%	percent
®	registered trademark
µg	microgram(s)
µL	microliter(s)
°C	degree Celsius
a/c	autoclaved
AAD	arbitrarily amplified dominant
Ab-GPA	glucose peptone agar with added antibiotics
AFLP	amplified fragment length polymorphisms
BC	British Columbia
bp	base pair(s)
dH_2O	distilled water
DNA	deoxyribonucleic acid
dNTP	deoxynucleotidetriphosphate
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol
EST	expressed sequence tag
g	gram(s)
G	relative centrifugal force
GPA	glucose peptone agar
GPB	glucose peptone broth

ISSR	inter-simple sequence repeat
Kb	kilo base pair(s)
LINES	long interspersed elements
M	molar
min	minute(s)
mg	milligram(s)
mL	milliliter(s)
mm	millimeter(s)
mM	millimolar(s)
mtDNA	mitochondrial DNA
Ng	nanogram(s)
PCR	polymerase chain reaction
RAMS	random amplified microsatellites
RAMPS	randomly amplified microsatellite polymorphisms
RAPD	random amplified polymorphic DNA
rDNA	ribosomal deoxyribonucleic acid
RFLP	restriction-fragment-length-polymorphism
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulphate
Sec	second(s)
SINES	short interspersed elements
SNP	single nucleotide polymorphism

SSR	simple sequence repeat
STR	simple tandem repeats
TAE	tris-acetate ethylenediaminetetraacetic acid
™	Trademark
Tris	tris(hydroxymethyl)aminomethane
U	unit
UV	ultraviolet
V	volts
VNTR	variable number of tandem repeats
v/v	volume to volume ratio
w/v	weight to volume ratio
X	times
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Acknowledgments

I would like to thank Dr. Will Hintz for starting me on the path to this thesis when I participated in the Malaysia Tropical Field Ecology course in the summer of 2010. He has been an excellent mentor and academic advisor for years now, and given me many opportunities to showcase my skills and learn more in an academic environment. I would like to thank Dr. Paul de la Bastide for guiding me with the *Saprolegnia parasitica* project and helping me with many of the culturing and lab techniques, as well as being the liaison between the hatcheries and myself. I would also like to thank Marine Harvest and their employees for providing us with samples and their expertise for this project. I would like to express my extreme gratitude to Dr. John Taylor and Dr. Juergen Ehlting for their support and guidance with my Master's thesis. They both provided valuable critique and feedback. I would certainly not have made it this far without the help of Webby Leung, who spent countless hours showing me different lab techniques, supporting me early on in my Master's work, and completing the first set of objectives for the *Saprolegnia parasitica* project. Jon LeBlanc has also been an excellent brain-stormer and sounding board from the beginning and to the end, for my various questions relating to my thesis and graduate schooling. Joyce, Irina, Erika, and Alex have all been wonderful lab mates who provided company and further advice and expertise when called upon. I would also like to give a special thanks to Chuck Groot, and his contacts Dr. Bert Buckley and Russ Pymm for their time and efforts on meeting me to discuss my data and data analysis. A big thank you also goes to my friends, family, and significant other, Nathan Groot, for supporting me through this two (plus) years' experience.

Dedication

This thesis is dedicated to my parents, Carol Naumann and Colin McDiarmid, both of whom are fellow scientists, with their own doctorates, who have encouraged my interest in science since I was a toddler. They have tirelessly helped me with many science fair projects, papers and oral presentations throughout my entire academic career. Without their constant guidance and support, my level of scholarly achievement and this thesis would not be possible.

Thesis

Introduction

Background

The order Saprolegniales represents a group of heterokonts that may assume a saprophytic or parasitic life habit and attack a wide range of hosts (Phillips *et al.*, 2007, Robertson *et al.*, 2009). Originally classified as true fungi, this group of organisms are now considered to be more closely related to brown algae and are classified as heterokonts within the Chromalveolate "super kingdom" (Beakes and Sekimoto, 2009, Beakes *et al.*, 2012). The ancestors of the order Saprolegniales were likely marine living and predominately parasites (Beakes *et al.*, 2012); modern day members of Saprolegniales live in fresh water or wet soils (Hughes, 1994). These diploid organisms reproduce and are dispersed through both sexual and asexual propagules (Robertson *et al.*, 2009) (Figure 1, pg. 2). During the asexual life cycle, mycelia grow in and on the surface of the infected host (Robertson *et al.*, 2009). Under appropriate conditions asexual sporangia develop on the hyphal tips which burst to release the apically biflagellate primary zoospores, thus dispersing clonal progeny from an original colony (Robertson *et al.*, 2009). The primary zoospores encyst to form primary cysts, which release laterally biflagellate and highly motile secondary zoospores. Once the zoospore forms a secondary cyst, it can develop into a new laterally biflagellate zoospore-infective unit or germinate to become an infective agent. The formation of a new zoospore stage is called repeated zoospore emergence (RZE) or polyplanetism and is an efficient method to find, adhere and infect a suitable host (Robertson *et al.*, 2009).

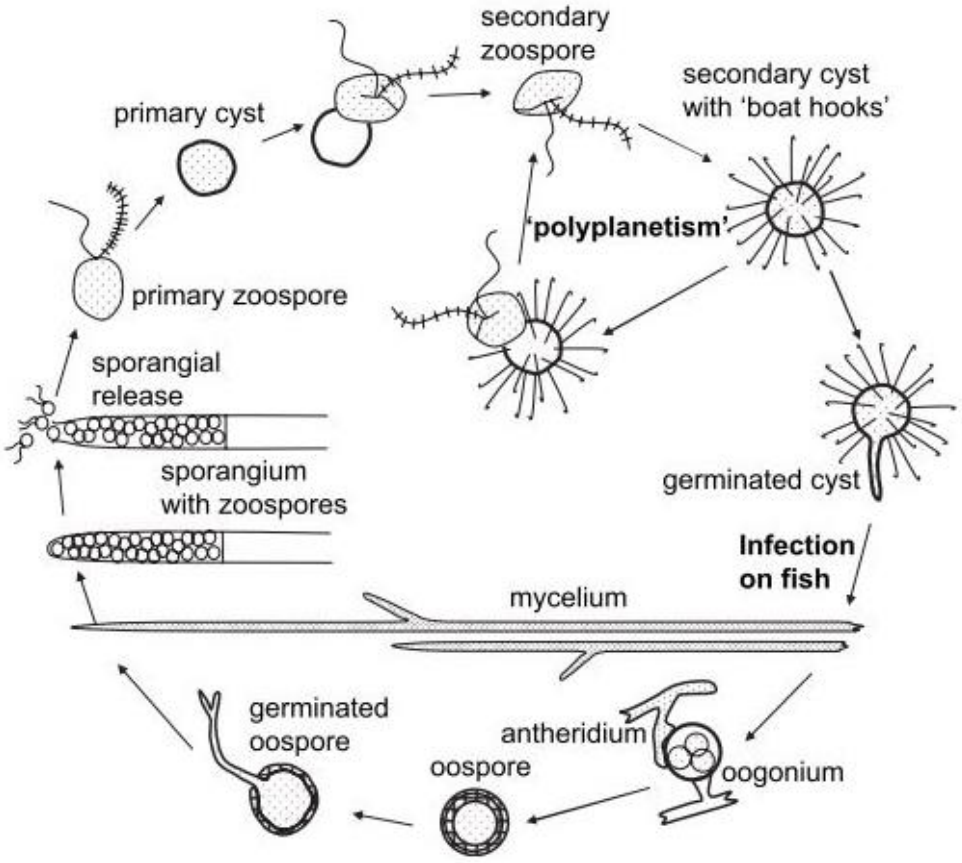


Figure 1. Schematic diagram of the lifecycle of *Saprolegnia parasitica* (Reproduced with permission from van West, 2006)

Mating is initiated by diffusible steroid hormones, which induce the production of haploid sexual structures called antheridia (nominally male structures) and oogonia (nominally female structures). Sexual reproduction occurs following gametangial contact resulting in fusion of haploid oospheres (eggs) produced in female gametangia with sperm carried by the antheridial hyphae. Following fusion of the gametes, the zygote divides and grows as a diploid hyphal filament, which eventually forms zoosporangia and produces asexual zoospores (Hughes, 1994). This complex lifecycle is summarized and depicted in Figure 1 (van West, 2006). Because of this capacity for sexual recombination, *Saprolegnia* species have the potential for significantly greater genetic variation than many parasites that reproduce through primarily asexual means (Judelson, 2009).

Many species within the Saprolegniales can cause diseases in animals, both in the wild and in captivity. Both *S. ferax* and *S. diclina* have been implicated in amphibian population declines (Blaustein *et al.*, 1994, Fernández-Benéitez *et al.*, 2008). *Saprolegnia diclina, S. salmonis* and *S. australis* are also considered significant pathogens of fish eggs (Hussein *et al.*, 2001, Robertson *et al.*, 2009), while *S. monoica* is the major cause of loss in sturgeon fish hatcheries (Phillips *et al.*, 2007). *Saprolegnia diclina* has been found to be the most prevalent species in Norwegian salmon hatcheries (Thoen, 2011); whereas, *Saprolegnia parasitica* (Figure 2) has been detected more commonly in BC hatcheries (Leung, 2012).

Saprolegnia parasitica is believed to be the primary pathogen leading to a condition known as saprolegniosis, a disease characterized by white or grey patches of cotton wool-like filamentous mycelia growth on epidermal lesions of the fish's body.

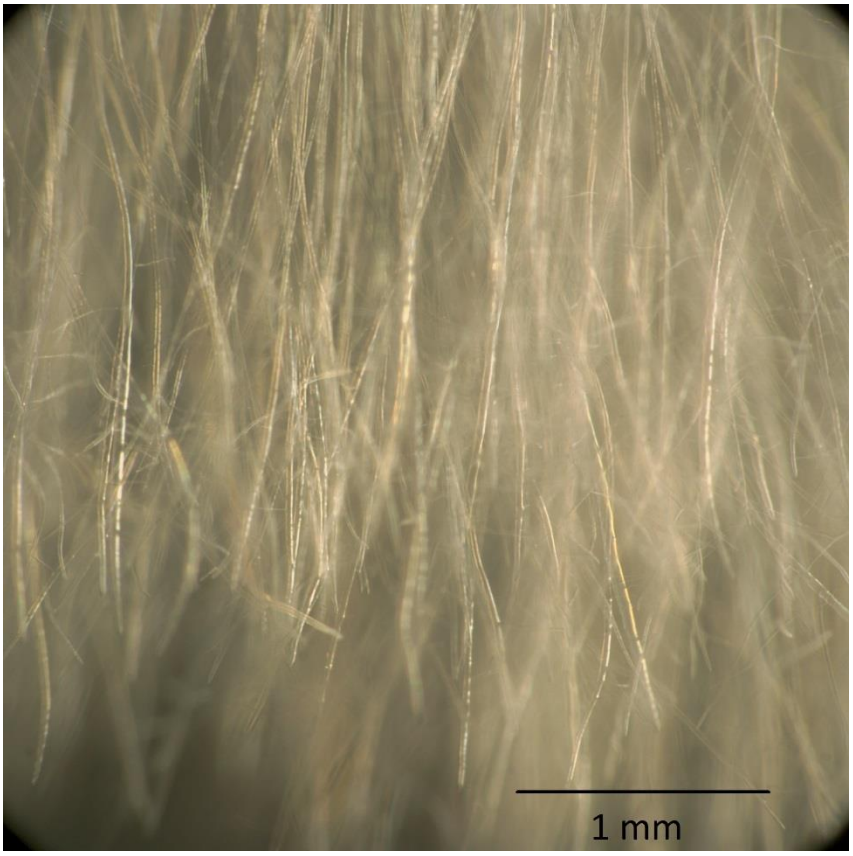


Figure 2. Image of leading edge of mycelial fibers of *Saprolegnia parasitica* cultured from hatchery water sample at 400X magnification under a compound microscope (Photo credit Cayla Naumann, 2013)

This disease can lead to fish death depending on the severity of infection, initial health of the fish and other factors (van West, 2006, Phillips *et al.*, 2007, Robertson *et al.*, 2009). The infection initially manifests on either the fish head, tail or fins, as seen in Figure 3 and then spreads to the rest of the body (Robertson *et al.*, 2009). Typically the infected fish succumb to imbalanced osmoregulation which results in hemodilution (Meyer, 1991, Robertson *et al.*, 2009). One of the characteristics that may distinguish pathogenic *S. parasitica* from closely related non-pathogenic species is the presence of grouped long, hooked hairs on the secondary cysts, compared to shorter singular hooks found in other species (Pickering *et al.*, 1979, Beakes, 1983, Beakes *et al.*, 1994, Fregeneda-Grandes *et al.*, 2000). This observation was also supported by Stueland *et al.* (2005) who indicated that the long hairs on the germinating sporocysts correlated with high initial growth rate and were indicative of pathogenicity on juvenile salmon. Theories as to why these grouped, long, hooked hairs may aid in pathogenicity include facilitating the adhesion of sporocysts to the host, advancing buoyancy to ensure presence in water column, and easing attachment to the host (Beakes, 1983).

Traditional classification of *Saprolegnia* species has been based on descriptions of sexual reproductive structures; however, many isolates fail to produce these structures *in vitro* (Hatai *et al.*, 1990, Stueland *et al.*, 2005, Diéguez-Uribeondo *et al.*, 2007). This has led to the inaccurate identification of *Saprolegnia* species and has complicated the taxonomy of this genus. The taxonomy and phylogeny of the family Saprolegniaceae has recently been improved through the analysis of internal transcribed spacer (ITS) nucleotides sequences and large subunit (LSU) of the ribosomal DNA repeat

Figure 3. ***Salmo salar*** **juvenile infected with** ***Saprolegnia*** **sp. (Photo credit Cayla Naumann, 2010)**

(Leclerc *et al.*, 2000, Diéguez-Uribeondo *et al.*, 2007, Hulvey *et al.*, 2007, Petrisko *et al.*, 2008, Ke *et al.*, 2009). Novel PCR primers for the ITS region have been successfully used to distinguish the genus *Saprolegnia* within the family Saprolegniaceae, but lack the resolution to distinguish individual species within this genus (Leung, 2012). Certain *puf* primers, 112 and 310, which amplify a portion of the *Pumilio* locus in *S. parasitica*, have been recently developed to identify *Saprolegnia parasitica* isolates (Leung, 2012). Although the physiology and life cycles of *Saprolegnia* species, and specifically *S. parasitica*, have been well described, the details of the mechanisms of its pathogenicity, host specificity and population structure are not well understood (Robertson *et al.*, 2009). This leaves a significant knowledge gap in understanding within-species variability for *S. parasitica* populations. Information on intraspecific variability could be very useful in understanding *S. parasitica* as a parasitic or opportunistic pathogen in aquaculture facilities.

Impact of *Saprolegnia parasitica*

Saprolegnia parasitica is one of the most devastating and destructive oomycete fish pathogens characterized and causes tens of millions of pounds of fish loss annually worldwide (van West, 2006). Fish loss in aquaculture facilities is primarily caused by bacterial diseases, but this is closely followed by fungal or fungal like infections, including loss due to saprolegniosis. Saprolegniosis accounts for approximately 10% of salmon loss in fish farms (Phillips *et al.*, 2007, Robertson *et al.*, 2009). Bacterial and fungal-like pathogens (oomycete) combined, these factors represent the greatest economic loss to the aquaculture industry (Meyer, 1991). *Saprolegnia parasitica* is also believed to cause significant losses in wild fish populations (van West, 2006).

Saprolegniosis is found exclusively in fresh water and could affect the fish eggs and juvenile and spawning fish. Although most Pacific salmon species die shortly after spawning (Altukhov *et al.*, 2000), saprolegniosis infection could result in death prior to spawning or reduce spawning fecundity. For example, pre-spawning salmon in the northwest have been found to suffer a 22% loss of mature returning salmon due to head wounds infected with bacterial and fungal pathogens such as *S. parasitica* (Neitzel *et al.*, 2004). Northwest wild salmon populations are also exposed to other environmental stresses, including anthropogenic factors such as warming waters due to human activities, making them even more susceptible to disease and infection (Driscoll, 2004). Wild Atlantic salmon populations are iteroparous (spawning multiple times before death) (Willson, 1997); therefore, saprolegniosis infections may result in losses in reproductive potential as described above, as well as loss in future reproductive events.

Current treatment options for saprolegniosis

In recent years commercial fish production and human consumption demand has become dependent on the fish farming industry in order to provide an adequate supply of fresh product (Robertson *et al.*, 2009). Global aquaculture production continues to increase and accounts for 47% of the total fish production for human consumption (Food and Agriculture Organization of the United Nations, 2012). This represents the world's fastest growing food sector. Current aquaculture losses directly attributable to *S. parasitica*, in combination with the potential for uncontrolled losses due to saprolegniosis, represent a significant risk for aquaculture industries around the world.

This oomycete pathogen was originally controlled using a chemical treatment known as malachite green. Although called malachite green, the compound is not related

to the mineral malachite, but is in fact classified in the dyestuff industry as triarylmethane dye. There has been a worldwide ban for the use of this chemical for food-related uses since 2002, when it toxicological and carcinogenic effects were realized (Robertson *et al.*, 2009).

Several alternatives to the use of malachite green have been implemented in fish hatcheries to control saprolegniosis, but all have demonstrated reduced efficacy compared to malachite green. These include formalin (Gieseker *et al.*, 2006), copper sulfate (Sun *et al.*, 2014), diquat bromide (Mitchell *et al.*, 2010), amphotericin B, hydrogen peroxide (Howe *et al.*, 1999, Robertson *et al.*, 2009), sodium chloride (Ali, 2005), bronopol (Pyceze®) (Pottinger and Day, 1999, Aller-Gancedo and Fregeneda-Grandes, 2007), and nikkomycin Z (Guerriero *et al.*, 2010). There is certainly a need for a novel and environmentally safe treatment, but there are currently no candidates available. From the perspective of an afflicted animal, one can only look towards the availability of a vaccine against *Saprolegnia* via injection of an antigen into the fish muscle tissue; however, this relies on the discovery of the correct antigen (Robertson *et al.*, 2009). This leaves few effective methods to control the fungus in aquaculture facilities, resulting in more saprolegniosis infections and continued losses for the industry.

Current research

In the fall of 2013, there was a conference with several presentations related to *Saprolegnia* infection including the immune response of salmon to infection, potential vaccines to prevent infection, controlling reproduction of *Saprolegnia*, potential bacterial control agents of *Saprolegnia*, and the cell wall and protein structure of *Saprolegnia*

(SAPRO: Sustainable Approaches to Reduce Oomycete Infection in Aquaculture, 2013). Thoen (2011) published a Doctoral thesis on *Saprolegnia* infections in Norwegian salmon hatcheries that presented an overview of the quantities and species of *Saprolegnia* in Norwegian salmon hatcheries, characterized the isolates, provided information on differences in virulence between isolates from eggs and parr of Atlantic salmon, as well as recommendations for managerial factors vital for prevention of saprolegniosis. The genomes and annotations of *S. parasitica* CBS 223.65, *Saprolegnia declina* VS20, and four *Phytophthora* species are available through the Broad Institute of Harvard and MIT (2010). The *S. parasitica* genome, the first oomycete pathogen of animals to be sequenced, is 53.09 Mb of complete genome sequence with a total contig length of 48.14 Mb and over 20, 000 genes and contains specific adaptations for its host (Jiang *et al.*, 2013). With the completion of the *Saprolegnia parasitica* genome, one of the next areas of study to greatly expand will be reverse genetics to determine the function of various genes and understand the biological importance of such genes (Bhadauria *et al.*, 2009).

These and similar projects centre on the problem created by *Saprolegnia* infections and how to reduce its impact or identify specific characteristics within the *S. parasitica* DNA structure, but fail to analyze *Saprolegnia parasitica* population structure or answer any questions on intraspecific species diversity, that may lead to a better understanding of the pathogen's spread and virulence within aquaculture facilities.

Current molecular methods and their potential use with *Saprolegnia parasitica*

Currently, hatchery managers are faced with a variety of decisions regarding management and treatment of saprolegniosis in a facility. One factor that needs to be clarified is to determine the source of the *S. parasitica* inoculum. There are a few ideas

on the causes of significant outbreaks of saprolegniosis in a given facility. One possibility is that a new community of the pathogen is introduced to a tank or facility, and that this novel inoculum causes a saprolegniosis outbreak, possibly because it is more virulent or the hosts are less adapted to resist the introduced isolates. A second option is that the pathogen is always present at a relatively low level, but some triggering environmental event either makes the fish more susceptible to infection or increases the concentration of inoculum present and consequently the rate of infection. Some support for this latter idea has been observed by Bly *et al.* (1992), who looked at winter saprolegniosis death in channel catfish and determined it was an immunodeficiency disease caused by unknown *Saprolegnia* species.

Support for either of these possibilities has different implications for management as to where to focus treatment and prevention efforts. If isolates of *S. parasitica* are determined to be relatively similar among various locations where they might be cross-contaminated, a new introduction of the pathogen causing an outbreak is likely simply due to the stress of the introduction of the fish, and not due to a new type of isolate. If, however, isolates are significantly different, it is possible that certain isolates of *S. parasitica* are more pathogenic than others, leading to the need to identify pathogenicity factors and the most pathogenic genotypes. If this is the case, hatchery mangers can target detection, treatment and prevention of the more pathogenic genotypes.

In order to assess whether disease outbreaks result from such pathogenic isolates, we need to first assess the overall genetic diversity of the population, and then determine how the population changes over time. Highly similar or clonal isolates may also support the hypothesis that a closed population is reproducing asexually. In a clonally

reproducing population, all of the isolates would have a very similar or identical genotype. In a sexually reproducing population, there would be a diversity of recombinant types within the population; however, the population diversity over time would remain the same, as long as no new genotypes were introduced. Only with the introduction of novel and new genotypes into a closed system, like a hatchery, would one see a dramatic increase in the genetic diversity of the population. By analysing the genotypes and diversity of isolates collected from various hatcheries over time, we can better understand how the pathogen is being introduced, why it is persisting and what might be the best method to prevent and treat infection in the future.

There are a variety of ways to measure intraspecific variation in a population, especially with the consistently decreasing cost and expertise required for many molecular techniques. They generally focus on ways to detect DNA polymorphisms, which are a form of genetic marker, particularly in the more variable non-coding regions of DNA. Genetic markers are heritable traits with allelic differences possible at a given location (Sunnucks, 2000). For a diploid organism, there are generally two possible alleles for each locus (Sunnucks, 2000).

Polymerase chain reaction (PCR) can be useful for detecting many different kinds of DNA polymorphisms. PCR very specifically amplifies nucleotide sequences from a sample DNA (Altukhov, 2006). This method relies on *Taq* DNA polymerase and (generally) two primers to create a complementary sequence of the sample DNA between the primer attachment sites. It is a more effective method than cloning, for short sequences of DNA (Altukhov, 2006). Products of PCR are usually visualized on agarose

or polyacrylamide gels and visualized using stains and specialized lighting depending on the type of fragment amplified and stain used.

A large proportion of nuclear DNA consists of tandemly repeated copies, often created by insertion or deletion mutations which alter the number of repeats. The variation in repeats can be used to differentiate genotypes. This variation is termed VNTR (variable number of tandem repeats) (Jeffreys *et al.*, 1985, Levinson and Gutman, 1987, Altukhov, 2006) and these variations are used to examine individual genome loci in population genetic studies (Altukhov, 2006). Most population biology studies focus on minisatellites of 9-100 base pairs or microsatellites of 1-6 base pairs (also called SSRs-simple sequence repeats or STRs-simple tandem repeats). They are common, highly variable and very common in eukaryotic genomes (Tautz and Renz, 1984). Molecular markers used to detect DNA polymorphisms can be roughly divided into dominant or co-dominant markers and specific or non-specific (*i.e.* arbitrary) markers.

Dominant markers allow for the study of multiple loci at once, are able to visualize many loci simultaneously and include random amplified polymorphic DNA markers (RAPD), amplified fragment length polymorphisms (AFLPs), inter-simple sequence repeats (ISSRs) and RAMS (randomly amplified microsatellites) (Sunnucks, 2000). Dominant markers can only be scored by their presence or absence, and it is impossible to determine a difference in zygosity from the amplifications (Sunnucks, 2000). For dominant marker scoring it is assumed that "presence" is one allele and "absence" is another (Zhivotovsky, 1999). Co-dominant markers measure a single locus where both alleles can be identified and include restriction-fragment-length-polymorphisms (RFLPs), microsatellites, and minisatellites (Sunnucks, 2000). Specific

primers may be designed to bind to a known target DNA sequence and require prior sequence knowledge to amplify the region of interest; whereas, non-specific arbitrary markers can be used on a wider variety of DNA samples and taxonomic groups because they do not require prior knowledge of the DNA target sequence.

RFLPs are one way of measuring DNA polymorphisms. Restriction endonucleases fragment the DNA at specific demethylated DNA sequences to create DNA sub-fragments (Altukhov, 2006). The DNA polymorphisms observed result from differences in the length of DNA fragments produced by the restriction endonucleases (Altukhov, 2006). They are very often used to determine nucleotide sequences after cloning, using a specific probe and Southern blotting, but are ineffective for separating a mixture of numerous fragments (Altukhov, 2006). It is also a time consuming process that requires large amounts of high quality DNA (Hantula *et al.*, 1996).

RAMS or RAMPs (randomly amplified microsatellite polymorphisms) amplify microsatellites and the sequence between them using non-specific primers, thus combining the universality of random amplified polymorphic DNA markers (RAPD) analysis and the benefits of microsatellites. It should be cautioned, however, that RAMS and RAMPs are dominant markers and are not co-dominant like the specifically-binding microsatellite primers. Wright & Bentzen (1994) outlined several advantages to using micro- or minisatellites including their high frequency of occurrence, random dispersal throughout the genome, rapid evolution, co-dominant Mendelian inheritance, and their location in mainly non-coding regions and, therefore, neutral selection. From a technical standpoint, use of PCR and these micro- or minisatellites requires little tissue, blood or DNA sample and automated analysis is possible.

RAPD priming sites are slightly longer than microsatellite primers at 10-20 base pairs of random sequences (Williams *et al.*, 1993, Altukhov, 2006). The amplified products depend on the length and sequence of the exact primer used and polymorphisms are expressed and quantified by presence or absence on a gel (Williams *et al.*, 1993, Altukhov, 2006). These markers have been criticized for over-estimating inter-specific genetic variation (Powell *et al.*, 1996, Zhivotovsky, 1999, Nybom, 2004).

Luikart *et al.* (2003) reviewed several molecular techniques including amplified fragment length polymorphisms (AFLPs), diversity array technology (DArT), microsatellites, single nucleotide polymorphisms (SNPs) and sequence data. These authors emphasized that AFLPs or modified techniques of AFLPs and microsatellites have the advantage of uncovering hundreds of polymorphic markers in an entire genome with ease and reasonable cost and high reliability because they generate dozens of bands (or amplicons) in a single gel lane. AFLPs operate by selective amplification of fragments of genomic DNA created by restriction enzymes (Altukhov, 2006). AFLP pattern generation involves RE digestion of the DNA, binding sticky fragment ends with oligonucleotides and then the use of PCR to selectively amplify the restriction fragment (Altukhov, 2006). As mentioned previously, AFLPs are also dominant markers (Luikart *et al.*, 2003). SNPs are represented by variable substitutions of a single nucleotide in a DNA sequence and have been extensively studied in the human genome (Altukhov, 2006); however, they are not as useful at measuring intraspecific variation due to their difficulty in measuring them.

Expressed sequence tags (ESTs) can be used to detect polymorphisms such as insertions or deletions within expressed coding genome sequences. They are fragments

or complete sequences of complementary DNA, obtained with reverse transcriptase from mRNA (Altukhov, 2006). Primers are used to amplify ESTs from genomic DNA and are examined using methods of amplification product analysis. ESTs are primarily used for gene mapping (Altukhov, 2006). Other elements include SINES (short interspersed elements) and LINES (long interspersed elements). These are repeated, unblocked and dispersed throughout genome sequences and are included in genomic transcripts of intracellular DNA (Altukhov, 2006). Mitochondrial DNA and mitochondrial control regions are also used to track uniparental transfer of DNA, particularly in humans and other vertebrates. These are genotyped by sequencing and generally lack recombination; therefore, they are limited in their use for measuring intraspecific genetic variation (Altukhov, 2006).

I used SSRs and RAMs, which particularly target regions with microsatellite repeats to detect genotypic variations in isolates of *Saprolegnia parasitica*. These primers are of particularly interest because these regions are highly variable, so they will show as much genetic variation as possible, rather than targeting other less variable regions or SNPs. SSR and RAM primers have relatively simple molecular techniques required and easily reproducible results. Many researchers have used microsatellite or SSRs along with statistical methods to measure intraspecific genetic variation. Hantula *et al.* (1996) used SSRs and RAM primers with degenerate ends to detect interspecific and intraspecific DNA-polymorphisms for six species of fungi. Wang *et al.* (2009) measured the genetic diversity of two different geographic populations of *Phytophthora sojae*, an oomycete and a soil-borne plant pathogen that causes stem and root rot of soybean, using 20 pairs of specifically targeting SSR primers to separate 83 isolates into seven clustering

groups based upon Nei's genetic distance. Aboukhaddour *et al.* (2011) also used Nei's genetic distance, as well as analysis of molecular variance (AMOVA), to distinguish the genetic diversity and relatedness of 80 isolates of *Pyrenophora tritici-repentis*, a wheat fungal pathogen, using thirty-one SSR markers. Similar techniques and statistical measures could be used to distinguish isolates of *S. parasitica* collected from BC aquaculture facilities.

When trying to measure the genotype of a given organism or population, we can randomly sample genetic loci to get an estimate of genetic variability (Nei, 1987). A sufficient number of isolates collected at a given time in a closed population can represent the genetic diversity of those individual isolates and may be extrapolated to the population diversity as a whole, and also allow for comparison of genetic diversity over time. Species diversity of the isolates can be assessed using various genetic markers, as described above, with the correct level of resolution to differentiate *S. parasitica* isolates. Certain genetic markers may resolve isolates only at the species level, while other markers may be able to distinguish sub-populations within a species.

When using SSRs or non-specific dominant primers (such as RAMS), each amplification is scored as "presence"; the genotypes that do not have an amplification for a given locus are scored as "absence". The presence or absence of alleles of all possible loci can then be combined into a single character string and analyzed using population genetic and diversity measures such as Nei's genetic distance, analysis of molecular variance (AMOVA), the construction of phylogenetic trees as well as the determination of the Shannon Information Index to assess diversity. Population genetics statistics such as Nei's distance are based on the average identities of randomly chosen markers within

and between populations or samples and are appropriate for populations with multiple alleles per locus or populations shaped by diverse evolutionary forces (Wang *et al.*, 2009). These population genetics statistics and phylogenetic analyses will provide information on the diversity and genotypic variation among isolates of *Saprolegnia parasitica* collected from BC aquaculture facilities that can be used to broaden the understanding of this oomycete fish pathogen.

Overall project objectives

While there are only limited genetic diversity studies of natural populations of *Saprolegnia* collected from the field virtually nothing is known about the genetic diversity of *S. parasitica* in contained aquaculture facilities. The levels of diversity between contained systems like hatcheries and the natural environment are likely quite different. Diversity in hatcheries may depend on the method of introduction (on fish or through groundwater), the spread of infection within the system, and whether an isolate is ever completely removed from the system with treatment. In order to better control saprolegniosis, a better understanding of the population diversity of *S. parasitica* is essential. This can be achieved through the development of a molecular marker system to evaluate the genetic structure of the pathogen population in Canadian fish hatcheries, and this could be extended to facilities worldwide.

In British Columbia, hatcheries potentially share isolates of *S. parasitica* through the transfer of eggs and juvenile salmon among facilities, which happens on a semi-annual basis (Boyce, personal communication). Several hatcheries are suppliers of fish eggs and juvenile fish to other facilities (Boyce, personal communication). Hatcheries also receive eggs or juvenile salmon from locations outside of British Columbia.

Because of the lack of quarantine when new eggs or fish are brought into a facility and because eggs and fish are moved throughout the facility as they grow or are vaccinated, the entire hatchery is vulnerable to the introduction and propagation of novel isolates of *S. parasitica* (Boyce, personal communication). This potential regular movement of *S. parasitica* genotypes throughout the facilities could possibly generate very high genetic diversity due to the regular exchange of different genotypes leading to a greater opportunity for sexual crossing of isolates. Alternately, there could be a very low apparent diversity as a limited number of genotypes might be evenly distributed amongst the various hatcheries from very few sources.

Previous work in our laboratory developed methods to rapidly and easily identify *S. parasitica* compared to other species in the genus (Leung, 2012). RAMS and SSR markers have been be used in other experiments to amplify variable regions of the DNA to create a unique profile of amplified characters, and these markers could work similarly to compare genotypes of *S. parasitica*. Presence or absence of these amplified characters can then be measured and used to highlight patterns in genotypic diversity. The assumption is that more closely related individuals will share a greater number of amplicons and more distantly related individuals will display unique amplicon characters. Unique amplicon characters may be derived from the presence or absence of a given primer binding site due to SNPs, variation in the length of the microsatellite repeat, or differences in zygosity between isolates. Phylogenetic trees and population statistical methods can then be used to determine if specific genotypes are related and correlated with sample collection factors. This will provide information on tracking the pathogen within and among facilities, as well as information on reproduction and differentiation of

the species at a given collection site. It is anticipated that a better understanding of population diversity for *S. parasitica* in hatcheries will contribute to the development of more effective disease management strategies and improved fish health.

Materials and Methods

Field sample collection

In order to measure the diversity of samples, a culture collection of *S. parasitica* needed to be developed. Samples were collected over time and from various locations to try and collect a diversity of genotypes that may be present in the various hatcheries at different temporal periods. Samples were collected predominantly by hatchery staff and shipped to the lab for processing. Dead fish collected from tanks, water samples from the tank water column, eggs with evidence of *Saprolegnia* infections and swabs of hard tank surfaces within the facilities were included in substrate types. Sample fish collected included those with obvious saprolegniosis and those without obvious saprolegniosis infection. Water samples included those from within the hatchery and sources that supplied the hatchery. Samples as described above were collected from various fish hatcheries and freshwater locations across the west coast of British Columbia, Canada (Table 1, 22 and Figure 4, 23). Samples were shipped in coolers maintained at 4°C and processed within 24 hours of receiving samples and within 72 hours of sample collection at the hatchery or field site.

Sample site descriptions

Nanaimo River is a private hatchery that raises eggs and juvenile salmon *Oncorhynchus gorbuscha*, *Oncorhynchus tshawtscha* and *Oncorhynchus keta* species (Pink, Chinook, and Chum salmon, respectively) for release to supplement sport fishing and wild fish stocks. Puntledge River is a Department of Fisheries and Oceans (DFO)

Table 1. Sample collection locations according to hatchery and geographical location. All locations are in British Columbia, Canada.

Location name	Abbreviation	Affiliation	Geographical location
Nanaimo River	NR	Private	South of Nanaimo
Puntledge River	PL	DFO	Courtney
Sayward Hatchery North	SN	Marine Harvest	Near Campbell River
Sayward Hatchery South	SS	Marine Harvest	Near Campbell River
United Hatchery	UH	Private	Fanny Bay
Ocean Falls	OF	Marine Harvest	Near Bella Coola
Georgie Lake	GL	Marine Harvest	Near Port Hardy
Stelling Hatchery	SH	Private	Fanny Bay
Upper Goldstream River	UG	Private	Near Langford

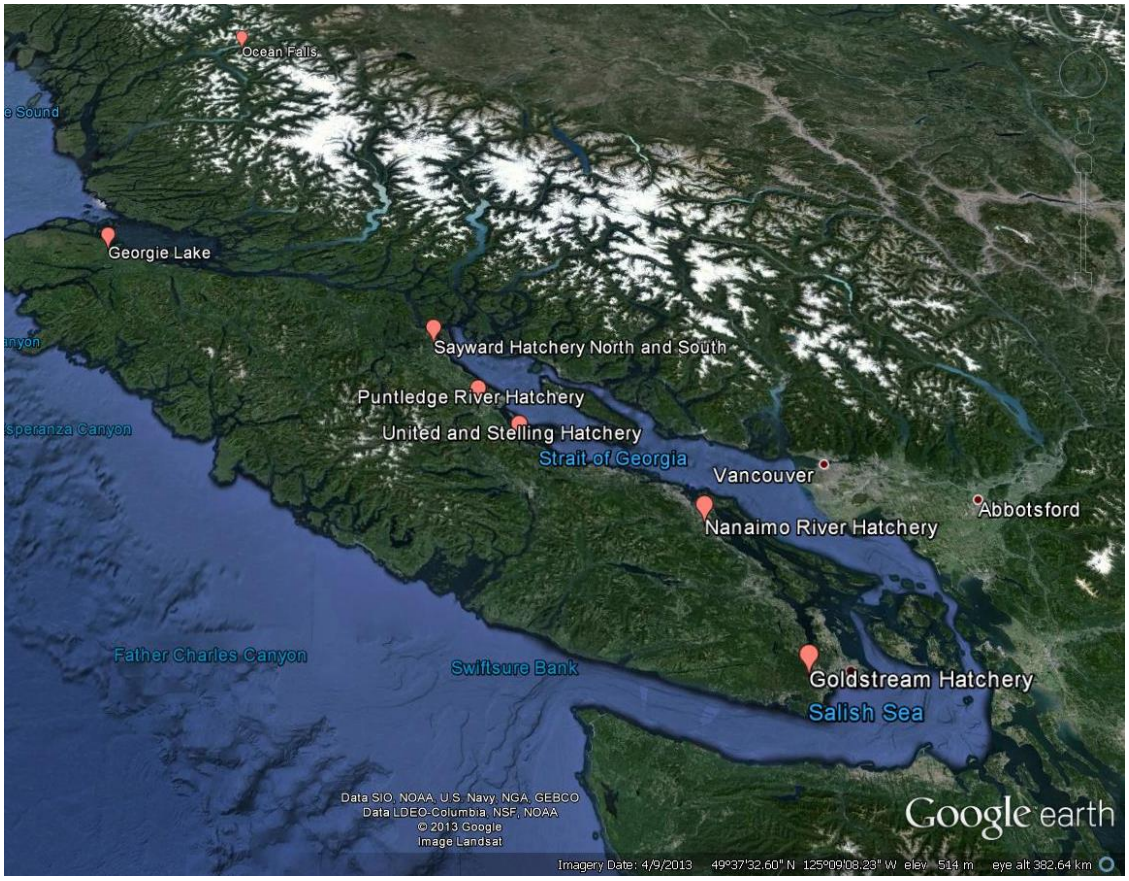


Figure 4. Map of sample collection locations. Approximate locations marked with red markers with white text labels adjacent.

hatchery that functions similar to Nanaimo River, hatching and growing wild salmon species until they can be released. Sayward Hatchery North and South, Ocean Falls, and Georgie Lake are Marine Harvest hatcheries that produce *Salmo salar* (Atlantic salmon) for human consumption. Georgie Lake is unique that is an open aquaculture system where the juvenile salmon are grown in net pens within the lake. United Hatchery and Stelling Hatchery supply Marine Harvest with Atlantic salmon eggs and fish fry (*Salmo salar)* to be grown into adult fish for human consumption. There is a hatchery at Goldstream River (http://www3.telus.net/gvsea/#) that functions similar to Nanaimo River and Puntledge River Hatcheries; however, we only collected samples from Goldstream River directly. Where possible, water samples were collected from outside the hatchery from the water source used at the hatchery.

Sample and isolate processing

Once samples were received and before infected tissue and samples were excised, dead fish were processed by gently rinsing with at least three exchanges of water to remove excess slime and mucous resulting from secondary bacterial infections. Each infected fish was carefully examined for external lesions and evidence of saprolegniosis (Figure 3). Sections of infected tissue were excised, rinsed three times with autoclaved distilled water (a/c dH_2O) and placed in sterile petri dishes (100 x 15mm) with 15mL of a/c dH_2O in an aseptic environment. Autoclaved hemp seeds were added as a bait substrate. The hemp seeds were visually observed for evidence of filamentous mycelial growth every 24 hours for up to two weeks. Once filamentous mycelial growth was observed, baited hemp seeds were aseptically transferred to glucose peptone agar (GPA, 3g/L D-glucose, 1.25g/L bacto peptone and 15g/L agar) plates augmented with four

antibiotics (Ab-GPA plates) to select for cultures uncontaminated by bacteria or ascomycetous (higher) fungi. Antibiotics added were: Rifampicin (Calbiochem, La Jolla CA, USA) at 50 µg/mL (stock of 50mg/mL in DMSO), Nystatin N1638 (Sigma-Aldrich, St. Louis MO, USA) at 10 µg/mL (provided as stock of 10,000 U/mL), Chloramphenicol (Sigma) at 25 µg/mL (stock of 25 µg/mL in 100% EtOH) and Streptomycin (Calbiochem) at 10µg/mL (stock of 10mg/mL in sterile dH_2O). Individual colonies were grown for three to five days whereupon a colony subsection measuring approximately 5 x 5mm was removed from the growing edge of the colony and transferred to a new Ab-GPA plate. This was repeated at least three times to ensure that the cultures were contaminant-free and were representative of a single genetic individual and not derived from mixed cultures.

Culturing from egg samples essentially followed the same protocol as was used for the fish samples. Obviously infected or suspect eggs were transferred to sterile petri dishes with 15 mL of water baited with hemp seeds. For water samples, 15-20 mL of the water sample was poured into a sterile petri dish and autoclaved hemp seeds bait was added. Four plates were poured per water sample. Once mycelial growth appeared on the hemp seeds, the same protocol as used for fish and fish egg samples was followed. In a few instances, swabs of hard surfaces or the outside of fish were taken and cultured by removing and leaving the end of the swab in a sterile petri dish, with 15-20 mL of sterile dH_2O and sterile hemp seeds as bait. Once a hemp seed showed signs of *Saprolegnia* growth it was transferred to an Ab-GPA plate and similarly processed. Only one hemp seed culture was removed from the baited plates, even if multiple hemp seeds showed growth.

Prior to DNA extraction, a square approximately 5 x 5mm was removed from the leading edge of the colony and transferred to a 125 mL Erlenmeyer flask containing 50 mL of glucose peptone broth (GPB, 3g/L D-glucose, 1.25g/L bacto peptone) and maintained at ambient temperature (23°C) until log phase growth was attained (approximately three to four weeks). Cultures were rinsed three times with a/c dH_2O, harvested by vacuum filtration, quick frozen in liquid nitrogen and immediately lyophilized for at least 48 hours. The freeze-dried samples were stored at -20° C prior to DNA extraction.

DNA isolation

DNA was extracted from freeze-dried mycelium using the protocol of Möller *et al.* (1992), with minor modifications. Approximately 30 to 60 mg of lyophilized mycelium was ground to a powder with 100 mg of a/c zirconium/silica beads (0.5mm diameter, Fisher Scientific, Canada) and 100 µL TES buffer (100 mM Tris pH 8.0, 10 mM EDTA, 2% SDS) inside a 1.5 mL microfuge tube by use of a bead beater (MINI Beadbeater™, Biospec Products). There were 3 rounds of 30 seconds beating which were interspersed with 10 second centrifugation at 13,000G to ensure complete maceration of the tissues. Once the mycelium was homogenized 400 µL TES and 50 µL Proteinase K (2 mg/ mL) were added and incubated at 55° C for 30 minutes. Salt solution was adjusted to 1.4 M by adding 140 µL of 5M NaCl and 65 µL 10 % CTAB (cetyltrimethylammoniumbromide) and incubated at 65° C for 10 minutes. Samples were centrifuged at 13,000G for 10 min. The supernatant was removed and combined with 700 µL SEVAG (chloroform: isoamylalcohol, 24:1) and placed on ice (4° C) for 10 min. Centrifugation, supernatant removal and combination with TES were repeated twice.

Supernatant layer was removed combined with 225 µL 5M NH_4Ac and iced (4° C) for 40 minutes. Samples were centrifuged for 10 minutes and the supernatant was combined with 510 µL isopropanol. Samples were iced (4° C) for 10 min and centrifuged at 13,000G for 5 min. Pellets were washed twice with 4° C EtOH. The extracted nucleic acids were re-suspended in 50 µL UltraPure™ distilled water (Invitrogen, Grand Island, New York, USA). The quality, concentration and ratio of DNA to RNA were analyzed using the Nanodrop® ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA), prior to preparing DNA template dilutions of 5 ng/µL and 10 ng/µL concentrations for PCR amplification. A ratio of ~1.8 is generally accepted for pure DNA and a ratio of ~2.0 for RNA (Thermo Fisher Scientific, 2014). A low ratio may be cause by residual phenol or other extraction reagents or a very low concentration of nucleic acids (<10 ng/µL). High ratios are not indicative of an issue (Thermo Fisher Scientific, 2014).

PCR amplification of the ITS region, nucleotide sequencing and isolate identification

The universal ITS region primer pair ITS4 (TCCTCCGCTTATTGATATGC) and ITS5 (GGAAGTAAAAGTCGTAACAAGG) (White *et al.*, 1990) was used to amplify the nucleotide sequence between the internal transcribed spacer 1 (ITS1) and 2 (ITS2) of the rRNA cistron, including the 5.8S region. The annealing sites of ITS4 and ITS5 are close together, with the 5’ end of ITS5 annealing two base pairs upstream of the 5’ end of ITS1, when using *S. parasitica* as a template. Each PCR reaction was performed in 20 µL final volumes using one unit of Fermentas DreamTaq DNA polymerase, a final concentration of 0.5 µM for each primer and 20.0 ng of genomic DNA per 20 µL reaction

volume. All PCR amplification reactions using these primers were performed using Eppendorf Mastercycler® gradient model 5331 and followed the reaction conditions described by Diéguez-Uribeondo *et al.* (2007). The DNA was initially denatured for 5 min at 94° C followed by five cycles of denaturation for 30 sec at 94° C, annealing for 30 sec at 55° C and extension for 1 min at 72° C. This was followed by 33 cycles of denaturation for 30 sec at 94° C, annealing for 30 sec at 48° C, and extension for 1 min at 72° C. There was a final extension for 10 min at 72° C and the PCR samples were held at 4°C until processed. A 5.0 µL volume of each amplification product was mixed 2 µL of 1:10 diluted loading dye (0.25% w/v bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol in H_2O) and loaded into each well of a 1.5% w/v agarose gel, separated by gel electrophoresis (97 volts for 1 hour 12 min) and stained with GelRed (3X staining solution from 10,000X stock, w/v) and visualized by illumination with UV light in a GelDocXR+ with Image Lab Software (version 4.1 build 16) (Bio-Rad Laboratories (Canada) Ltd., Mississauga, OT).

Amplified products (ITS5 and ITS4) of template DNA were initially (February to July 2010) sent without purification to the Macrogen direct sequencing service (Macrogen, Rockville, USA). Samples collected August 2010 and later were purified using QIAquick PCR purification Kit (Qiagen, Germantown, WI, USA) and sent to Eurofins mwglOperon (Operon, Hunstville, AL, USA) for direct DNA sequencing. Sequencing results were visually analysed and manipulated using the BioEdit Sequencing Alignment Editor (version 7.0.9.0) (Hall, 1999). Each sample was sequenced in both directions and both reaction sequences were compared to create a consensus sequence. Sequences were subjected to a blastn (nucleotide query/nucleotide database search

option) search, using default parameters of the National Centre for Biotechnology Information (NCBI, web resource) database and the Identification Engine under the category of "Fungal identification—ITS search" of the Barcode of Life Data System v2.5 (BOLD, web resouce). Only samples confirmed as *S. parasitica*, based on the results obtained from both databases, were used for further analysis. After February 2011 and the development of the techniques by Leung (2012), samples putatively assigned as *S. parasitica* according to ITS sequence were positively identified using *puf* primers 112 and 310.

Primer development and selection

Degenerate primers designed and used by Hantula *et al.* (1996) were used to genotype isolates of *S. parasitica* (Table 2, 31). Additional degenerate primers were developed according to the abundance of simple sequence repeats in expressed sequence tag (EST) libraries of fungal and oomycete genomes, and those used by others for similar genotyping analysis (Van der Nest *et al.*, 2000, Karaoglu *et al.*, 2005, Lee and Moorman, 2008) (Table 3, 31). The newly developed primers used the same degenerate 5' ends as those of Hantula *et al.* (1996). Primers of interest were tested across a gradient of annealing temperatures (Table 4, 32) to determine optimum annealing temperature, based on distinct and consistent band amplifications for as wide of a temperature range as possible. Primers were also selected and used for final genotype comparison based on their ability to show clear, distinct amplifications and some variability in amplification profiles between isolates. Each PCR reaction was performed in 10 µL final volume using one unit of Invitrogen *Taq* DNA polymerase, a final concentration of 0.5 µM and 5.0 ng of genomic DNA per 10 µL reaction volume. All PCR amplification reactions using

these primers were performed using Eppendorf Mastercycler® gradient model 533 using the PCR reaction of Hantula *et al.* (1996). The DNA was denatured for 10 min at 95° C, followed by 35 cycles of denaturation for 30 sec at 95°C, annealing for 45 sec at the gradient temperature listed in Table 4, and extension for 2 min at 72°C, and a final extension for 7 min at 72°C. Post reaction, samples were held at 4° C until processed. A 5.0 µL volume of each PCR product was mixed with 2 µL of 1:10 diluted loading dye (0.25% w/v bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol in H_2O) and loaded into each well.

Table 2. Degenerate primers designed and used by Hantula *et al.* (1996) and van der Nest *et al.* (2000) and tested in this experiment.

SSR	Number of repeats	Primer (5'-3')*	%GC	T_m (° C)
GT	5	VHV GTG TGT GTG TGN	54.4	49.1
CCA	5	DDC CAC CAC CAC CAC CA	62.7	58.1
ACA	5	BDB ACA ACA ACA ACA ACA	37.0	47.8
CGA	5	DHB CGA CGA CGA CGA CGA	62.9	59.1

***The following designations are used for degenerate sites: V (A, C, or G), H (A, C, or T), N (any base), D (A, G, or T), and B (C, G, or T).**

Table 3. Degenerate primers developed and tested during this project. SSRs were chosen based on frequency within existing EST libraries. Degenerate ends were chosen based on the work of Hantula *et al.*, (1996).

SSR	Abbreviated name	Number of repeats	Primer (5'-3')**	%GC	T_m (° C)
CAG	BCAG	5	BDB CAG CAG CAG CAG CAG	64.8	59.5
CAG	DCAG	5	DDC AGC AGC AGC AGC AG	62.7	57.5
AAG	BAAG	5	BDB AAG AAG AAG AAG AAG	37.0	44.0
AAG	DAAG	5	DDA AGA AGA AGA AG AAG	33.3	40.9
AGG	BAGG	5	BDB AGG AGG AGG AGG AGG	64.8	55.7
AGG	DAGG	5	DDA GGA GGA GGA GGA GG	62.7	53.3
AGC	BAGC	5	BDB AGC AGC AGC AGC AGC	64.8	59.6
AGC	DAGC	5	DDA GCA GCA GCA GCA GC	62.7	57.6

**The following designations are used for degenerate sites: D (A, G, or T), B (C, G, or T).

Table 4. Temperatures used to test degenerate primers on a gradient PCR to determine optimum annealing temperature for PCR reaction.

Primer	Gradient(° C)
ACA	49°±10°
CCA	61°±10°
GT	58°±5°
CGA	57°±7°
BCAG	64°±5°
DAGG	64°±5°
DAAG	50°±10°
DAGC	64°±5°

Isolate comparison using gel electrophoresis

Isolates for final genotypic analysis (87 total) were selected to represent a wide variety of hatchery locations, types of samples (*i.e.* water, fish, egg and swab), and multiple samples from the same location over time to monitor changes that may occur in the same hatchery. Amplified DNA was electrophoresed on a 50 lane gel of 2.5% w/v agarose gel, with 100 bp DNA ladder (New England BioLabs, Inc., Ipswich, MA), at 150 volts for 1 hour 36 min and visualized by staining with GelRed (3X staining solution from 10,000 X stock, w/v) for 60 minutes, followed by illumination under UV light and electronic image capture using Bio-Rad GelDoc XR+ with Image Lab Software. Each primer and isolate combination was repeated at least three times with the same DNA extracted from the original isolate growth to ensure reproducibility and consistency in the final results. Unfortunately due to the large volume of sample collected and processed,

there were not the resources (time, man power, money, and culture material) to do a unique DNA extraction for each replicate.

Genotyping of isolates

Each gel image was scored independently. Minor image adjustments and lane selections were completed using Image Lab Software to optimize band visualization. Bands were manually selected and the size of the fragment was calculated using the 100 bp ladder standard and analysis toolbox of Image Lab Software to determine the molecular weight (bp) and absolute quantity (ng) of DNA in each amplicon. Analysis table data from Image Lab Software was exported to an Excel spreadsheet (Microsoft, 2007) and presence or absence of bands at a given locus for a given isolate was recorded. All bands were manually verified to ensure appropriate scoring of presence (1) or absence (0). For the complied results of all isolates genotyped, replicate amplification profiles of each unique isolate and primer combination were compiled and aligned to create a single amplification profile that represented all the bands amplified in at least one of the replicates. Bands from each replicate were aligned based on calculated base pair size, relative position to other bands, overall band pattern and band intensity. Band patterns were lined up across gels relying mostly on the distinct and consistent bands amplified for all or most isolates. The calculated base pair size was also used to line up amplified products across gels. The average base pair sizes and alignment from the combination of replicates and aligned gels were used to score the presence or absence of a given character for each isolate.

For the analysis of each group of isolates, the best representative gel image or two for a given primer and set of isolates was scored for the presence and absence of a given

character. For population genetic analysis, all of the character trait scores for a given isolate and primer combination were ordered (in primer alphabetical order) into a single string of characters.

Population genetics analysis

Phylogenetic trees were derived using distance matrix data from POPGENE Version 1.32 (32-bit) (Yeh *et al.*, 1997), which also calculated gene frequency, allele number, effective allele number, polymorphic loci, genetic distance, gene diversity, Shannon Diversity Index, Homogeneity test, F-statistics, gene flow, neutrality test and a dendrogram using Nei's genetic distance. Characters were assumed to be dominant markers from a diploid data set in Hardy-Weinberg equilibrium, and the hierarchical structure was set to multiple populations. Maximum parsimony trees were created using pars of PHYLIP Version 3.695 (Felsenstein, 1989). Parsimonious trees were made using default settings except for: saving 1000 trees and randomizing input order (jumble=10). Consense of PHYLIP was used to make a consensus tree of the parsimonious trees. Bootstrap consensus trees were calculated using seqboot, pars and consense of PHYLIP Version 3.695. Seqboot was used to produce 1000 replicates. Parsimonious trees of bootstrapped data were created using pars on the same settings as above, except data sets=1000. A majority consensus tree of the bootstrapped data was created using consense of PHYLIP. MEGA (Build#:4028) (Tamura *et al.*, 2007) and FigTree (version 1.4.0) (Rambaut, 2012) were used to display and label the trees produced by POPGENE and PHYLIP.

Results

Isolate library composition

Since November 2009, samples were collected from a number of fish hatcheries and water sources on Vancouver Island (Table 1, 22 and Figure 4, 23) to test for the presence of *Saprolegnia* species. Because at the outset the source of infection was completely unknown, all possible sources of *Saprolegnia* inoculum, including the source well water of a facility, pre- and post- filtering systems, and even the fish feed, were assayed for the presence of members of the Saprolegniales at each hatchery site. For each water sample collected, four plates were baited to try and ensure that if *Saprolegnia* species were present, they were detected. The large majority of water sample tested negative for *Saprolegnia* species; however, the isolates used in this analysis are representative of *Saprolegnia parasitica*: isolated. Isolates collected from fish and the water column within the hatcheries predominantly tested as *S. parasitica*, fewer tested positive from eggs and the physical tank apparatuses. Very few water samples from outside of the hatcheries tested positive for *Saprolegnia parasitica* growth. Based on the lack of culturing of *Saprolegnia* species from most of these tested sources, we determined the most likely source for the introduction of *Saprolegnia* into hatcheries was through the introduction of previously infected fish or eggs. The majority of the samples used in the following phylogenetic analysis were collected from fish and water samples within the facilities, plus a few collected from eggs or swabs taken within the facilities, and a few water collections from water sources that fed the hatchery facilities. Between November 2009 and November 2012, 580 samples collected from BC hatcheries (Table 1, 31 and Figure 4, 23) or water sources surrounding hatchery facilities, showed signs of

Saprolegnia growth. Of these, 350 were confirmed as *S. parasitica* using the sequences compared to a blastn search and the Identification Engine under the category of "Fungal identification—ITS search" of the Barcode of Life Data System v2.5 (BOLD, web resouce). After February 2011 and the development of the techniques by Leung (2012), samples were positively identified using *puf* primers.

Of the 350 confirmed *S. parasitica* samples, 8 were collected from eggs (2%), 177 from fish (51%), 12 from swabs (3%), 152 from water samples (43%), and 1 from the bio-filter system of a hatchery. The majority of the samples were collected at different locations and times within Sayward Hatchery North to obtain a thorough survey of isolates over time.

Primer development and selection

DNA was extracted from all cultures of confirmed *S. parasitica* isolates to compare the genotype of isolates. Extracted DNA concentrations ranged from 5-2000 ng/ µL with a purity ranging from 1.8-2.16 A_{260}/ A_{280}; however, most samples had a DNA concentration of 200-500 ng/ µL and a purity of 1.9-2.1 A_{260}/ A_{280}. Samples with low concentrations and purity were reserved for final genotype comparison, and samples with higher concentrations and purity were used during trials to determine optimum primers and methods for genotype comparison.

In order to determine which primers might be most suitable for detecting DNA polymorphisms and distinguishing intraspecific variation, a series of screening experiments was conducted to determine the resolving power of primers of interest and to determine which primers might be most informative for detecting genetic variation among isolates. PCR and gel electrophoreses was used to distinguish informative

primers, those considered as showing distinct amplicons with some variation in the number or size of amplicons between different isolates. Four reference isolates (24, 42, 131, and 306) (Table 5, 38) from four different locations and four different collection times were used to maximize the likelihood that if there was a difference in genotypes it would be observable and quantifiable in the amplification profile.

Primers GT, CCA and ACA were determined to be informative with clear, distinct bands showing variation between isolates (Figure 5, 40). Primer CGC was not used in further study due the amplification products appearing as smears rather than discrete bands on the gel (Figure 5, 40). Primers BCAG, DAGG, DAGC and DAAG were also informative with clear, distinct bands showing variation between isolates (Figure 6, 41). Having confirmed that seven of these primers could resolve this small subset of isolates, these seven primers were used for a wider genotypic analysis.

Table 5. Group 1 isolates used for genotypic analysis. Condensed label for trees indicates (from left to right): group a, b, or x (repeated in both groups of isolates), unique isolate id number (3 digits), date of collection (year/month/day), location of collection (Table 1, 22), hatchery tank information and sample substrate type (e=eggs, w=water, f=fish, s=swab)

Isolate ID	Date collected	Location collected	Tank information	Sample substrate	Condensed label for trees
9	22-Dec-09	Nanaimo River		Egg	a009_091222_NR_____e
24	18-May-10	Puntledge River		Egg	a024_100518_PL_____e
30	13-May-10	Puntledge River		Egg	a030_100513_PL_____e
34	Nov-09	Upper Goldstream		Water	a034_0911___UG_____w
42	22-Dec-09	Nanaimo River	Trough	Water	a042_091222_NRTroughw
84	31-Aug-10	Sayward North	10M-01	Water	a084_100831_SN10M-01w
87	07-Sep-10	Sayward North	10M-01	Fish	a087_100907_SN10M-01f
88	07-Sep-10	Sayward North	06M-01	Fish	a088_100907_SN06M-01f
93	13-Sep-10	Sayward North	10M-01	Water	a093_100913_SN10M-01w
98	07-Sep-10	Sayward North	10M-01	Fish	a098_100907_SN10M-01f
104	20-Sep-10	Sayward North	06M-05	Fish	x104_100920_SN06M-05f
105	20-Sep-10	Sayward North	06M-05	Fish	a105_100920_SN06M-05f
107	28-Sep-10	Sayward North	06M-05	Water	a107_100920_SN06M-05w
110	28-Sep-10	Sayward North	06M-05	Fish	a110_100928_SN06M-05f
119	28-Sep-10	Sayward North	10M-04	Water	a119_100928_SN10M-04w
121	28-Sep-10	Sayward North	06M-05	Water	a121_100928_SN06M-05w
123	28-Sep-10	Sayward North	06M-05	Fish	a123_100928_SN06M-05f
125	29-Sep-10	Georgie Lake		Water	a125_100929_GL_____w
131	04-Oct-10	Sayward North	06M-05	Water	a131_101004_SN06M-05w
134	04-Oct-10	Sayward North	06M-05	Fish	a134_101004_SN06M-05f
147	28-Sep-10	Sayward North	06M-05	Water	x147_100928_SN09M-01w

161	28-Sep-10	Sayward North	10M-04	Fish	a161_100928_SN10M-04f
168	06-Oct-10	Sayward North	Tank 02	Fish	a168_101006_SHTank02f
169	12-Oct-10	Sayward North	10M-05	Fish	a169_101012_SN10M-05f
179	19-Oct-10	Sayward North	06M-05	Fish	x179_101012_SN06M-05f
183	19-Oct-10	Sayward North	Seducer	Water	a183_101019_SNSeducrw
200	24-Nov-10	Sayward North	Pre-treatment	Water	a200_101124_SNPretrtw
216	07-Dec-10	Sayward North	10M-01	Fish	a216_101207_SN10M-01f
218	07-Dec-10	Sayward North	10M-01	Water	a218_101207_SN10M-01w
221	07-Dec-10	Sayward North	09M-03	Fish	x221_101207_SN09M-03f
231	07-Dec-10	Sayward North	U-03	Fish	x231_101207_SNU-03__f
256	13-Dec-10	Sayward North	U-Sump	Water	a256_101213_SNU-Sumpw
279	11-Jan-11	Sayward North	06M-09	Fish	a279_110111_SN06M-09f
293	11-Jan-11	Sayward North	10M-01	Water	a293_110111_SN10M-01w
306	19-Jan-11	Sayward South	10M-01	Fish	a306_110119_SS10M-01f
307	19-Jan-11	Sayward North	10M-01	Water	a307_110119_SN10M-01w
310	19-Jan-11	Sayward North	U-01	Water	a310_110119_SNU-01__w
348	08-Feb-11	Sayward North	U-02	Water	a348_110208_SNU-02__w
352	15-Feb-11	Sayward North	09M-03	Fish	a352_110215_SN09M-03f
365	08-Feb-11	Sayward North	09M-01	Fish	a365_110208_SN09M-01f
402	14-Mar-11	Sayward North	09M-01	Water	a402_110314_SN09M-01w
403	14-Mar-11	Sayward North	09M-01	Water	a403_110314_SN09M-01w
404	14-Mar-11	Sayward North	09M-01	Water	a404_110314_SN09M-01w
433	28-Mar-11	Sayward North	Post-treatment	Swab	a433_110328_SNPostrts
438	04-Apr-11	Sayward North	U-02	Water	a438_110404_SNU-02__w
492	29-Aug-11	Sayward South	Anesthetic bath	Fish	a492_110828_SSAnes__f

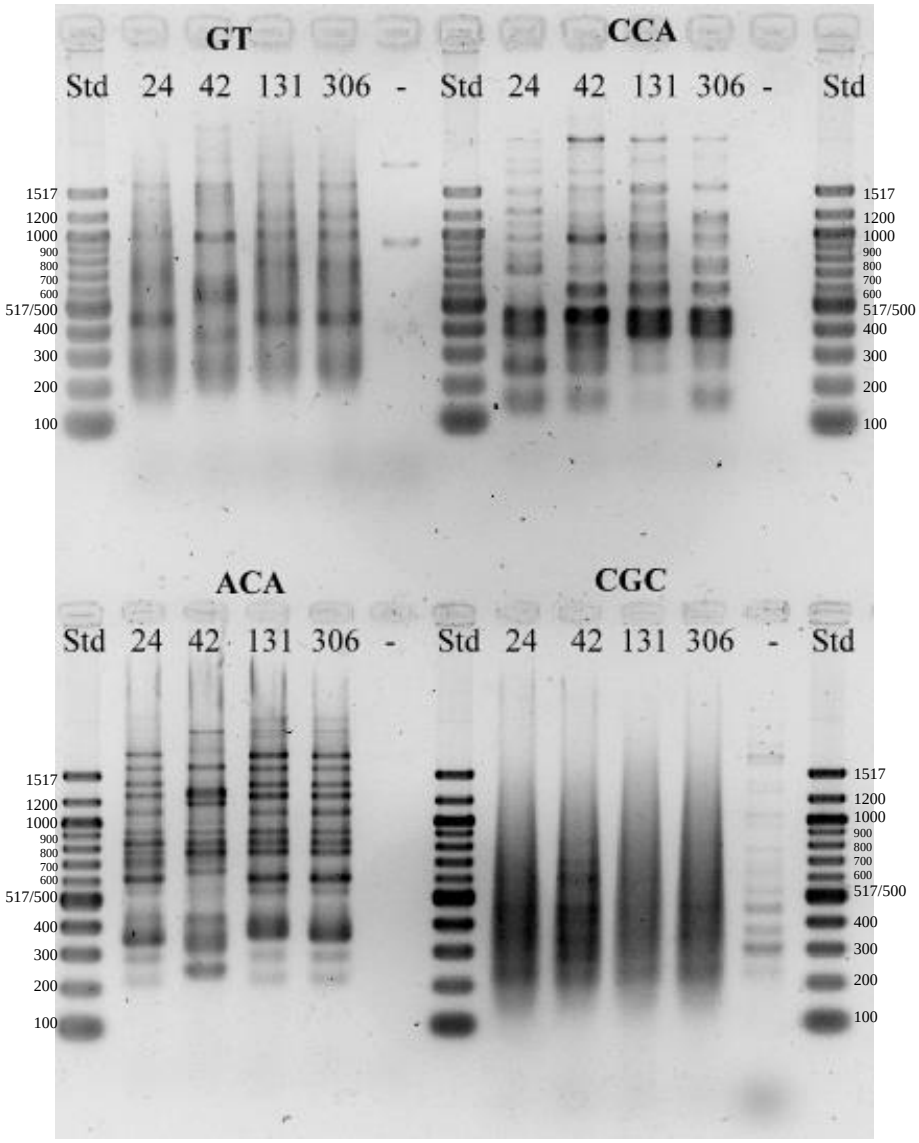


Figure 5. 2% agarose gel of the PCR products of *Saprolegnia parasitica* isolates 24, 42, 131, 306 and negative control (dH_2O) with degenerate primers GT, CCA, ACA, and CGA (Table 2, 31 and Table 3, 31) Each 5 µL PCR sample was mixed with 2 µL 1:10 diluted loading dye. Each ladder was 5 µL of 1:6 100 bp DNA ladder (BioLabs). Run at 97 V for 72 minutes in 1x TAE buffer. Visualized with GelRed and UV light. The amplifications in the negative control lane were likely exogenous DNA contamination due to human error. These PCR amplifications were repeated to ensure contaminant free products before moving on to subsequent steps. The amplification profiles here were only to compare the effectiveness of the primers and not for genotypic comparison.

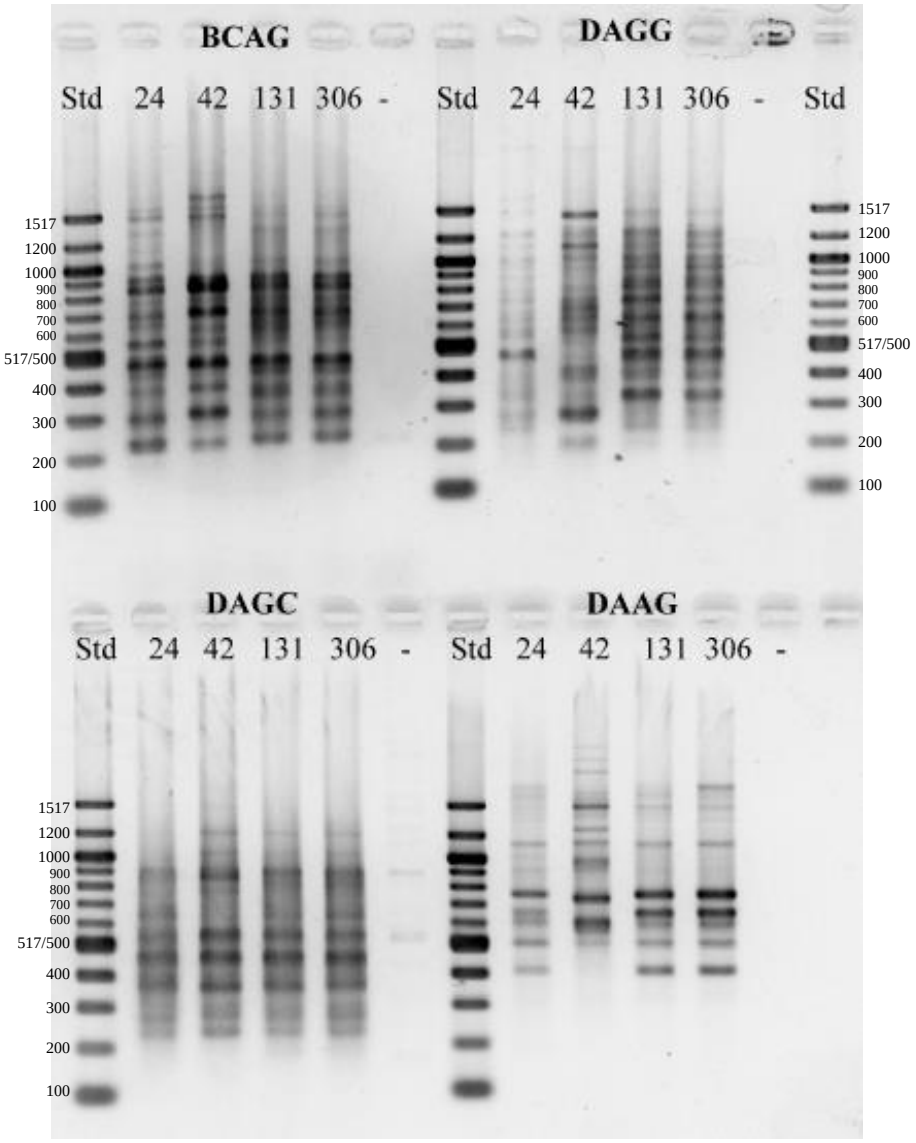


Figure 6. 2% agarose gel of the PCR products of *Saprolegnia parasitica* isolates 24, 42, 131, 306 and negative control (dH_2O) with degenerate primers BCAG, DAGG, DAGC, and DAAG (Table 2, 31 and Table 3, 31). Each 5 µL PCR sample was mixed with 2 µL loading dye. Each ladder is 5 µL of 1:6 100 bp DNA ladder (BioLabs). Run at 97 V for 72 minutes in 1x TAE buffer. Visualized with GelRed and UV light. The amplifications in the negative control lane were likely exogenous DNA contamination due to human error. These PCR amplifications were repeated to ensure contaminant free products before moving on to subsequent steps. The amplification profiles here were only to compare the effectiveness of the primers and not for genotypic comparison.

Genotyping of isolates

The differences in banding patterns and variation in "amplification profiles" among isolates resulted from each primer binding to multiple sites along a DNA strand and replicating multiple regions of the DNA (**Figure** 7, 43). Because there was only one non-specific primer per PCR reaction acting in both forward and reverse directions, each primer targeted and amplified regions of DNA between two inverted SSR repeats. The region of DNA amplified could be a continuation of the repeat, or other DNA base pair sequences in between the repeats. Although **Figure** 7 shows only one amplification region, in actuality across the genome multiple primer binding sites and regions of replication exist, creating a mixture of lengths of DNA replicated in each PCR reaction. The mixture of amplicons was separated on an agarose gel to determine the unique "amplification profile" for a given isolate and primer combination.

In this experiment, each amplicon may have resulted from differences in the genetic sequence of isolates at primer binding sites or differences in the length of the regions between primer binding sites; therefore, each band on a gel is referred to as an "amplicon" and each row as a "character" for which a given isolate was present or absent. From the 350 collected samples of *S. parasitica*, an initial group of 46 isolates was selected for diversity assessment (Table 5, 38). Certain isolates (9, 24, 30, 34, and 42 in Table 5) were intentionally included to ensure coverage of all locations and types of samples, but because of the large number of isolates collected from Sayward Hatchery North, most isolates were randomly selected using a random number generator (http://www.random.org/). The same selection process was completed for the

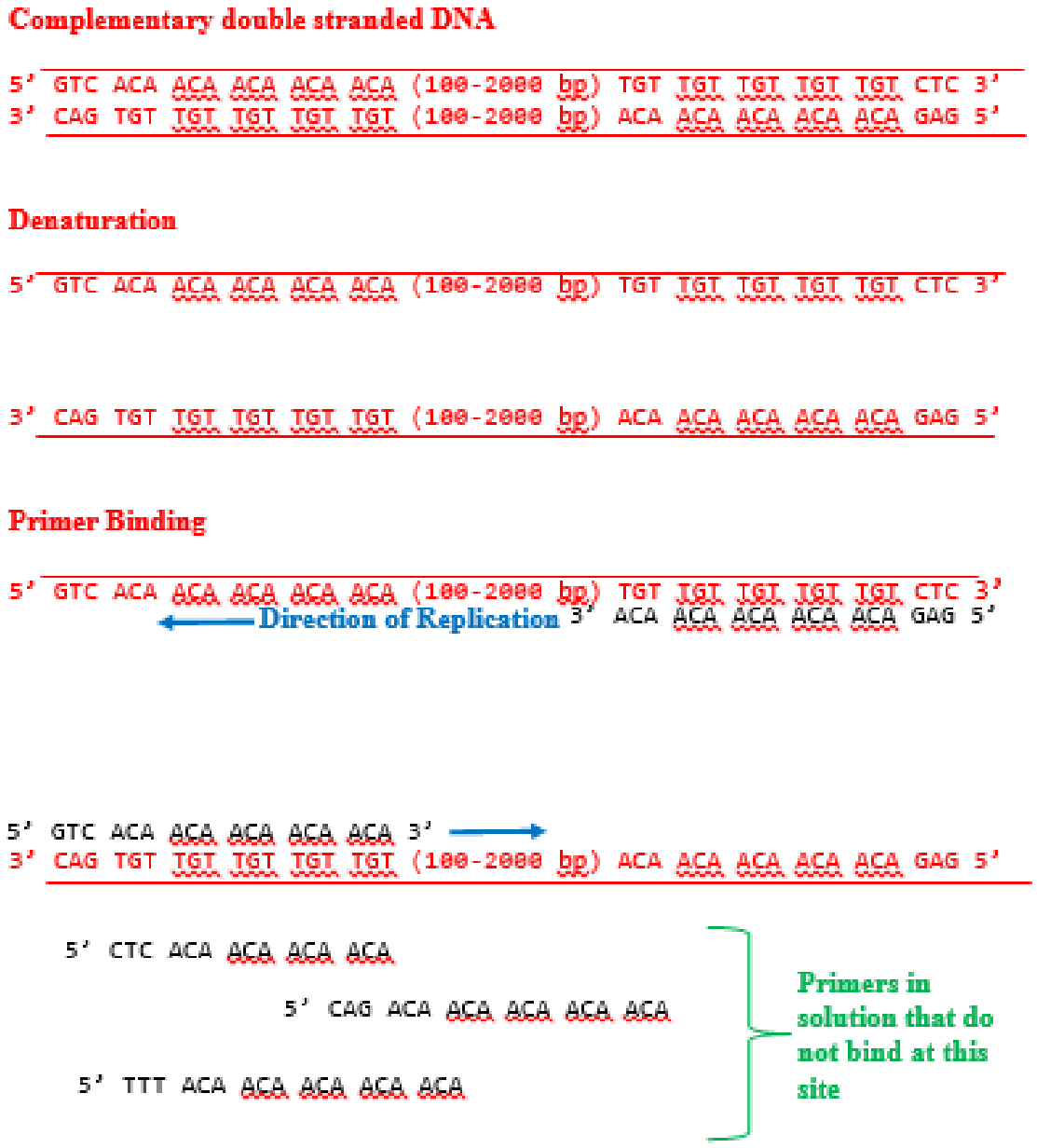


Figure 7. Example diagram of primer ACA (Table 2, 31) bound to a section of DNA and replicating the segment of DNA between the two binding sites. In actuality the primer bound to multiple sites along the DNA strand and amplified multiple regions of DNA between the inverted forward and reverse bound site.

second group of isolates, intentionally including isolates 25, 41, 66, and 72 (Table 7, 46). This resulted in two experimental replicate sets of isolates, with five isolates (104, 147, 179, 221, 231) present in each group, that could be combined into a single data set or independently assayed. The summary of where the isolates were collected from and the source of the sample collection are summarized in Table 6 (45). These two groups constituted the two data sets whose amplification products were separated on large (50 lane) gels which were scored for the presence or absence for any given character (Figure 8, 48 and Figure 9, 49).

Initially, each group of isolates was scored independently taking the best replicate or two of a given primer and isolate combination. The number of characters or amplicon rows amplified by a given primer ranged from 23 to 44, with fragment sizes from 145 to 3000 base pairs.

A second analysis was completed by first aligning the calculated base pair sizes of three or four replicates of a given isolate and primer combination (Figure 10, 51), and then assaying all isolates as a large group by lining up distinct bands that carried across most isolates. The summarized results of all amplicons scored for each primer can be seen in Appendix 1 (Table 9, 93-Table 15, 103). The combined analysis was to compare all isolates and to determine if a smaller group of isolates (46) could be used to draw the same conclusions as a larger group of isolates (87) in the future.

Table 6. Breakdown of total number of isolates collected at each location and the sample substrates for the isolates used for genotypic analysis in this experiment. This breakdown should not be considered to be representative of the entire library of *Saprolegnia parasitica* samples collected during the period of sample collection.

Location collected	**Total number of samples used for genotyping**	**Sample substrate**			
		water	eggs	fish	swabs
Nanaimo River	3	2	1	0	0
Puntledge River	3	0	3	0	0
Sayward Hatchery North	67	33	0	30	4
Sayward Hatchery South	4	0	0	4	0
United Hatchery	2	0	0	2	0
Ocean Falls	2	0	0	2	0
Georgie Lake	3	1	0	2	0
Stelling Hatchery	2	1	0	1	0
Upper Goldstream River	1	1	0	0	0
Total		**38**	**4**	**41**	**4**

Table 7. Group 2 isolates used for genotypic analysis. Condensed label for trees indicates (from left to right): group (a, b, or x (repeated in both groups of isolates)), unique isolate id number (3 digits), date of collection (year/month/day), location of collection (Table 1, 22), hatchery tank information and sample substrate type (e=eggs, w=water, f=fish, s=swab)

Isolate ID	Date collected	Location collected	Tank information	Sample substrate	Condensed label for trees
25	18-May-10	Puntledge River		Egg	b025_100518_PL_____e
41	22-Dec-09	Nanaimo River		Water	b041_091222_NR_____w
63	20-Aug-10	Ocean Falls		Fish	b063_100820_OF_____f
66	20-Aug-10	Ocean Falls		Fish	b066_100820_OF_____f
72	27-Aug-10	United Hatchery		Fish	b072_100827_UH_____f
81	31-Aug-10	Sayward North	10M-01	Fish	b081_100831_SN10M-01f
96	27-Aug-10	United Hatchery		Fish	b096_100827_UH_____f
104	20-Sep-10	Sayward North	06M-05	Fish	x104_100920_SN06M-05f
112	29-Sep-10	Georgie Lake	Penn3	Fish	b112_100929_GLPenn3_f
114	29-Sep-10	Georgie Lake	Penn3	Fish	b114_100929_GLPenn3_f
147	28-Sep-10	Sayward North	09M-01	Water	x147_100928_SN09M-01w
164	20-Sep-10	Sayward North	10M-04	Fish	b164_100920_SN10M-04f
171	27-Oct-10	Sayward North	09M-01	Fish	b171_101027_SN09M-01f
179	12-Oct-10	Sayward North	06M-05	Fish	x179_101012_SN06M-05f
189	27-Oct-10	Stelling Hatchery	Tank K7	Water	b189_101027_SHK7____w
193	12-Oct-10	Sayward North	09M-01	Water	b193_101012_SN09M-01w
211	24-Nov-10	Sayward North	10M-04	Swab	b211_101124_SN10M-04s
213	01-Dec-10	Sayward North	Hatch Tank 24	Water	b213_101201_SNHtch24w
219	07-Dec-10	Sayward North	10M-01	Water	b219_101207_SN10M-01w
221	07-Dec-10	Sayward North	09M-03	Fish	x221_101207_SN09M-03f
231	07-Dec-10	Sayward North	U-03	Fish	x231_101207_SNU-03__f
232	07-Dec-10	Sayward North	U-03	Water	b232_101207_SNU-03__w

240	07-Dec-10	Sayward North	06M-11	Water	b240_101207_SN06M-11w
244	07-Dec-10	Sayward North	U-03	Water	b244_101207_SNU-03__w
253	07-Dec-10	Sayward North	09M-03	Water	b253_101207_SN09M-03w
264	20-Dec-10	Sayward North	Vaccination Tank	Swab	b264_101220_SNVacc__s
303	19-Jan-11	Sayward North	09M-02	Water	b303_110119_SN09M-02w
312	19-Jan-11	Sayward North	09M-02	Fish	b312_110119_SN09M-02f
314	01-Feb-11	Sayward North	09M-01	Fish	b314_110201_SN09M-01f
320	01-Feb-11	Sayward North	10M-09	Fish	b320_110201_SN10M-09f
326	01-Feb-11	Sayward North	UR-01	Fish	b326_110201_SNUR-01_f
330	01-Feb-11	Sayward North	UR-01	Fish	b330_110201_SNUR-01_f
333	01-Feb-11	Sayward North	UR-01	Water	b333_110201_SNUR-01_w
339	08-Feb-11	Sayward North	U-02	Water	b339_110208_SNU-02__w
381	21-Feb-11	Sayward North	09M-03	Water	b381_110221_SN9M-03_w
394	21-Feb-11	Sayward North	U-03	Water	b394_110221_SNU-03__w
409	14-Mar-11	Sayward North	U-02	Water	b409_110314_SNU-02__w
411	28-Mar-11	Sayward North	Pre-treatment	Swab	b411_110328_SNPretrts
418	28-Mar-11	Sayward North	Pre-treatment	Water	b418_110328_SNPretrtw
432	14-Mar-11	Sayward North	06M-15	Fish	b432_110314_SN06M-15f
449	04-Apr-11	Sayward North	10M-02	Fish	b449_110404_SN10M-02f
489	10-Aug-11	Sayward North	10M-01	Fish	b489_110810_SN10M-01f
502	29-Aug-11	Sayward South	Tank R31	Fish	b502_110829_SSTnkR31f
507	05-Jul-11	Sayward North	U-03	Fish	b507_110705_SNU-03__f
517	28-Nov-11	Sayward North	U-01	Fish	b517_111128_SNU-01__f
554	27-Feb-12	Sayward South	Tank R10	Fish	b554_120227_SSTnkR10f

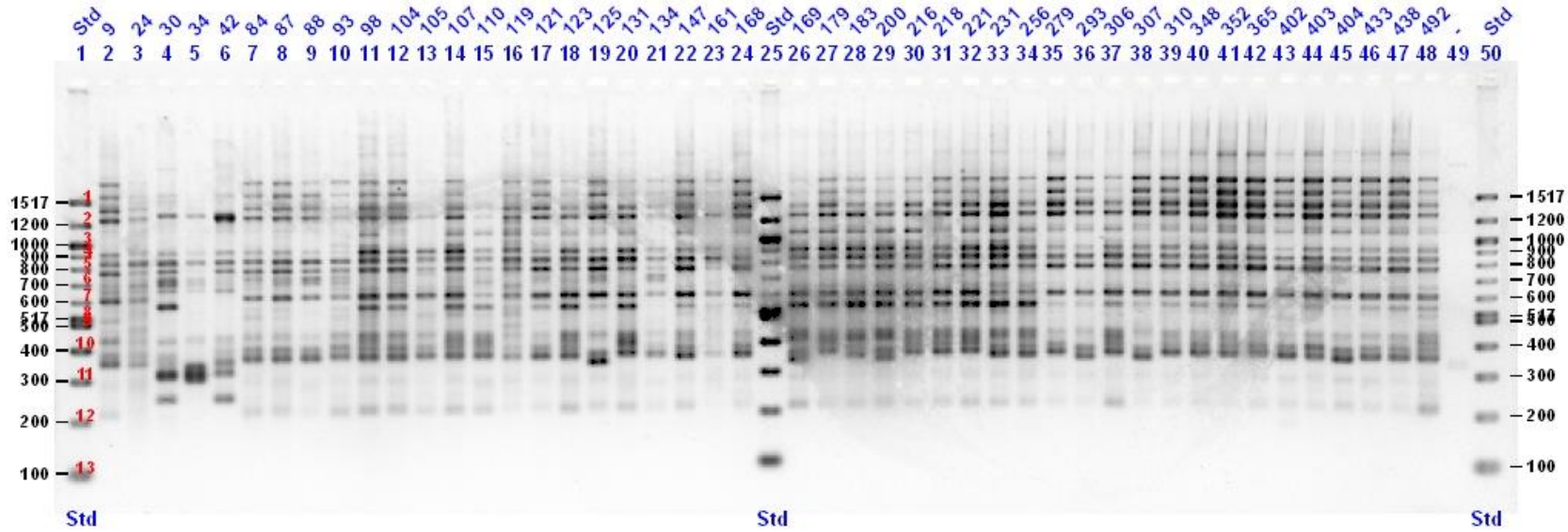


Figure 8. Example of a 50 lane gel (2.5 % agarose) of the PCR products using primer ACA with template DNA of 46 *Saprolegnia parasitica* isolates (group 1). Isolates were collected from sample sites in BC hatcheries between November 2009 and August 2011 including seven hatchery sites. These amplification profiles were scored for presence and absence of characters to compare genetic profiles of the isolates. Lane 1, 25 and Lane 50 were standard 100 bp DNA ladders. Gels were run at 150 volts for 1 hour 36 min and visualized by staining with GelRed and illuminated under UV light using a GelDoc+ with Image Lab software.

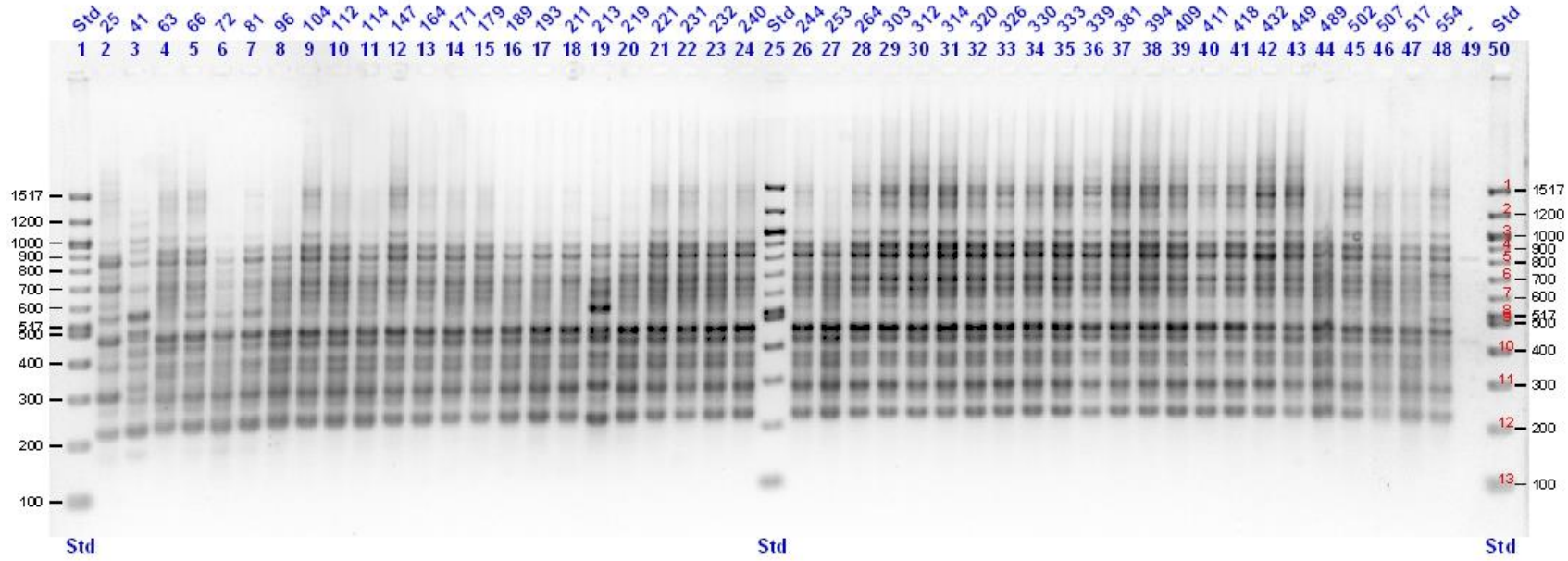
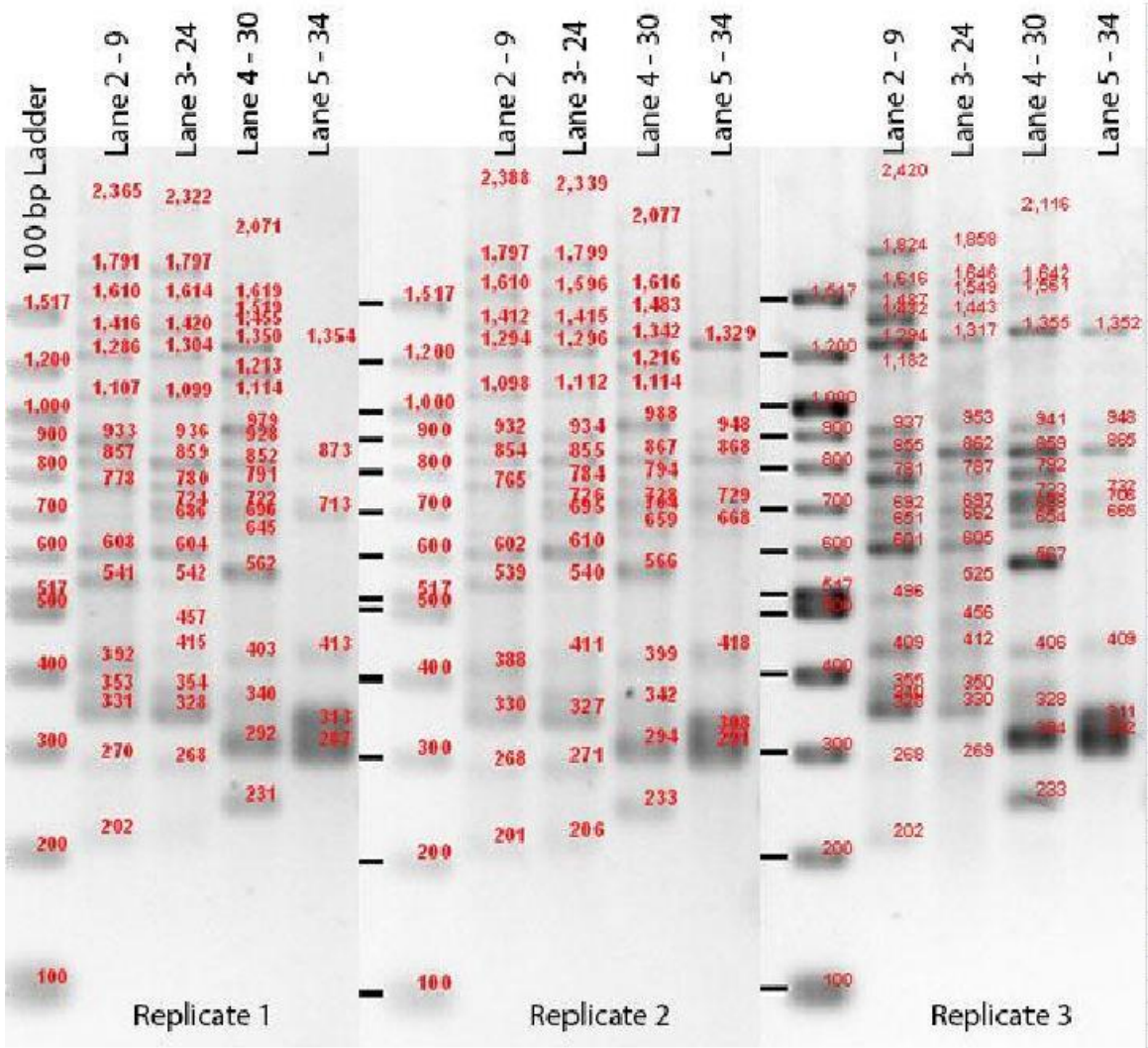


Figure 9. Example of a 50 lane gel (2.5 % agarose) of the PCR products using primer BCAG with template DNA of 46 *Saprolegnia parasitica* isolates (group 2). Isolates were collected from sample sites in BC hatcheries between November 2009 and August 2011 including seven hatchery sites. These amplification profiles were scored for presence and absence of characters to compare genetic profiles of the isolates. Lane 1, 25 and Lane 50 were standard 100 bp DNA ladders. Gels were run at 150 volts for 1 hour 36 min and visualized by staining with GelRed and illuminated under UV light using a GelDoc+ with Image Lab software.



Average base size of replicates							
Lane 2 - 9		Lane 3 - 24		Lane 4 - 30		Lane 5 - 34	
2390	1	2330	1		0		0
	0		0		0		0
	0		0	2090	1		0
	0		0		0		0
1805	1	1820	1		0		0
1610	1	1620	1	1625	1		0
	0	1550	1	1540	1		0
1485	1		0	1485	1		0
1420	1	1425	1	1455	1		0
	0		0	1350	1	1345	1
	0	1305	1		0		0
1290	1		0		0		0
	0		0	1215	1		0
1180	1	1165	1		0		0
1100	1	1105	1	1115	1		0
	0		0		0		0
	0		0	985	1		0
	0		0		0		0
935	1	940	1	935	1	950	1
	0		0		0	870	1
855	1	860	1	860	1		0
775	1	785	1	790	1		0
	0		0		0		0
	0	725	1	725	1	730	1
690	1	690	1	695	1	710	1
650	1	660	1	655	1	665	1
605	1	605	1		0		0
	0		0	565	1		0
540	1	540	1		0		0
495	1		0		0		0
	0	455	1		0		0
	0		0		0		0
395	1	410	1		0	415	1
	0		0	405	1		0
	0		0		0		0
	0		0		0		0
355	1	350	1		0		0
	0		0	340	1		0
	0		0		0		0
330	1	330	1		0		0
	0		0		0	310	1
	0		0		0		0
	0		0	295	1	290	1
	0		0		0		0
270	1	270	1		0		0
	0		0	230	1		0
200	1	205	1		0		0
	0		0		0		0

Figure 10. (Previous page) Example of gel scoring: primer ACA group 1 replicates. Each panel is one of three separate PCR reactions and gel visualizations creating replicate gel images of primer ACA (Table 2, 31). Replicate 1 PCR was February 18, 2013, Replicate 2 was October 31, 2012 and Replicate 3 was completed October 30, 2012. Each replicate is annotated in red with the base pair size for each amplicon (calculated and reported by ImageLab software). The labels at the top of the gels indicate the 100bp ladder, lane number and isolate identification numbers (9, 24, 30, and 34). The table on the right shows the average base pair size of the amplicons for the three replicates (replicate 1, 2 and 3), round to the nearest five base pairs. Band presence or absence was scored based on the entire 50 lane gel, not just the four lanes indicated in this figure.

Population genetics analysis

As described in the methods, two groups of isolates were each scored independently for the presence or absence of each character amplicon. The two groups of isolates, 46 in each group with five repeated in each group, were also combined into a single combined isolate data set of 87 unique isolates. The groups were scored independently to determine if a smaller group of isolates (46) would show similar trends to a large group of isolates (87) that could not be compared on the same gel electrophoresis. Character scoring data and accompanying population genetics statistics and phylogenetic tree analyses were completed for each group of isolates separately and the combined group of isolates. More in depth analysis of the allele frequency breakdown was only completed for the larger data set to show the overall trend without being overwhelmed with the amount of data. When each group of isolates was scored independently for all seven primers used, a total of 222 characters were scored for group 1 and 243 characters for group 2 (Table 8, 57). When all isolates were scored as a single group, including three replicates of each primer and isolate combination, by aligning the two groups of isolates, a total of 309 characters were scored. The number of characters scored for each primer ranged from 33 to 52 for the combined isolate data.

All of the characters (amplicons for each primer) showed a gradual, rather than an exponential decrease in the allele frequency for all the characters scored (**Figure 11**, 54). This indicates there is a relatively even distribution of the presence or absence for the characters scored, *i.e.* there are not just a few characters that separate out the isolates, compared to the rest of the characters which lump all of the isolates together. All amplicons can therefore be assumed to be informative. When each primer is examined

individually, there are small differences in the distribution of the allele frequency reflected in the slope of the line (**Figure 11**, 54). Primers ACA, DAAG, and DAGG have similar numbers of characters scored and similar slopes; they have very gradual decreases in the number of isolates separated out for each character. In contrast, primers BCAG and DAGC have fewer characters scored, and steeper slopes. Primer BCAG only reaches about a 60-40 split in allele frequency, whereas the other primers approach the 50-50 split in allele frequency (this is indicated by the lowest deviation from 0.5). Primer BCAG has the lowest allele frequencies for the greatest number of characters (plateau along the top of the graph), and then allele frequency increases sharply (deviation decreases). Primer CCA has the most gradual and consistent decrease in the deviation from 0.5 allele frequency, indicating that there is an even distribution of characters that separate out different numbers of isolates.

The frequency distribution for the characters for the combined isolate data indicates that overall there is a slight peak in the number of characters with a deviation of 0.4-0.5 from 0.5 (separates out >80% of the isolates), indicating a few more minor alleles than the average for the rest of the characters (Figure 12, 56). There is also a peak in the number of characters that deviate by about 0.25 from 0.5 (separates out about 75% of the isolates)

As calculated according to Nei (1987), isolates represented by group 1 had 90% polymorphic (both presence and absence scored for that character) loci (n=200), while group 2 had 93% (n=226), and the combined group had 96% polymorphic loci (297) (Table 8, 57). This portion of polymorphic loci is used to measure the genetic variation of a population when studying a large number of loci (Nei, 1987). The observed number of

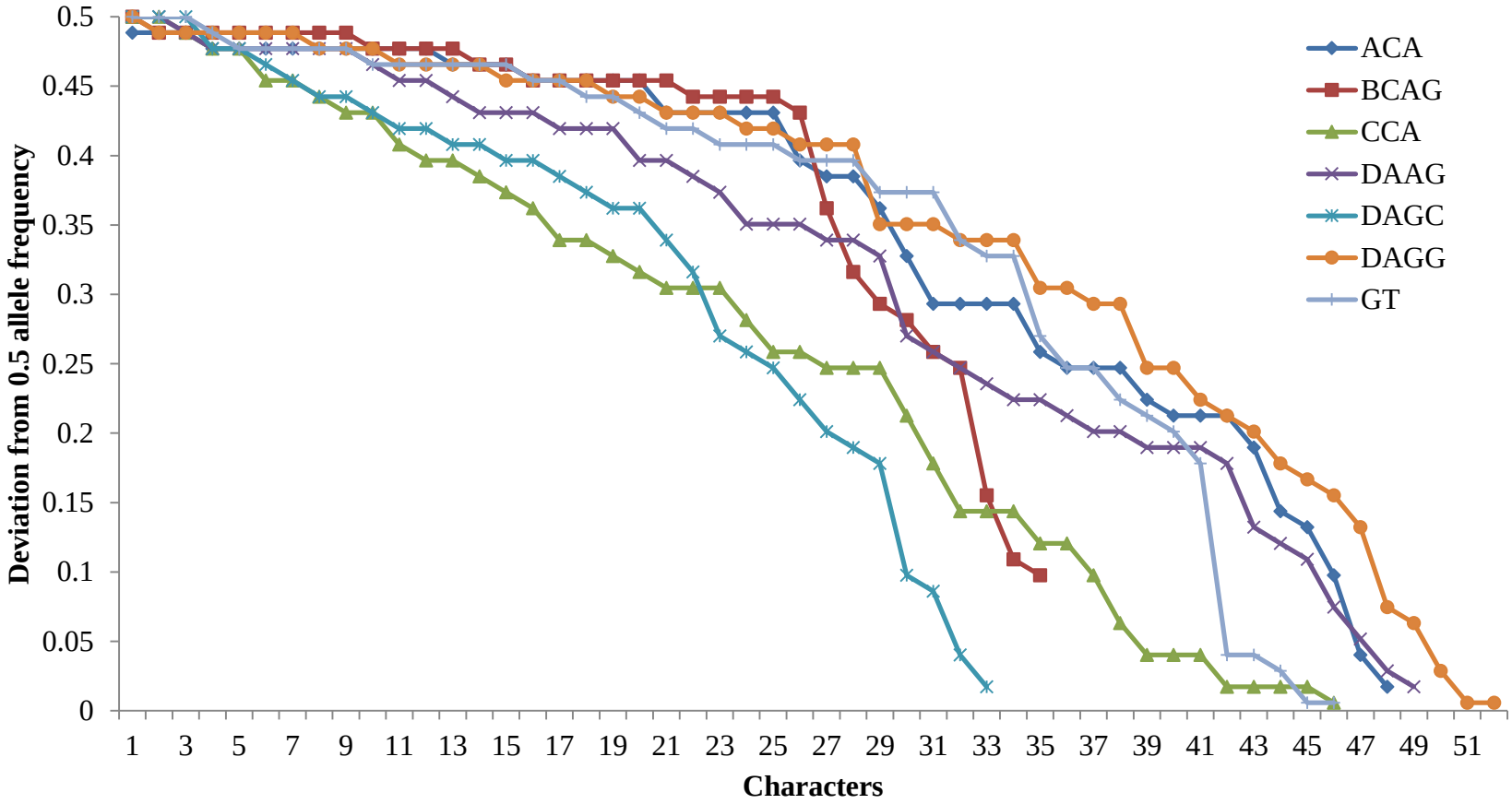


Figure 11. Allele frequency for each of the RAM primers used for genotype comparison. Deviation from an equal (0.5) allele frequency (presence and absence) was calculated for each character, for each primer sequence. Characters along the x axis were ordered by their deviation, not their character string order. Characters that were present in all of the isolates were indicated by the leftmost points; minor characters (high deviation from 0.5) tend to the left, and characters that have roughly equal presence and absence distribution tend to the right of the figure.

alleles (Nei, 1987) for the population in group 1 was 1.90±0.30, 1.93±0.25 for group 2, and 1.96±0.19 when comparing all isolates as a single group, indicating that most characters had many more isolates present for that character than absent. It is calculated by counting the number of alleles present at each locus. In this case a value of one would indicate that only the "presence" allele was observed for that character; whereas, a value of two indicates that both presence and absence of the allele were observed for that character.

The effective number of bands, or reciprocal of homozygosity (Kimura and Crow, 1964, Nei, 1987), is the number of alleles that can be present in a population (De Vicente *et al.*, 2003). In this experiment it has a maximum value of 2.0 because there are only two possible alleles: "presence" or "absence". The average effective number of alleles for all loci (n_e) was 1.32±.33, 1.33±0.31, and 1.36±.33 for group 1, group 2, and all isolates combined, respectively. The effective number of bands or alleles is a measure of true diversity within the population (Jost, 2008).

Nei's genetic diversity (H_e) is the probability that at any given locus any two alleles chosen at random from the population are different from each other and is a measure of the extent of genetic variability in the population (Nei, 1973, De Vicente *et al.*, 2003). For this experiment it was 0.20±0.17 for group 1 isolates, 0.18±0.16 for group 2 isolates, and 0.23±0.17 for all isolates combined. Shannon's information index (Lewontin, 1972) was 0.33±.22 for group 1 isolates, 0.30±0.22 for group 2 isolates, and 0.36±0.22 for all isolates combined. These population genetic statistics, summarized below, showed a very good correspondence between the two groups of data individually and when they were re-scored as a single large data set (Table 8, 57).

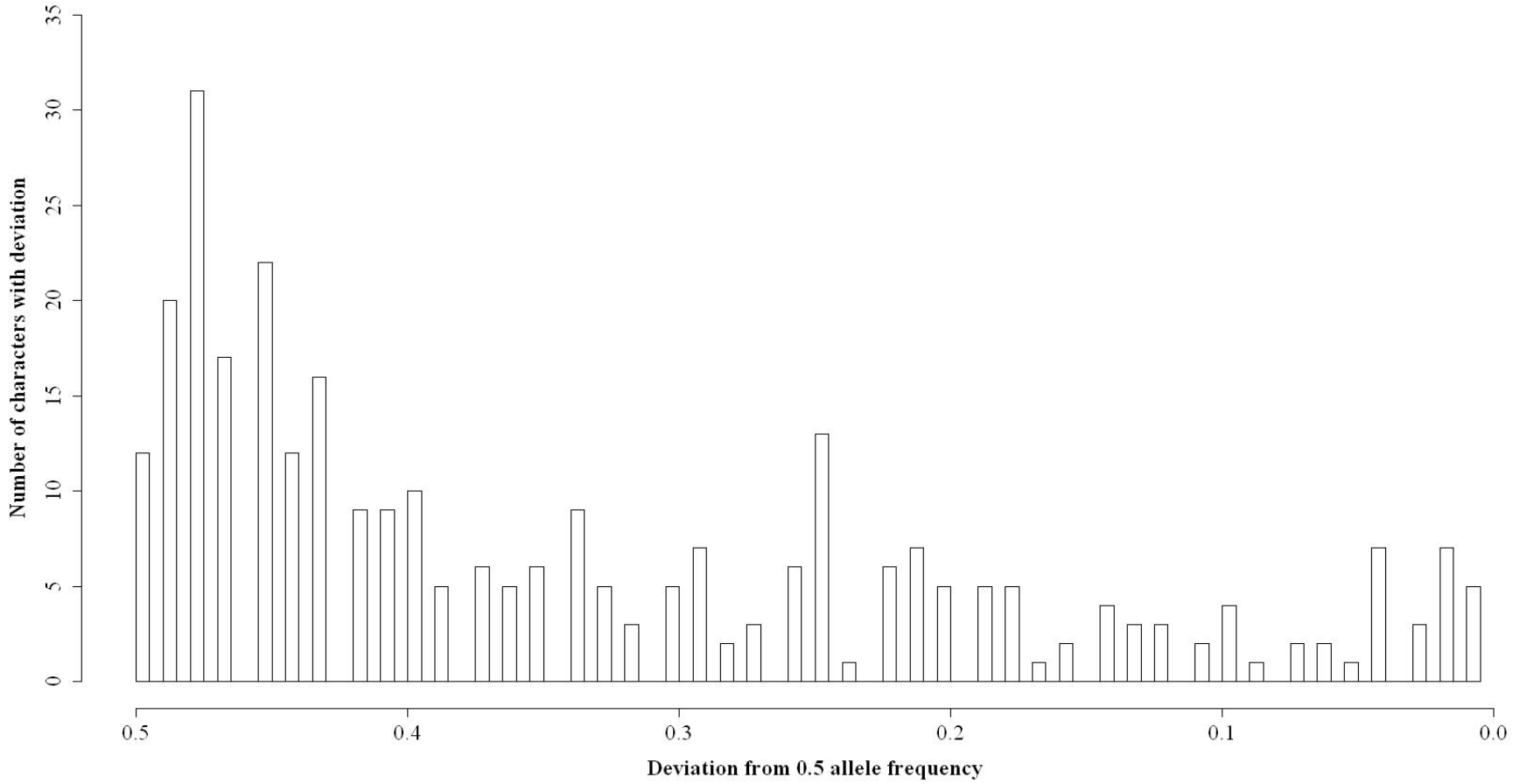


Figure 12. Frequency distribution for combined isolate data of the number of characters that deviate, based on how far they deviate from an allele frequency of 0.5. Characters that are present in all isolates are the leftmost column; minor characters (high deviation from 0.5) tend to the left, and characters that have roughly equal presence and absence distribution tend to the right.

Table 8. Summary statistics for combined primer data of scored amplification profiles of *Saprolegnia parasitica* isolates collected from BC hatcheries. Values represent the average for all loci or characters scored. Group 1 and Group 2 isolates are independent groups of isolates. "All isolates" is groups 1 and 2 aligned and re-scored for presence or absence of each allele. Five isolates were repeated in both groups to aid in alignment.

	Group 1 Isolates	Group 2 Isolates	All Isolates
Number of Isolates	46	46	87
Total number of characters	222	243	309
Number of polymorphic alleles	200	226	297
Percentage of polymorphic alleles	90%	93%	96%
Observed number of alleles (n_a)	1.90±0.30	1.93±0.25	1.96±0.19
Effective number of alleles (n_e)	1.32±.33	1.33±0.31	1.36±.33
Nei's genetic diversity (H_e)	0.20±0.17	0.18±0.16	0.23±0.17
Shannon's Information Index (I)	0.33±.22	0.30±0.22	0.36±0.22

Phylogenetic trees

Phylogenetic trees were constructed for the two groups separately and then as a combined data set, to try to determine the relatedness of collected isolates and to look for patterns in relatedness among all of the collected isolates (*e.g.* populations corresponding to specific locations). The trees also helped determine how diverse or similar the isolates may be. Closely related or isolates with low diversity would be expected to form a large clade with few distinct branches. If isolates collected from different locations or times shared similarities, they would be expected to group into different clades based on those that are most closely related. The data were also bootstrapped to test the robustness of

the data and indicate the percentage of times the given tree is an accurate representation of all possibilities according to the bootstrapped data. The bootstrap values infer the number of conflicting phylogenetic trees that would otherwise not be displayed through parsimonious or distance matrix methods. Consensus trees of the most parsimonious trees of the bootstrapped data were derived. First each group will be discussed individually and then the analysis of both groups as a combined data set will follow. A maximum parsimony tree with bootstrap values and colored by location was assessed for group 1 (Figure 13, 61). Overall the bootstrap values were very low, but those with a value >0.60 are indicated in black and labelled with A-F (Figure 13, 61).

Both samples from Puntledge River clustered together (in clade A) with high significance (0.93), whereas the samples from Nanaimo River were not clustered. The two isolates collected from Sayward South were not clustered together. The isolates from Georgie Lake, Upper Goldstream and Stelling Hatchery had no other isolates from the same location to compare to, but isolate 125 from Georgie Lake paired with isolate 121 from Sayward Hatchery North (clade E, boot strap value = 0.63). There was a group of isolates (131, 107, 98, 105, and 104) from Sayward North (clade C) that formed a clade for all of the majority consensus trees. It should also be noted that four of the five isolates came from the same tank (6M-05) within the facility.

There was no correlation between the source of isolate (*e.g.* fish, water) and clustering, as best indicated by clade C which has three isolates collected from fish and two collected from water samples. Most of the clades (B-E) with significant bootstrap values were collected within a relatively short time frame. Clade A was collected between November 2009 and May 18th, 2010. Clade B was collected between October

12, 2010 and November 24, 2010. Clade C was collected between September 7th and October 10th 2010. Clade D was collected between December 7th and 13th 2010. Clade E was collected on September 28th and 29th 2010.

The Nei's genetic distance tree reconstructed using UPGMA, colored by location, for group 1 (Figure 14, 62) agreed with the maximum parsimony tree for clusters by location and source of isolate. The isolate that clustered together in the maximum parsimony tree with labelled bootstrap values, also paired together in the Nei's genetic distance tree (A-E, Figure 13, 61 and Figure 14, 62), with the exception of isolate 24 (A) which forms a tighter clade in the majority consensus tree, but is paired more closely with isolate 134 in the Nei's genetic distance tree. Isolates 30, 34, 42, 24, and 134 form a clade in both trees.

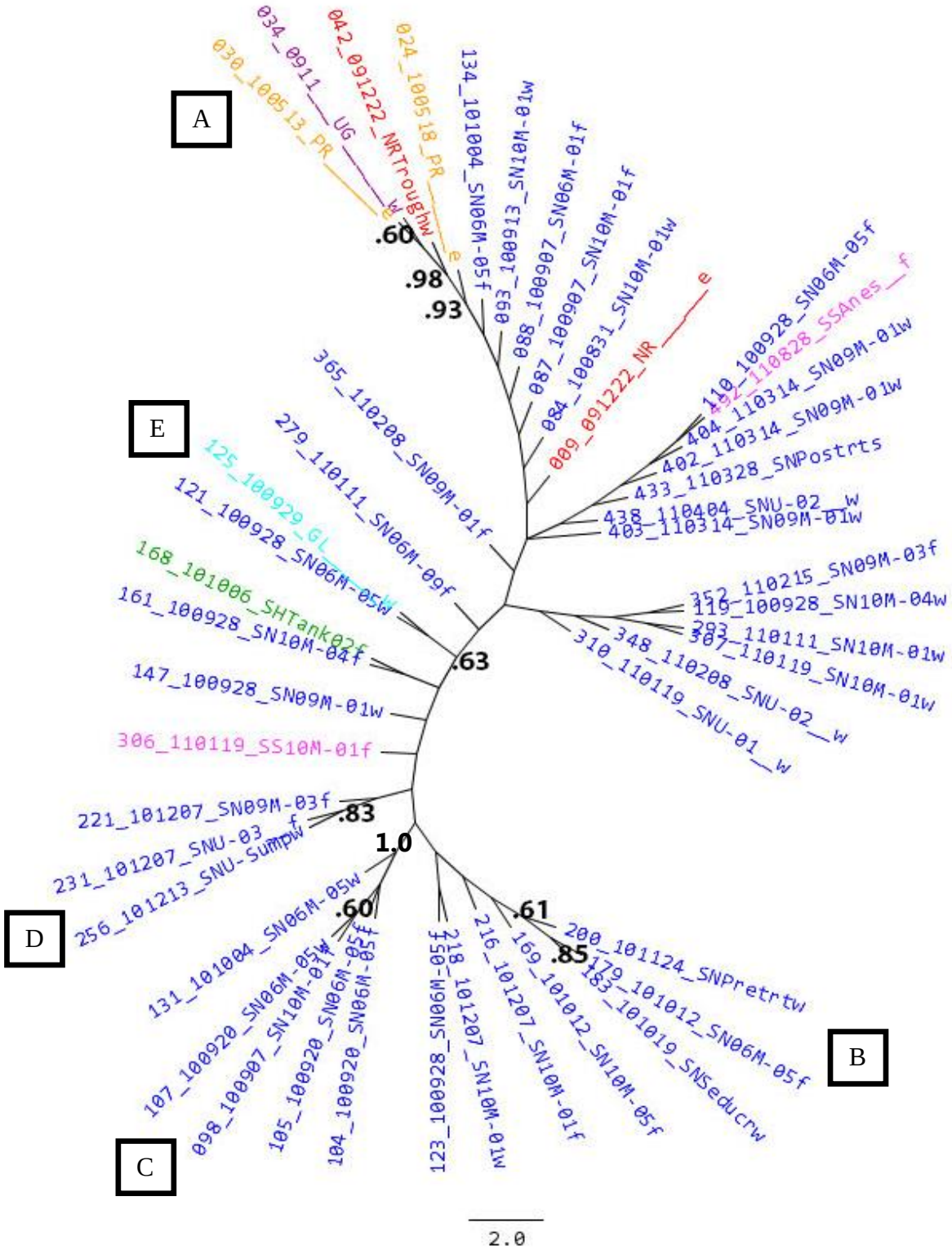

A
E
D
C
B
.60
.98
.93
.63
.83
1.0
.60
.61
.85
034_0911__UG
042_091222_NRTroughw
024_100518_PR
030_100513_PR
134_101004_SN06M-05f
093_100913_SN10M-01w
088_100907_SN06M-01f
087_100907_SN10M-01f
084_100831_SN10M-01w
009_091222_NR
110_100928_SN06M-05f
492_110828_SSAnes
404_110314_SN09M-01w
402_110314_SN09M-01w
433_110328_SNPostrts
438_110404_SNU-02
403_110314_SN09M-01w
365_110208_SN09M-01f
279_110111_SN06M-09f
125_100929_GL
121_100928_SN06M-05w
168_101006_SHTank02f
161_100928_SN10M-04f
147_100928_SN09M-01w
306_110119_SS10M-01f
352_110215_SN09M-03f
119_100928_SN10M-04w
283_110111_SN10M-01w
307_110119_SN10M-01w
348_110208_SNU-02
310_110119_SNU-01
221_101207_SN09M-03f
231_101207_SNU-03
256_101213_SNU-Sumpw
131_101004_SN06M-05w
107_100920_SN06M-05w
098_100907_SN10M-01f
105_100920_SN06M-05f
104_100920_SN06M-05f
123_100928_SN06M-05f
218_101207_SN10M-01w
216_101207_SN10M-01f
169_101012_SN10M-05f
179_101012_SN06M-05f
183_101019_SNSeducrw
200_101124_SNPretrtw
2.0

Figure 13. (Previous page) Majority consensus of maximum parsimony analysis of 1000 replicate bootstrapped data for group 1 isolates. Colored according to sample collection location: blue=Sayward North, red=Nanaimo River, green=Stelling Hatchery, aqua=Georgie Lake, orange=Puntledge River, purple=Upper Goldstream, pink=Sayward South. Numbers in black indicate bootstrap values of >0.60. Red braces and green circle are to compare arrangement to Nei's genetic distance tree (Figure 14, 62)

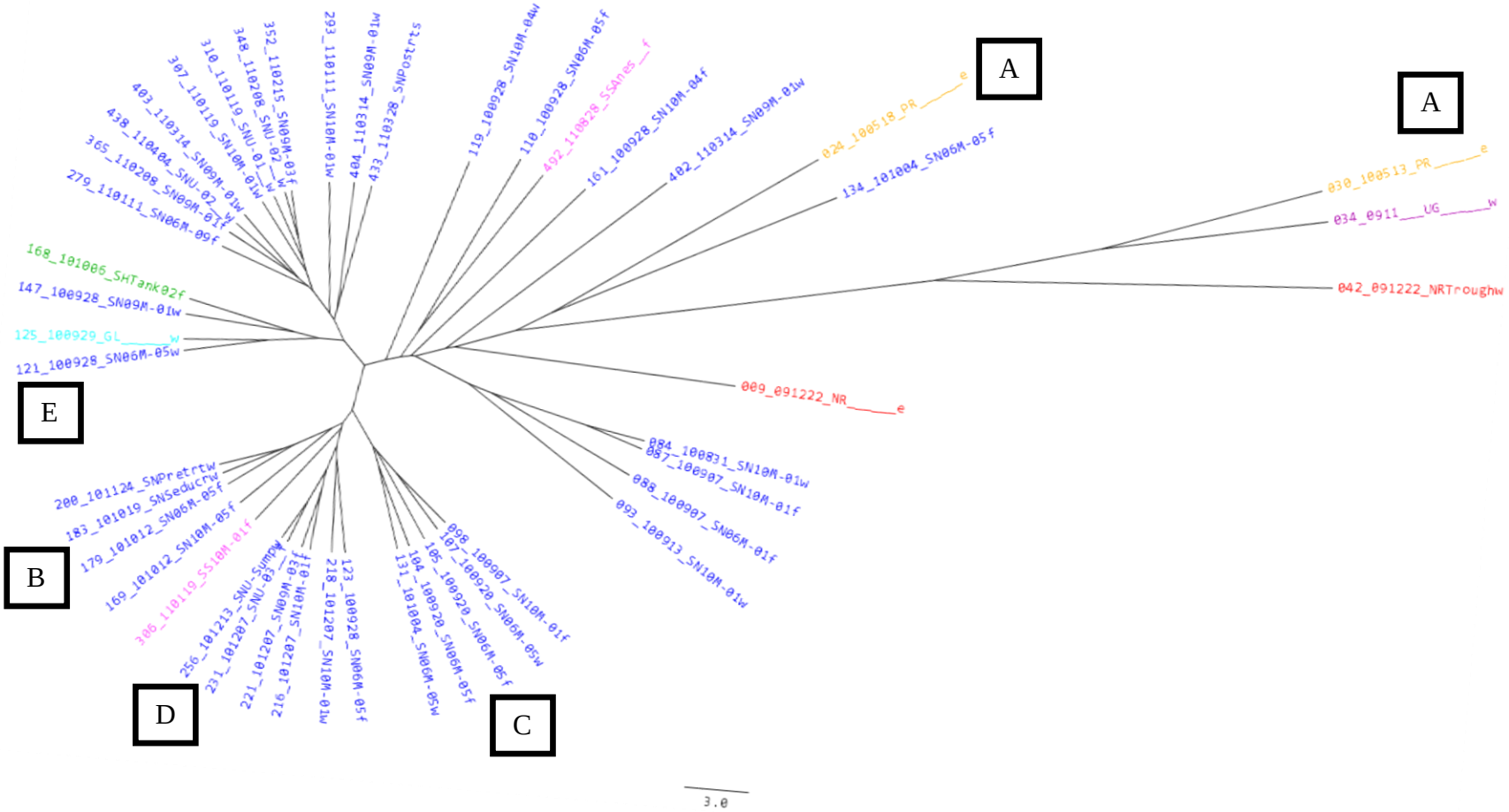


Figure 14. Nei's genetic distance matrix tree of combined primer data for group 1 isolates. Colored according to sample collection location: blue=Sayward North, red=Nanaimo River, green=Stelling Hatchery, aqua=Georgie Lake, orange=Puntledge River, purple=Upper Goldstream, pink=Sayward South.

A maximum parsimony tree with bootstrap values and colored by location for group 2 isolates was assessed (Figure 15, 64). Again, most bootstrap values were quite low, but bootstrap values >0.60 are indicated in black and clades of isolates referred to are identified with unique letters in Figure 15. Isolates from Georgie Lake paired together (0.90, F). Isolates from Sayward South did not pair together. Isolates from United Hatchery did not pair together. Isolate 25 from Puntledge River and 41 from Nanaimo River paired together (0.66, G). Several pairs of isolates collected from Sayward hatchery North paired together: 381 and 394 (0.83, H), 449 and 432 (0.83, I), 411 and 418 (0.78, J), and 221 and 231 (0.96, K). As was observed for group 1, no pattern in isolate clustering for source of sample (*e.g.* water or fish) was observed. Again, isolates grouped closely together were most closely related by date. Pair F was collected September 29th, 2010. Isolates in clade G were collected between December 22, 2009 and August 27th, 2010. Isolates in clade H were collected February 21, 2011. Isolates in clade I were collected April 4th and March 14th 2011. Isolates in clade J were collected March 28th, 2011. Isolates in clade K were collected December 7th, 2010.

The genetic distance tree based on UPGMA, colored by location, for group 2 (Figure 16, 65) agreed with the maximum parsimony tree dividing the isolates into three clades (indicated with corresponding circle and brace on Figure 15, 64 and Figure 16, 65); however, there were differences in the arrangement within those clades. No patterns in isolate source were observed.

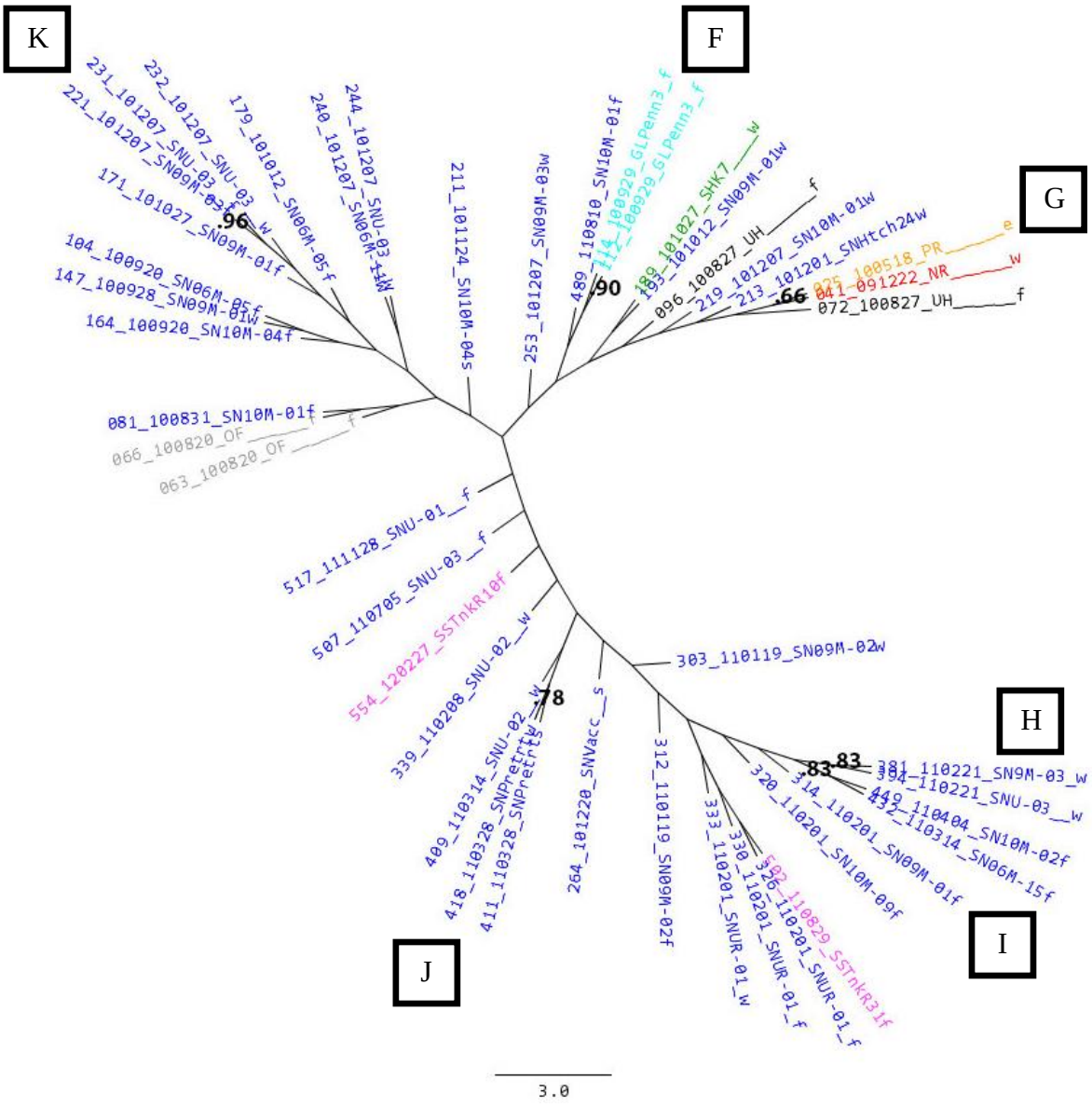


Figure 15. Majority consensus of maximum parsimony analysis of 1000 replicates of bootstrapped data for group 2 isolates. Colored according to sample collection location: blue=Sayward North, red=Nanaimo River, green=Stelling Hatchery, aqua=Georgie Lake, orange=Puntledge River, pink=Sayward South, grey=Ocean Falls and black=United Hatchery. Numbers in black indicate bootstrap values >0.60.

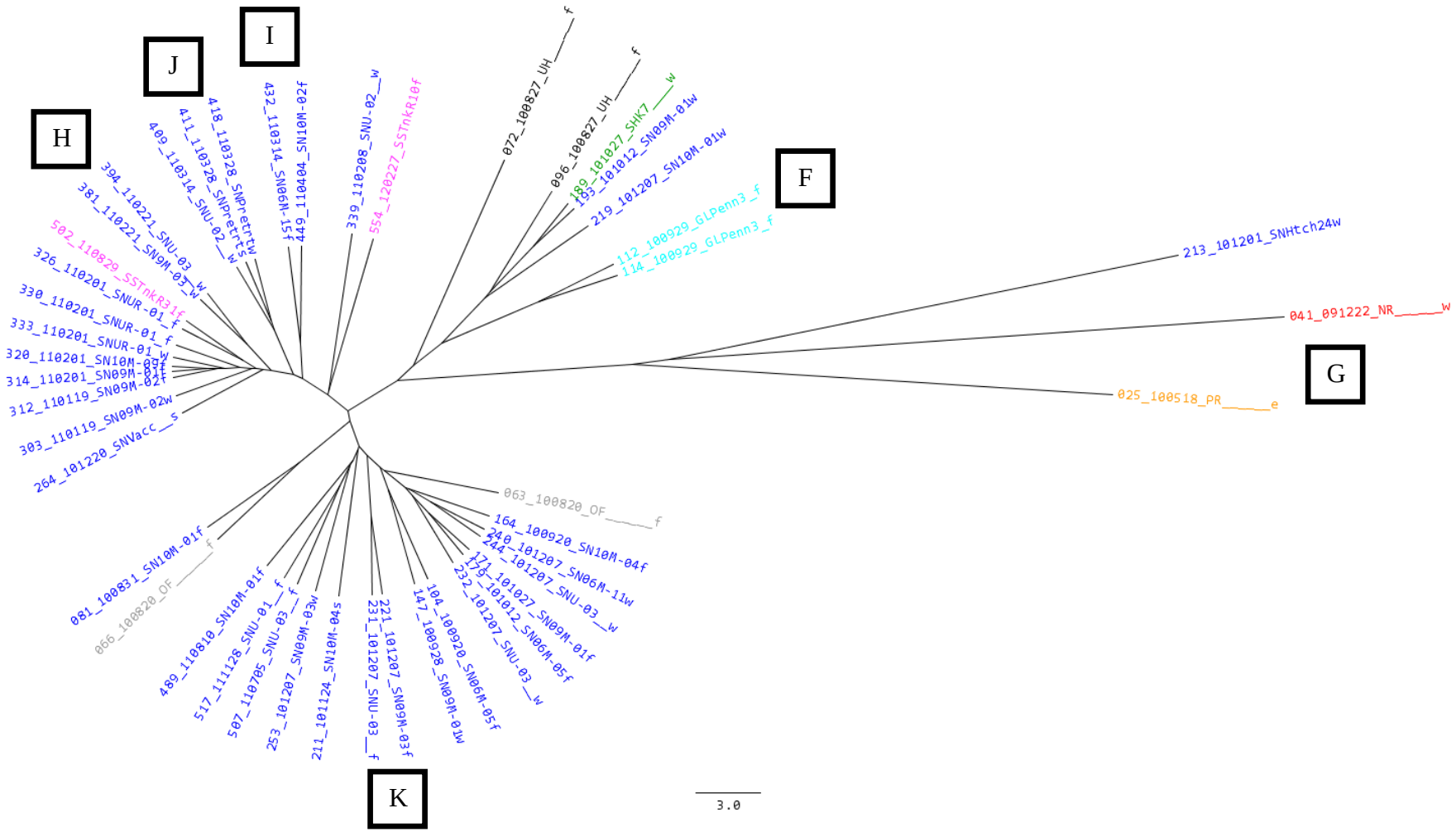


Figure 16. Nei's genetic distance matrix based tree of combined primer data for group 2 isolates. Colored according to sample collection location: blue=Sayward North, red=Nanaimo River, green=Stelling Hatchery, aqua=Georgie Lake, orange=Puntledge River, pink=Sayward South, grey=Ocean Falls and black=United Hatchery.

When the two groups of isolates were combined and re-scored as a single large data set, similar trends were observed in the maximum parsimony with consensus and bootstrap values tree (Figure 17, 67) and the Nei's genetic distance calculated using UPGMA trees (Figure 18, 68), as were observed in the data sets of each group independently; however, there were a couple of interesting differences. Again, most bootstrap values were low, but those >0.60 are indicated in black and identified with a unique clade letter. One of the biggest differences between the group 2 trees and the combined trees was that isolates 72 and 81 collected August 27th, 2010 from United Hatchery and August 31, 2010 from Sayward North grouped together (0.69, O). Also of particular note is that clade A is composed of isolates 30 (Puntledge River), 34 (Upper Goldstream), and 42 (Nanaimo River) from group 1 and isolate 25 (Puntledge Lake) from group 2 (0.83). Isolates 411 and 418 (0.90, M), 432 and 449 (0.69, N), 112 and 114 (0.99, P) and 221 and 231 (Q) continued to cluster together. There was not a correlation observed in sample substrate and genotype of the isolate.

Similar trends were seen between the maximum parsimony and Nei's genetic distance tree (Figure 18, 68), although some re-structuring of the isolates that compose each clade occurred. The clades (L-Q) identified in the maximum parsimony tree (Figure 17, 67), were correspondingly marked in the Nei's genetic distance tree. For example, clade M and N are quite distinct in the maximum parsimony tree; however, they make up a larger clade in the Nei's genetic distance tree. Clades L and O are quite distantly related in the Figure 17 (67), but make up part of a larger clade in Figure 18 (68).

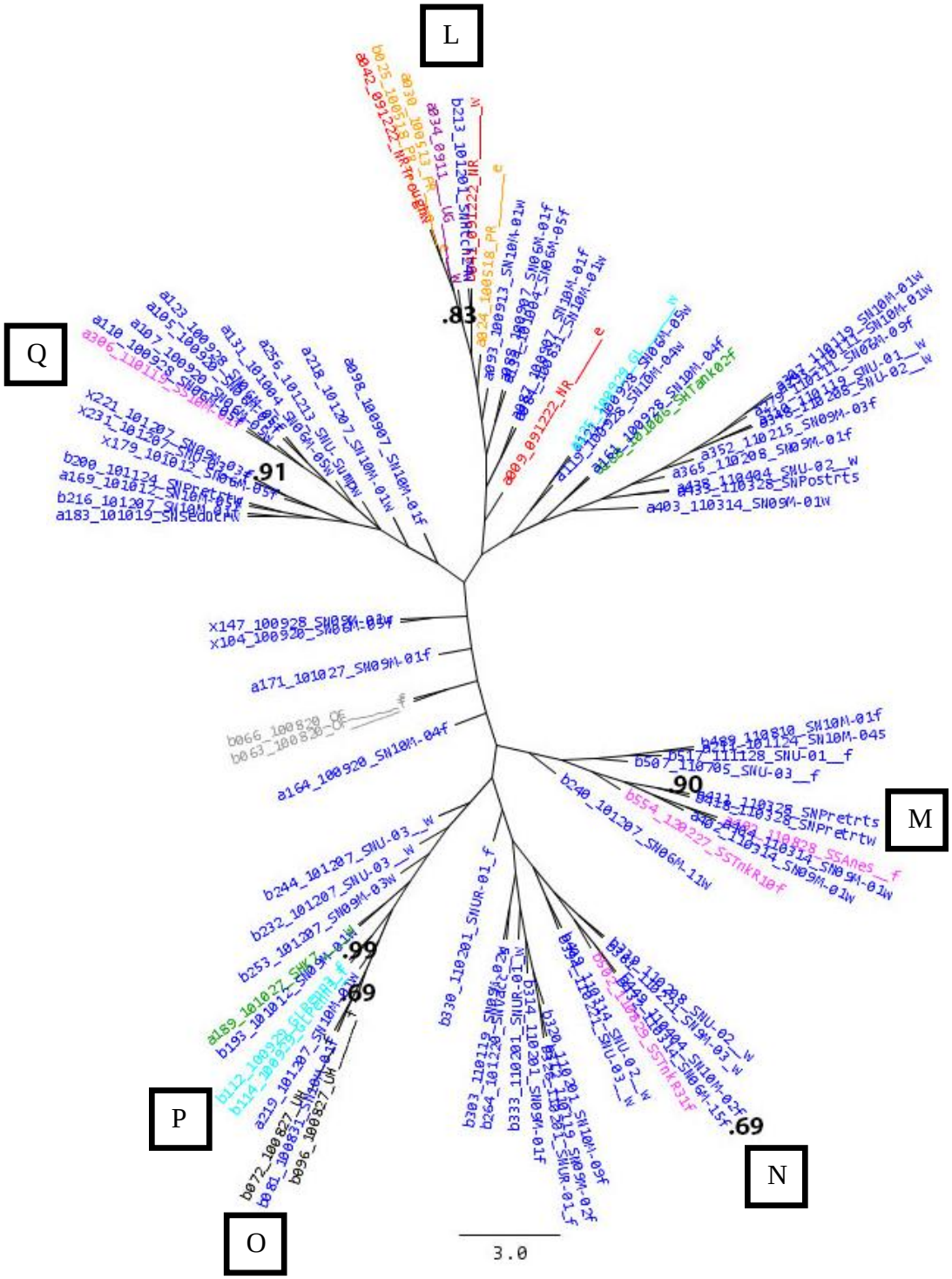


Figure 17. Majority consensus of maximum parsimony analysis of 767 replicates of bootstrapped data for all isolates. Colored according to sample collection location: blue=Sayward North, red=Nanaimo River, green=Stelling Hatchery, aqua=Georgie Lake, orange=Puntledge River, purple=Upper Goldstream, pink=Sayward South, grey=Ocean Falls and black=United Hatchery. Numbers in black indicate bootstrap values >0.60.

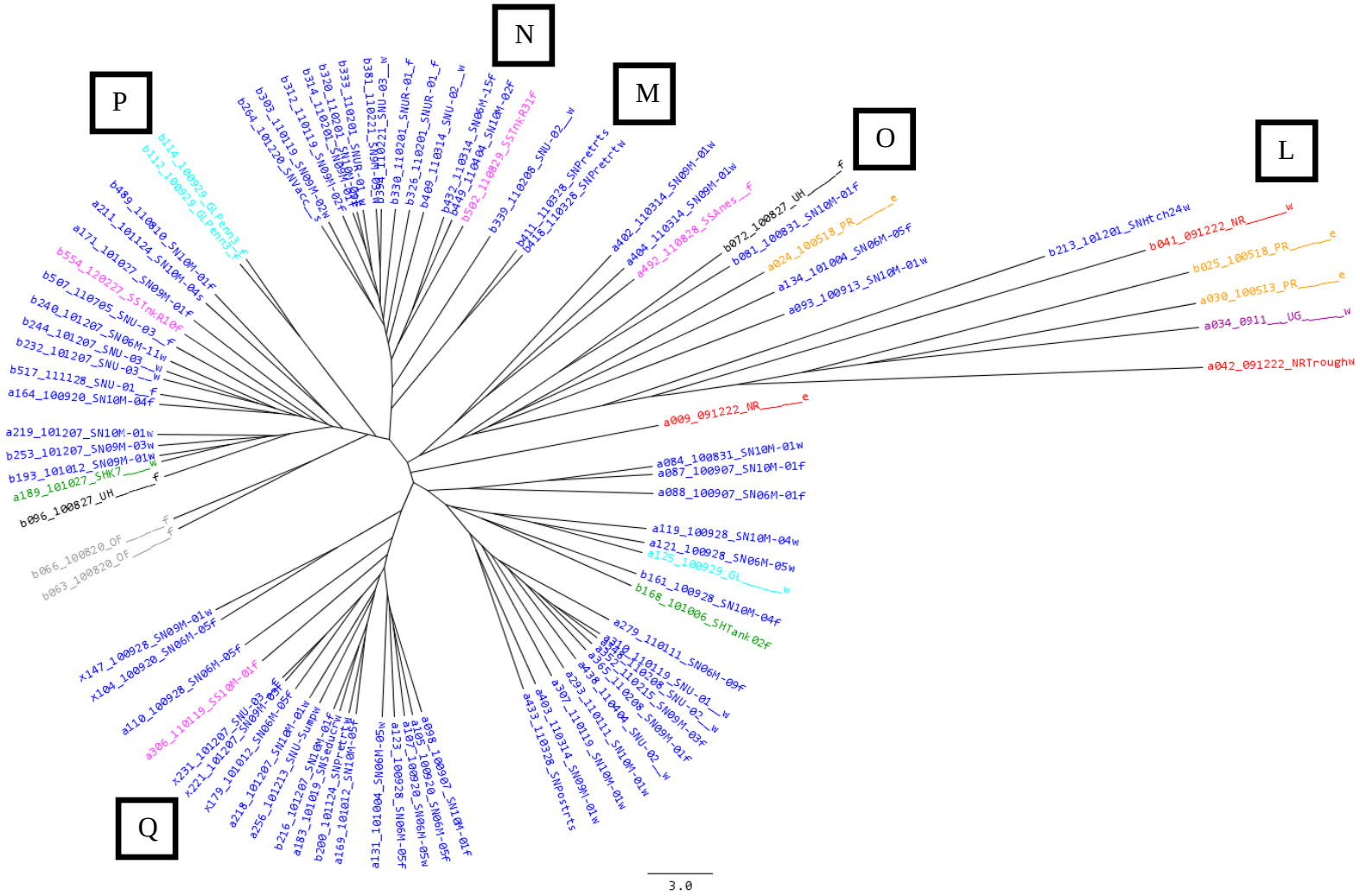


Figure 18. Nei's genetic distance matrix-based tree of all isolates. Colored according to sample collection location: blue=Sayward North, red=Nanaimo River, green=Stelling Hatchery, aqua=Georgie Lake, orange=Puntledge River, pink=Sayward South, grey=Ocean Falls and black=United Hatchery.

In order to further explore the correlation between collection date and genetic distance, a graph of genetic distance between pairs of isolates and the number of days between sample collection at Sayward Hatchery North was generated (Figure 19, 70). Due to the low number of sample pairs with greater than 210 days between sample collections, the graph only included sample collections with up to 210 days difference. This allowed us to visualize the range in genetic distances between isolates for a given number of days between isolate collections. There appeared to be a strong clustering of points between 0 and 200 days difference in collection and 0.1-0.3 genetic distance. The variability in genetic diversity between samples decreased as the number of days between their collections increased; however, the high variability in genetic distance for samples collected closer together resulted in a low fit of a linear trend for the data (R^2=0.0921). Overall it appears there was an increasing trend in genetic distance with time, but there are likely other factors contributing that do not result in a perfectly linear trend.

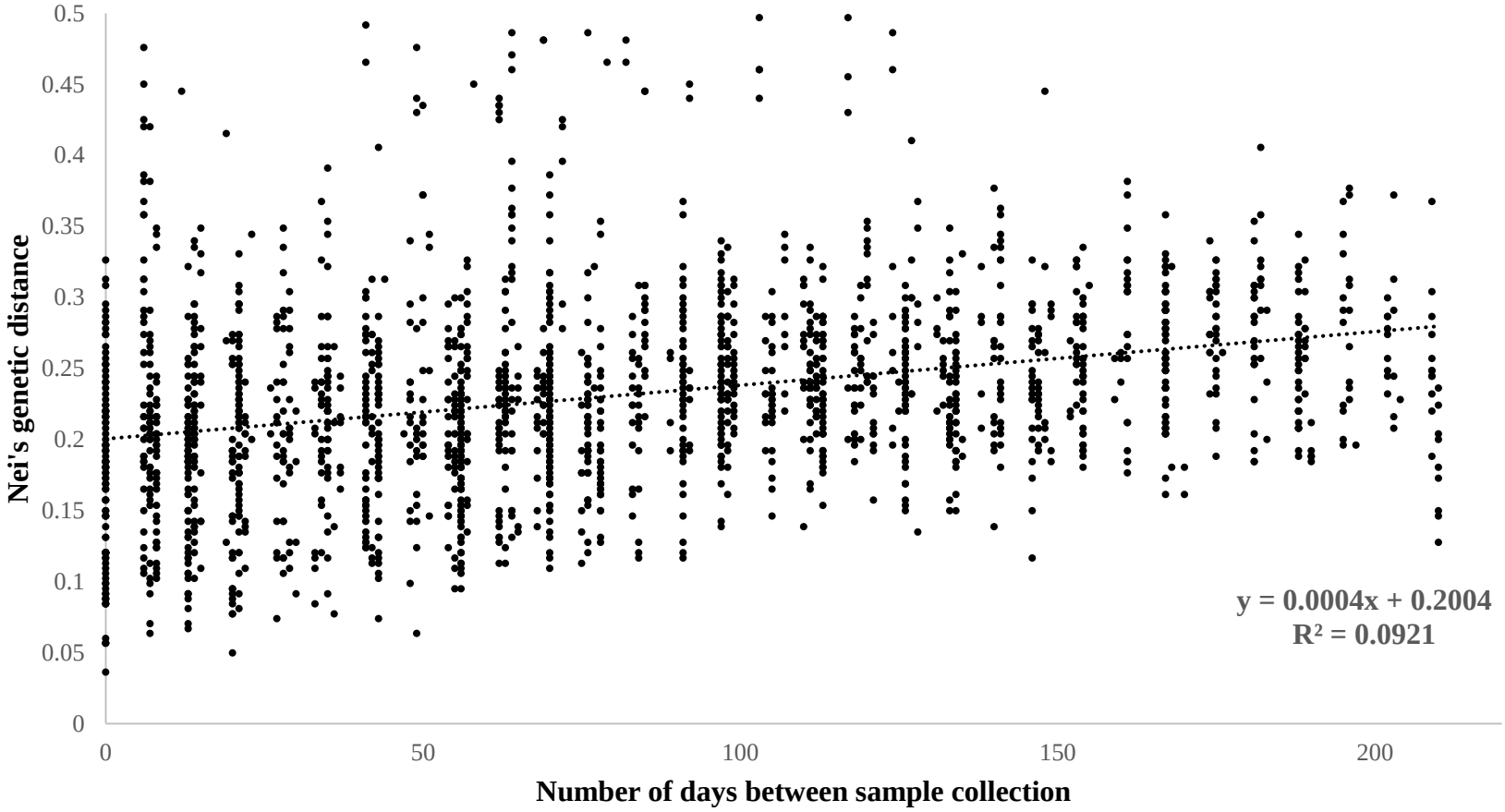


Figure 19. Scatter plot of pair wise Nei's genetic distance and number of days between sample collections for isolates collected at Sayward Hatchery North between August 31st, 2010 and November 28th, 2011. Best fit line equation of y=0.0004x + 0.2004 and an R^2 value of 0.0921, as indicated in lower right hand corner of graph. Pairs of isolates greater than 210 days apart were not included due to low sample sizes.

Discussion

Isolate composition

All of the *S. parasitica* isolates used in this study were collected on the west coast of British Columbia. Although we had a wide variety of overall substrates for the samples collected (eggs, fish, water and swabs), they were not very evenly distributed among the locations sampled (Table 6, 45). Through this sampling we intended to obtain a maximal diversity of genotypes while not necessarily obtaining a representative sample of all genotypes at a given facility. We collected a thorough time series of isolates at Sayward Hatchery North to compare changes in *S. parasitica* diversity and population composition over time and included isolates from a variety of other locations to determine the diversity of *S. parasitica* isolates throughout BC hatcheries. All of these hatcheries potentially share isolates of *S. parasitica* through the transfer of egg and juvenile salmon among facilities, which occurs on a semi-annual basis (Brad Boyce, Marine Harvest Canada, personal communication). United hatchery is a known supplier of fish eggs and juvenile fish to other facilities (Boyce, personal communication); hence, infected fish could possibly provide a source of *S. parasitica* inoculum for other facilities. The hatcheries surveyed also receive eggs or juvenile salmon from other locations outside of those sampled in this study. Because of the potential for regular exchange in genotypes among facilities we could not reasonably divide the isolates into specific populations. As a result, we considered all of the isolates to belong to a single founder population.

Comparison of genotypes

Each gel showed amplicons derived from a single primer with degenerate positions at the 5' end and a short microsatellite at the 3' end. These microsatellite targets were used because of their abundance in the genome. Each primer acted as both a forward and reverse primer, amplifying the sequence between the repeat. Variation in the amplification of a given band, or what we are referring to as an "allele" or amplicon, could result from the presence or absence of a primer binding site due to SNPs or variation in the length of the simple sequence repeats, adjacent microsatellite loci, or region between two degenerate microsatellite primer binding sites (Zhivotovsky, 1999, Altukhov, 2006).

The primers developed by Hantula *et al.* (1996) (Table 2, 31) and primer DAGC (Table 3, 31) were designed such that they annealed outside of the SSR repeat targeted, The individual nucleotides at the 5'end were all the possible base pair combinations except those of the repeat. Due to the possible 5' binding sites, primers BCAG, DAAG, and DAGG (Table 3) could have annealed within the simple sequence repeat. This could lead to inconsistent amplification lengths of the same DNA sequence region, depending on the length of the SSR and where the primer happened to anneal within the repeat region. When comparing the seven primers, there was not a consistent trend in the deviation from 0.5 allele frequency or the number of characters scored for the three primers, compared to the rest of the primers used, so the exact primer binding site probably did not impact the results of this analysis. This makes sense because these are anonymously binding primers that targeted many regions of the genome, hence there

were similar amplification profiles seen overall, even if for one replicate the primer bound at a slightly different location along the DNA sequence.

Due to the nature of the primers there were limitations to the interpretation of the population genetic data. Because these were dominant markers we could not distinguish homozygote or heterozygote genotypes (Zhivotovsky, 1999, Altukhov, 2006). In addition, individual DNA fragments migrating the same distance could have resulted from different and unique sequences. For the purpose of this analysis only the base pair size of each amplicon was scored. This meant that if multiple DNA sequences of the same base pair length were amplified they would be indistinguishable on the agarose gels used and be counted as a single allele. We observed that the staining intensity of the bands could vary quite significantly between amplicons, and this may have resulted from multiple sequences of the same length amplifying, migrating and being scored together. Because quantitative PCR techniques were not used, there was no way to validly include this information to further distinguish alleles or differentiate amplification profiles.

While this general approach was useful for observing intraspecific variation, the methods used in this experiment are only able to detect genotype differences, but do not tell us whether these sequences are part of a coding function that may influence the success, viability or productivity of the individual. The genotype information on its own does not predict whether the polymorphisms may ecologically significant or whether they provide a selective benefit to the population. They could occur due to chance. The only way to assess this would be to determine the function of these sequences and whether they affect fitness or pathogenicity in various isolates.

Population genetics analysis

The overall diversity of *S. parasitica* samples was much lower than anticipated, and can be broadly visualized in Figure 8 (48) and Figure 9 (49) which demonstrated the overall similarity in amplification profiles for the isolates surveyed. There were a high number of polymorphic loci (90% for group 1, 93% for group 2 and 96 % for all isolates combined) for the characters surveyed. For all of the primers combined (total 309 characters surveyed), there was a very broad and even range in the allele frequency for each character scored, ranging from alleles present in all of the isolates to being very minor alleles for the isolates surveyed. The overall trend averages out the differences observed in the slope and intersection of the curve for each of the primer's deviation from 0.5 allele frequency for each of the characters (**Figure 11**, 54). The effective number of loci was closer to one than two (1.32±.33 for group 1, 1.33±0.31 for group 2 and 1.36±0.33 for all isolates combined), indicating that there was a higher abundance of minor alleles (Figure 12, 56). The effective number of alleles in this experiment was slightly lower than the values of 1.58, 1.59 and 1.62 found in populations of 83 isolates of *Phytophthora sojae,* an oomycete soybean pathogen, when surveyed with 20 SSR markers (Wang *et al.*, 2009).

The combination of broad and even distribution in deviation from 0.5 allele frequency and the only slightly higher number of characters with minor allele frequencies indicated that the overall low bootstrap values in the maximum parsimony trees resulted from the phylogenetic relationship of the isolates differing, depending on which characters were used for the phylogenetic tree arrangement. Basically, the low bootstrap

values in this case indicate that some of the characters were contradictory in how they arranged the isolates.

The low Shannon's information index, 0.33±0.22 for group 1 0.30±0.22 for group 2 isolates and 0.36±0.22 for all isolates combined compared to the high percentage of polymorphic loci, 90% for group 1, 93% for group 2 and 96 % for all isolates combined suggests there are a small number of isolates creating most of the diversity in the groups. The genetic distance of three isolates (30, 34, and 42 in group 1 and 25, 41 and 213 in group 2) compared to the rest is clearly demonstrated by Nei's genetic distance, where three of the isolates clustered together but at a significant distance from the rest of the isolates in the tree (Figure 14, 62; Figure 16, 65). A similar trend was seen when all the isolates were combined in a single Nei's genetic distance tree (Figure 18, 68), as isolates in Clade L (213, 41, 25, 30, 34, 42) were significantly further away from the rest of the isolates. The calculated observed number of bands (1.90±0.30 for group 1, 1.93±0.25 for group 2, 1.96±0.19 for all isolates combined) indicated that the majority of characters had variation in the presence or absence of that character, and this can also be observed in Figure 12 (56) by the number of characters (about 12) of the 309 that deviate by 0.5 from 0.5 allele frequency.

When comparing the two trees of bootstrapped data, most of the clades with significant bootstrap values were isolates collected within a short time frame from the same location (Figure 13, 61 and Figure 15, 64). These isolates collected from the same location in a short period of time could be asexual replicates of the same individual. However, because of the differences in genotypes over time at the same location (several different clades of isolates collected from Sayward Hatchery North at different times)

there is likely sexual recombination occurring, and possible some introduction of new genotypes over time. It does not appear that the isolates were clonal replicates. It is impossible to determine how much asexual versus sexual reproduction is occurring in the facilities surveyed; however, our data suggest that there is likely a combination of both.

Another group of isolates of interest are those that make up clades A, G and L in the maximum parsimony and Nei's genetic distance trees. These isolates were collected between November 2009 and May 2010 at three very different locations (Puntledge River, Nanaimo River, and Upper Goldstream). The samples from Nanaimo River and Upper Goldstream were collected from bodies of water outside of hatchery facilities. Possibly why they are grouping together is that the majority of the isolates collected in this experiment have a much more similar genotype to each other than any of these isolates have to the rest. By all being so different from the other isolates, this group of isolates groups together, away from the rest of the isolates

There was no correlation with the source of the isolate (*i.e.* fish, water, egg, and swab) and genotype. One concern for hatchery managers is that there might be specific *S. parasitica* genotypes targeting fish or eggs compared to those generally found in the water column. From our survey it does not appear that genotypes are specific to the type of substrate they infect, but that *S. parasitica* is ubiquitous in the environment. It appears, therefore, that any suitable host within the environment is susceptible to the pathogen. This can make management of the pathogen in aquaculture facilities more difficult, as once the pathogen is introduced it may be very difficult to remove from the system completely.

The reason an analysis of molecular variance (AMOVA) was not attempted, as it usually it is for similar population studies, is that we did not want to make any assumptions about the population groups of the isolates. There was exchange of isolate genotypes within and among facilities as they transport fish eggs and juvenile fish between tanks and hatcheries, and during the time period our samples were collected, likely several exchanges and rearrangements occurred. We also did not want to make the assumption that a given isolate is entirely representative of a given location, with so few isolates collected from some of the locations. This limited our ability to compare populations of *S. parasitica* for all the locations tested. Our series of isolates collected from Sayward Hatchery North, however, provided us a relatively good representation of the changes to the population over time.

For the time series data in Sayward Hatchery North, we know that during that time period there were likely several introductions of new fish and eggs to the facility (Boyce, personal communication) and that the juvenile fish were moved throughout the facility as they grew and were vaccinated. Because of this movement within the facility, the Sayward North Hatchery can be considered as one population of *S. parasitica*, until new isolates are introduced when new fish or eggs are brought to the facility. There was only a slight relationship between the number of days between sample collection and the Nei's genetic distance between the isolates (Figure 19, 70); therefore, there was very little change in genetic distance between isolates collected on the same day and isolates collected many months later. This shows that significantly novel or new genotypes were not being introduced to the hatchery system. This could be because there are not that many differences in *S. parasitica* genotypes or that they are already evenly distributed

among the hatchery locations. There is likely a sufficient amount of sexual recombination occurring within the population at Sayward Hatchery North to create significant variation in genetic distance for isolates collected at a similar time, and the total amount of variation remains consistent for isolates collected close together and isolates collected far apart.

There have been no published reports of *S. parasitica* diversity from isolates collected from the wild, likely due to the extreme difficulty of identifying isolates to the species level and the generally overwhelming abundance of propagules of other *Saprolegnia* species that can be present in the water column, particularly during peak salmon spawning times. Recent work in our lab has provided a set of genetic tools for the specific and rapid identification of *S. parasitica* (Leung, 2012) thus enabling comparative studies. This preliminary analysis suggests that comparing isolates from more geographically disparate field sites might yield greater genetic diversity than those collected from artificial systems such as hatcheries where there can be sharing of isolate genotypes between facilities through the transport of fish or eggs.

Challenges and implications of techniques used

Originally, we compared our two groups of isolates separately, due to the challenges in trying to align isolates across gels. Having the two groups of isolates, in effect, allowed us to replicate our experiment and determine whether a comparison of 46 isolates indicated similar trends as 87 isolates. Once scoring bands and aligning the isolates was more familiar, it was easier to align all 87 isolates as a single group. First the three replicates of a given isolate and primer combination were combined, and then the two groups were aligned using bands amplified consistently across all isolates and the

distinct main bands for each isolate. The comparison of the three data analyses (each group separately and then combined) provided some insight into the benefits and robustness of each data set. It was much easier and less time consuming to score a single group of isolates (46) than the combined replicates and aligned groups of isolates (87); yet there were not many more significant or unique results with the combined data sets than the smaller data sets. This could be due to the isolates being difficult to distinguish as they were so highly similar or because the smaller sample size already gives a relatively even representation of the population surveyed.

It was difficult to interpret the alignment of specific amplicons between the two groups when separated on separate gels, even with five reference isolates (104, 147, 179, 221, 231) included in both amplification sets. Sometimes the bands were quite faint or variable in their location due to distortion in the gel or the "reference" bands for the isolate on either side were also out of alignment or missing. There was significant variation in the calculated (by the ImageLab software) base pair size of each amplicon, which required that each amplicon's alignment be visually verified, and not just aligned based on calculated base pair size. The variability in base pair sizes for each character can be inferred by the calculated mean, standard deviation, maximum and minimum base pair sizes for each character in Appendix 1 Table 9, 93-Table 15, 103. The calculated variation in size was greatest for gels with obvious distortions in the agarose gel (interpreted by the distortions in the ladder across the gel). Some of the bands that ran across all or most isolates showed little variation side by side, but a variation of 10-20 base pairs (or more) across the whole gel, yet they were distinctly recognizable as the same amplicon if followed across the whole gel. For the purposes of scoring the

amplicons, it was assumed that an amplicon in approximately the same relative location for the three replicates was the same sequence of amplicon each time; however, it is possible because of the degenerate primers, that different sequences of approximately the same length could have been amplified in different replicates. The best way to confirm what was amplified for any of the given amplicons would be to excise the bands and sequence the amplifications

In order to minimize the influence of previous assessment of replicates, each gel was first scored independently. When combining the replicates, if a band had only been counted on one of the replicates it was scrutinized to ensure that inclusion was justified. There was a certain level of interpretation for the scoring of bands and for assessing whether bands truly constituted unique bands or were simply small variants in amplification of the same allele. Even misinterpreting or missing a single amplicon across multiple gels of a data set could sufficiently skew the data that two sets would appear more distantly related and divide the isolates into two separate clades corresponding to their assigned experimental groups. It became obvious that with the primers used there were certain limitations of the interpretation of the migration patterns, and that individual bands could only be scored for presence or absence. It was not possible to include variations in intensity of the amplifications (DNA concentration) in the final analysis because we did not complete quantitative PCR reactions and only scored presence or absence of each amplification. Despite these limitations there was very good replication and consistency in gel replicates of the same group of isolates run on separate gels, as was seen in the example Figure 10 (51).

Part of what may contribute to the low bootstrap values for the maximum parsimony trees was not factoring an error rate for scoring of the amplicons. Even following the most exacting character protocol, a particular amplicon had to be classified as either "present" or "absent" and be classified as a certain character, either grouping it with other similarly sized amplicons, or creating a new character category for the size of amplification. The replicates certainly aided in the classifications and determining whether a band was "present" or "absent", but there is likely still a level of error in classification that resulted in noise in the data and creating segregations in the isolates where they may not actually exist. A possible way to address this would be to calculate an error rate, based on the similarity in the replicates, and then apply this error rate to the data as a whole, ignoring phylogenetic separations of the isolates that fall outside of a reasonable error rate. For example, if we allowed for an error rate of 5%, which would be equivalent to approximately 15 of the 309 characters, we would only allow phylogenetic divisions that had at least fifteen differences in character string scoring, and group the isolates that had fewer than fifteen differences together. That is probably a reasonable error rate, as it represents about two characters per primer, but several error rate levels could be compared to see how the phylogenetic trees differ. Another way to map this would be to make a phylogenetic tree and map on the number of character differences between isolates, so you could determine how many character differences actually separate out the individual isolates.

I expected the same number of total characters for group 1, group 2 and all isolates scored together; however this was not the case. The groups of isolates were selected to try and be as representative of all possible genotypes and as random as

possible, and to show the maximum amount of genetic diversity possible in the locations sampled during the given time frame. However, the greater the number of lanes compared, the greater the number of possible characters, because of the increased likelihood of sampling rare characters. This explains why the two groups combined resulted in more characters than either group individually. To explain the difference in the number of characters between group 1 and group 2, group 2 must have had more "unique" genetic profiles. This is supported by the fact that group 2 has more clades that have high bootstrap values, suggesting they are less dependent on which alleles were sampled for determining the phylogenetic tree.

The primers used in this experiment to detect genetic variation among isolates were degenerate at the 5' end and targeted repeat regions of the genome. Because the primers were degenerate there were several possible binding sites for each primer and the amplicons were anonymous as the repeated regions were randomly distributed throughout the genome. Identifying the characters that are most important in distinguishing different genotypes, sequencing those amplified bands, and designing primers specific to those characters would make it easier to distinguish genotypes and clarify the amplifications.

If similar future experiments are to be completed, it would be extremely beneficial to include an appropriate out-group for phylogenetic analysis. Originally, there was an isolate included in group 1 from genus *Aphanomyces*, which is mostly a pathogen of aquatic animals and some plants (Kamoun, 2003). It was initially included in the round 1 isolates, but it because apparent that to ensure more accurate calculations of amplicon base pair sizes, a third ladder in the middle of the large gels was necessary. The outgroup was the easiest isolate to drop without losing an isolate and possible unique

genotype. Because the outgroup isolate was not included for all of the primers tested, the genotype information could not be included in the final analysis.

Future directions

An important aspect for the management of *S. parasitica* in aquaculture facilities is understanding the load at which the pathogen starts to cause significant losses for aquaculture facilities. Because of the low diversity, and lack of specialization of isolates to specific substrates, this pathogen will likely always be present at a low level with the facilities. By knowing the tipping point of when the pathogen starts to cause significant fish loss, we can aid hatchery managers in monitoring the pathogen load level and keeping it below the trigger level. We can also see from these analyses that there were several different genotypes present that could have different aggressivity, which should be monitored to ensure there is not a specific genotype re-occurring and causing significant losses. These questions could be answered by completing challenge studies *in vitro* to control the environment and apply specific genotypes at specific loads to determine the pathogenicity to the fish. *In vitro* experiments would also allow us to control some stress factors on the fish (crowding, temperature, handling, immunization protocols) and how they affect the susceptibility to infection.

While this analysis provides insight into the population genetic structure of *S. parasitica* within hatcheries on the west coast of Canada, it does not address the question of regional differences between isolates. *Saprolegnia parasitica* has a world-wide distribution and may also be unintentionally spread through the shipping of live fish to other facilities. An expanded study could compare isolates collected from a wider geographic area, such as Chile and Norway, which also have prominent aquaculture

industries. Of particular interest would be the comparison of isolates collected from hatcheries in Australia because their Atlantic salmon stock was originally taken from Canadian hatcheries (Dr. Barbara Nowak, Australian Maritime College, personal communication). Presumably, the fish underwent a thorough quarantine during immigration, but it would be interesting to see if there appears to be any close relatedness between *S. parasitica* in Australia and Canada.

Conclusions

The population genetic diversity of *S. parasitica* in BC hatcheries appears to be quite low. This uniformity of genotypes might be due to significant mixing of isolate genotypes due to the movement of fish and eggs between hatchery locations while there may be diversification of genotypes by sexual recombination of founding isolates within a given facility. Isolates collected from outside the hatcheries (Nanaimo River, Upper Goldstream) were distantly related compared to those collected within the hatcheries (Sayward North, Sayward South, Stelling Hatchery). There was a detectable change in the amount of genetic diversity within Sayward Hatchery North over the time course of this study, suggesting that there were either multiple clonally propagating genotypes being introduced over time through hatchery practices or that there was a significant amount of sexual recombination within a facility leading to a diversification of genotypes.

Based on our analysis, *S. parasitica* is a ubiquitous, persistent pathogen that would be best treated by ongoing monitoring and infection prevention methods. We also addressed several questions for hatchery managers indicating that *S. parasitica* appears to be primarily transported through the exchange of fish and fish eggs among facilities. It

also indicates that the lack of quarantine and regular transfer of fish to various tanks and hatcheries contributes to the spread and persistence of *S. parasitica*. The best treatment and prevention methods going forward would be to screen fish and their water before movement, quarantine any incoming fish or eggs, and determine the maximum allowable spore load for fish health. Continued research in effective treatments for saprolegniosis is also necessary as there will likely always be a baseline level of infection. This study provided a good assessment of the population structure of *S. parasitica* in BC hatcheries, but comparison to isolates not contained by the hatchery system, such as field collected populations across the province, is necessary to better understand the pathogenicity of *S. parasitica*.

Bibliography

Aboukhaddour, R., Cloutier, S., Lamari, L., and Strelkov, S.E. 2011. Simple sequence repeats and diversity of globally distributed populations of *Pyrenophora tritici-repentis*. Canadian Journal of Plant Pathology. **33**(3): 389–399.

Ali, E.H. 2005. Morphological and biochemical alterations of oomycete fish pathogen Saprolegnia parasitica as affected by salinity, ascorbic acid and their synergistic action. Mycopathologia. **159**(2): 231–43.

Aller-Gancedo, J.M. and Fregeneda-Grandes, J.M. 2007. Comparative efficacy of Pyceze® (bronopol) in controlling mortality of brown trout *Salmo trutta* eggs. Aquaculture Research. **38**(6): 618–624.

Altukhov, Y.P. 2006. *Intraspecific Genetic Diversity: Monitoring, Conservation and Management*. Berlin, Heidelberg, Germany: Springer-Verlag.

Altukhov, Y.P., Salmenkova, E., and Omelchenko, V. 2000. *Salmonid fishes: population biology, genetics and management*. Oxford, U.K.: Blackwell Science.

Beakes, G. 1983. A comparative account of cyst coat ontogeny in saprophytic and fish-lesion (pathogenic) isolates of the *Saprolegnia diclina - parasitica* complex. Canadian Journal of Botany. **61**(603-625): 603–625.

Beakes, G.W., Glockling, S.L., and Sekimoto, S. 2012. The evolutionary phylogeny of the oomycete "fungi." Protoplasma. **249**(1): 3–19.

Beakes, G.W. and Sekimoto, S. 2009. The evolutionary phylogeny of oomycetes--insights gained from studies of holocarpic parasites of algae and invertebrates. *In*: K. Lamour and S. Kamoun, eds. *Oomycete Genetics and Genomics: Diversity, Interactions, and Research Tools*. Hoboken, New Jersey: John Wiley & Sons Inc., 1–24.

Beakes, G.W., Wood, S.E., and Burr, A.W. 1994. Features which characterize *Saprolegnia* isolates from Salmonid Fish Lesions--A review. *In*: *Salmon Saprolegniasis*. Portland, Oregon: Report to Bonneville Power Administration, 33–66.

Bhadauria, V., Banniza, S., Wei, Y., and Peng, Y.-L. 2009. Reverse genetics for functional genomics of phytopathogenic fungi and oomycetes. Comparative and functional genomics. **2009**: 380719.

Blaustein, A.R., Grant Hokit, D., O'Hara, R.K., and Holt, R.A. 1994. Pathogenic fungus contributes to amphibian losses in the pacific northwest. Biological Conservation. **67**(3): 251–254.

Bly, J.E., Lawson, L.A., Dale, D.J., Szalai, A.J., Durborow, R.M., and Clem, L.W. 1992. Winter saprolegniosis in channel catfish. Diseases Of Aquatic Organisms. **13**: 155–164.

BOLD. Barcode of Life data system 2013. Saprolegnia parasitica {species} - Heterokontophyta; Oomycota; Saprolegniales; Saprolegniaceae; *Saprolegnia*; [online]. Available from: http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=179294 [Accessed 6 Dec 2013].

Broad Institute of Harvard and MIT 2010. No Title [online]. Available from: http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/GenomesIndex.html [Accessed 15 Mar 2014].

Diéguez-Uribeondo, J., Fregeneda-Grandes, J.M., Cerenius, L., Pérez-Iniesta, E., Aller-Gancedo, J.M., Tellería, M.T., Söderhäll, K., and Martín, M.P. 2007. Re-evaluation of the enigmatic species complex *Saprolegnia diclina-Saprolegnia parasitica* based on morphological, physiological and molecular data. Fungal genetics and biology : FG & B. **44**(7): 585–601.

Driscoll, J. 2004. What's killing Klamath's fish? [online]. The Eureka Times-Standard. Available from: http://www.oregonwild.org/waters/klamath/the-klamath-river/klamath-fish-kills/what-s-killing-klamath-s-fish [Accessed 18 Jul 2012].

Felsenstein, J. 1989. PHYLIP--Phylogeny Inference Package (Version 3.2). Cladistics. **5**(2): 163–166.

Fernández-Benéitez, M.J., Ortiz-Santaliestra, M.E., Lizana, M., and Diéguez-Uribeondo, J. 2008. *Saprolegnia diclina*: another species responsible for the emergent disease "*Saprolegnia* infections" in amphibians. FEMS microbiology letters. **279**(1): 23–9.

Food and Agriculture Organization of the United Nations 2012. *The State of World Fisheries and Aquaculture 2012*. Rome: Food and Agriculture organization of the United Nations.

Fregeneda-Grandes, J.M., Fernández Díez, M., and Aller Gancedo, J.M. 2000. Ultrastructural analysis of *Saprolegnia* secondary zoospore cyst ornamentation from infected wild brown trout, *Salmo trutta L.*, and river water indicates two distinct morphotypes amongst long-spined isolates. Journal of Fish Diseases. **23**(2): 147–160.

Gieseker, C., Serfling, S., and Reimschuessel, R. 2006. Formalin treatment to reduce mortality associated with *Saprolegnia parasitica* in rainbow trout, *Oncorhynchus mykiss*. Aquaculture. **253**(1-4): 120–129.

Guerriero, G., Avino, M., Zhou, Q., Fugelstad, J., Clergeot, P.-H., and Bulone, V. 2010. Chitin synthases from *Saprolegnia* are involved in tip growth and represent a potential target for anti-oomycete drugs. PLoS pathogens. **6**(8): e1001070.

Hantula, J., Dusabenyagasani, M., and Hamelin, R.C. 1996. Random amplified microsatellites (RAMS) — a novel method for characterizing genetic variation within fungi. European Journal of Forest Pathology. **26**: 159–166.

Hatai, K., Willoughby, L.G., and Beakes, G.W. 1990. Some characteristics of *Saprolegnia* obtained from fish hatcheries in Japan. Mycological Research. **94**(2): 182–190.

Howe, G.E., Gingerich, W.H., Dawson, V.K., and Olson, J.J. 1999. Efficacy of Hydrogen Peroxide for Treating Saprolegniosis in Channel Catfish. Journal of Aquatic Animal Health. **11**: 222–230.

Hughes, G.C. 1994. Saprolegniasis, Then and Now: A Retrospective. *In*: G.J. Mueller, ed. *Salmon Saprolegniasis*. Portland, Oregon: Report to Bonneville Power Administration, 3–32.

Hulvey, J.P., Padgett, D.E., and Bailey, J.C. 2007. Species Boundaries within *Saprolegnia* (Saprolegniales , Oomycota) Based on Morphological and DNA Sequence Data. Mycological Society of America. **99**(3): 421–429.

Hussein, M.M., Hatai, K., and Nomura, T. 2001. Saprolegniosis in salmonids and their eggs in Japan. Journal of wildlife diseases. **37**(1): 204–7.

Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Hypervariable "minisatellite" regions in human DNA. Nature. **314**(7): 67–73.

Jiang, R.H.Y., de Bruijn, I., Haas, B.J., Belmonte, R., Löbach, L., Christie, J., van den Ackerveken, G., Bottin, A., Bulone, V., Díaz-Moreno, S.M., Dumas, B., Fan, L., Gaulin, E., Govers, F., Grenville-Briggs, L.J., Horner, N.R., Levin, J.Z., Mammella, M., Meijer, H.J.G., Morris, P., Nusbaum, C., Oome, S., Phillips, A.J., van Rooyen, D., Rzeszutek, E., Saraiva, M., Secombes, C.J., Seidl, M.F., Snel, B., Stassen, J.H.M., Sykes, S., Tripathy, S., van den Berg, H., Vega-Arreguin, J.C., Wawra, S., Young, S.K., Zeng, Q., Dieguez-Uribeondo, J., Russ, C., Tyler, B.M., and van West, P. 2013. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. PLoS genetics. **9**(6): e1003272.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Molecular Ecology. **17**(18): 4015–4026.

Judelson, H.S. 2009. Sexual reproduction in oomycetes: biology, diversity, and contributions to fitness. *In*: K. Lamour and S. Kamoun, eds. *Oomycete Genetics and*

Genomics: Diversity, Interactions, and Research Tools. Hoboken, New Jersey: John Wiley & Sons Inc., 121–138.

Kamoun, S. 2003. Molecular Genetics of Pathogenic Oomycetes. American Society for Microbiology. **2**(2): 191–199.

Karaoglu, H., Lee, C.M.Y., and Meyer, W. 2005. Survey of simple sequence repeats in completed fungal genomes. Molecular biology and evolution. **22**(3): 639–49.

Ke, X.L., Wang, J.G., Gu, Z.M., Li, M., and Gong, X.N. 2009. Morphological and molecular phylogenetic analysis of two *Saprolegnia* sp. (Oomycetes) isolated from silver crucian carp and zebra fish. Mycological research. **113**(5): 637–44.

Kimura, M. and Crow, J.F. 1964. The number of alleles that can be maintained in a finite population. Genetics. **49**: 725–738.

Leclerc, M.C., Guillot, J., and Deville, M. 2000. Taxonomic and phylogenetic analysis of Saprolegniaceae (Oomycetes) inferred from LSU rDNA and ITS sequence comparisons. Antonie van Leeuwenhoek. **77**(4): 369–77.

Lee, S. and Moorman, G.W. 2008. Identification and characterization of simple sequence repeat markers for *Pythium aphanidermatum*, *P. cryptoirregulare*, and *P. irregulare* and the potential use in *Pythium* population genetics. Current genetics. **53**(2): 81–93.

Leung, W.L. 2012. The oomycete *Saprolegnia parasitica*: molecular tools for improved taxonomy and species identification. University of Victoria.

Levinson, G. and Gutman, G.A. 1987. Review Article Slipped-Strand Mispairing: A Major Mechanism for DNA Sequence Evolution. Molecular biology and evolution. **4**(3): 203–221.

Lewontin, R.C. 1972. The apportionment of human diversity. Evolutionary Biology. **6**: 381–398.

Luikart, G., England, P.R., Tallmon, D., Jordan, S., and Taberlet, P. 2003. The power and promise of population genomics: from genotyping to genome typing. Nature reviews. Genetics. **4**(12): 981–94.

Meyer, F.P. 1991. Aquaculture disease and health management. Journal of Animal Science. **69**(69): 4201–4208.

Mitchell, A.J., Straus, D.L., Farmer, B., and Carter, R. 2010. Comparison of Percent Health and Fungal Infestation in Channel Catfish after Copper Sulfate, Diquat Bromide, Formalin, and Hydrogen Peroxide Treatment. North American Journal of Aquaculture. **72**: 201–206.

Möller, E.M., Bahnweg, G., Sandermann, H., and Geiger, H.H. 1992. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nucleic acids research. **20**(22): 6115–6.

NCBI. National Centre for Biotechnology Information. 2013. NCBI. National Centre for Biotechnology Information. [online]. Available from: http://www.ncbi.nlm.nih.gov/ [Accessed 6 Dec 2013].

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America. **70**(12): 3321–3.

Nei, M. 1987. Genetic variation within species. *In*: *Molecular Evolutionary Genetics*. New York City: Columbia University Press, 176–187.

Neitzel, D.A., Elston, R.A., and Abernethy, C.S. 2004. *Prevention of Prespawning Mortality: Cause of Salmon Headburns and Cranial Lesions*. Portland, Oregon.

Van der Nest, M.A., Steenkamp, E.T., Wingfield, B.D., and Wingfield, M.J. 2000. Development of simple sequence repeat (SSR) markers in *Eucalyptus* from amplified inter-simple sequence repeats (ISSR). Plant Breeding. **119**: 443–436.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular ecology. **13**(5): 1143–55.

Petrisko, J.E., Pearl, C.A., Pilliod, D.S., Sheridan, P.P., Williams, C.F., Peterson, C.R., and Bury, R.B. 2008. Saprolegniaceae identified on amphibian eggs throughout the Pacific Northwest, USA, by internal transcribed spacer sequences and phylogenetic analysis. Mycologia. **100**(2): 171–80.

Phillips, A.J., Anderson, V.L., Robertson, E.J., Secombes, C.J., and van West, P. 2007. New insights into animal pathogenic oomycetes. Trends in microbiology. **16**(1): 13–19.

Pickering, A.D., Willoughby, L.G., and McGrory, C.B. 1979. Fine structure of secondary zoospore cyst cases of *Saprolegnia* isolates from infected fish. Transactions of the British Mycological Society. **72**(3): 427–436.

Pottinger, T.G. and Day, J.G. 1999. A *Saprolegnia parasitica* challenge system for rainbow trout: assessment of Pyceze as an anti-fungal agent for both fish and ova. Diseases of aquatic organisms. **36**(2): 129–41.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding. **2**(3): 225–238.

Rambaut, A. 2012. FigTree [online]. Available from: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed 15 Aug 2012].

Robertson, E.J., Anderson, V.L., Phillips, A.J., Secombes, C.J., Diéguez-Uribeondo, J., and van West, P. 2009. *Saprolegnia* — fish interactions. *In*: K. Lamour and S. Kamoun, eds. *Oomycete Genetics and Genomics: Diversity, Interactions, and Research Tools*. Hoboken, New Jersey: John Wiley & Sons Inc., 407–424.

SAPRO: Sustainable Approaches to Reduce Oomycete Infection in Aquaculture 2013. Novel Solutions for *Saprolegnia* Control. *In*: *Conference Proceedings: Novel Solutions for Saprolegnia Control*. Glasgow, Scotland: The SAPRO Network.

Stueland, S., Hatai, K., and Skaar, I. 2005. Morphological and physiological characteristics of *Saprolegnia* spp. strains pathogenic to Atlantic salmon, Salmo salar L. Journal of fish diseases. **28**(8): 445–53.

Sun, Q., Hu, K., and Yang, X.-L. 2014. The Efficacy of Copper Sulfate in Controlling Infection of *Saprolegnia parasitica*. Journal of the World Aquaculture Society. **45**(2): 220–225.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in ecology & evolution. **15**(5): 199–203.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular biology and evolution. **24**(8): 1596–9.

Tautz, D. and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic acids research. **12**(10): 4127–4138.

Thermo Fisher Scientific 2014. Assessment of Nucleic Acid Purity [online]. Wilmington, Delaware, USA. Available from: http://www.nanodrop.com/Library/T042-NanoDrop-Spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf [Accessed 28 Apr 2014].

Thoen, E. 2011. Doctoral Thesis: *Saprolegnia* infections in Norwegian salmon hatcheries. Occurance, characteristics and pathogenicity. Norwegian School of Veterinary Science.

De Vicente, M.C., Lopez, C., and Fulton, T. 2003. *Genetic diversity analysis with molecular markers data: Learning module*.

Wang, Z.-Y., Wang, Z.-X., Lu, H.-X., and Shen, J. 2009. Genetic relationships between Anhui and Heilongjiang populations of *Phytophthora sojae* assessed by SSR markers. Mycosystema. **28**(5): 698–704.

Van West, P. 2006. Saprolegnia parasitica, an oomycete pathogen with a fishy appetite: new challenges for an old problem. Mycologist. **20**(3): 99–104.

White, T.J., Bruns, T., Lee, S., and Taylor, J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In*: M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White, eds. *PCR protocols: A guide to methods and applications*. New York City: Academic Press, Inc., 315–322.

Williams, J.G., Hanafey, M.K., Rafalski, J.A., and Tingey, S. V. 1993. Genetic Analysis Using Random Amplified Polymorphic DNA Markers. Methods in Enzymology. **218**: 704–739.

Willson, M.F. 1997. *Variation in salmonid life histories: patterns and perspectives*. Res. Pap. . Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Wright, J.M. and Bentzen, P. 1994. Microsatellites: genetic markers for the future. Reviews in Fish Biology and Fisheries. **4**(3): 384–388.

Yeh, F.C., Yang, R.-C., Boyle, T.B.J., Ye, Z.H., and Mao, J.X. 1997. POPGENE, the user-friendly shareware for population genetic analysis [online]. Available from: http://www.ualberta.ca/~fyeh/popgene.html [Accessed 7 May 2012].

Zhivotovsky, L.A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Molecular ecology. **8**(6): 907–13.

Appendix 1

Table 9. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer ACA.

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
2410	60	2598	2284	0.71	2	1.69	0.41	0.60
2235	38	2304	2159	0.25	2	1.61	0.38	0.57
2135	42	2209	2088	0.11	2	1.26	0.20	0.36
1965	51	2090	1867	0.29	2	1.69	0.41	0.60
1855	49	1955	1770	0.89	2	1.26	0.20	0.36
1640	41	1759	1556	0.93	2	1.15	0.13	0.25
1555	24	1584	1540	0.03	2	1.07	0.07	0.15
1480	24	1516	1413	0.24	2	1.58	0.37	0.55
1445	31	1501	1377	0.93	2	1.15	0.13	0.25
1355	18	1378	1345	0.03	2	1.07	0.07	0.15
1310	23	1357	1260	0.93	2	1.15	0.13	0.25
1270	29	1291	1250	0.02	2	1.05	0.04	0.11
1215	1	1215	1214	0.02	2	1.05	0.04	0.11
1155	21	1189	1056	0.52	2	2.00	0.50	0.69
1105	17	1132	1052	0.95	2	1.10	0.09	0.19
990	21	1026	938	0.25	2	1.61	0.38	0.57
980	6	984	975	0.02	2	1.05	0.04	0.11
950	18	987	915	0.21	2	1.49	0.33	0.51
930	10	956	903	0.99	2	1.02	0.02	0.06
860	8	869	854	0.03	2	1.07	0.07	0.15
845	9	863	824	0.98	2	1.05	0.04	0.11
775	9	793	754	0.99	2	1.02	0.02	0.06
725	18	753	694	0.17	2	1.40	0.29	0.46
715	9	731	702	0.21	2	1.49	0.33	0.51
685	15	712	657	0.29	2	1.69	0.41	0.60
635	15	667	578	0.79	2	1.49	0.33	0.51
595	12	622	567	0.95	2	1.10	0.09	0.19
555	14	565	545	0.02	2	1.05	0.04	0.11

***Rounded to nearest 5 base pairs**

Table 9 (con't). Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer ACA.

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
530	12	558	510	0.95	2	1.10	0.09	0.19
495	16	519	445	0.46	2	1.99	0.50	0.69
465	8	480	451	0.31	2	1.75	0.43	0.62
435	12	464	412	0.25	2	1.61	0.38	0.57
400	10	425	377	0.86	2	1.31	0.24	0.40
385	11	408	363	0.60	2	1.93	0.48	0.67
365	15	380	354	0.02	2	1.05	0.04	0.11
355	8	375	335	0.63	2	1.87	0.47	0.66
350	3	354	343	0.07	2	1.15	0.13	0.25
340	12	359	316	0.28	2	1.67	0.40	0.59
330	8	349	314	0.64	2	1.85	0.46	0.65
320	9	331	295	0.21	2	1.49	0.33	0.51
310	4	310	305	0.02	2	1.05	0.04	0.11
305	4	308	302	0.02	2	1.05	0.04	0.11
290	3	293	287	0.05	2	1.10	0.09	0.19
270	3	274	269	0.01	2	1.02	0.02	0.06
265	6	284	255	0.93	2	1.15	0.13	0.25
230	5	235	224	0.05	2	1.10	0.09	0.19
205	4	215	196	0.90	2	1.23	0.19	0.33
175	2	175	173	0.02	2	1.05	0.04	0.11

***Rounded to nearest 5 base pairs**

Table 10. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer BCAG

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
1900	46	2018	1843	0.40	2	1.93	0.48	0.67
1725	37	1801	1680	0.34	2	1.82	0.45	0.64
1600	32	1644	1535	0.61	2	1.91	0.48	0.67
1525	26	1582	1476	0.86	2	1.31	0.24	0.40
1370	23	1432	1325	0.76	2	1.58	0.37	0.55
1245	27	1289	1183	0.25	2	1.61	0.38	0.57
1060	12	1081	1020	0.95	2	1.10	0.09	0.19
955	10	961	946	0.02	2	1.05	0.04	0.11
940	9	956	921	0.94	2	1.12	0.11	0.22
910	5	917	908	0.03	2	1.07	0.07	0.15
860	7	877	846	1.00	1	1.00	0.00	0.00
785	11	802	743	0.98	2	1.05	0.04	0.11
750	13	759	740	0.02	2	1.05	0.04	0.11
705	19	739	667	0.82	2	1.43	0.30	0.48
690	8	705	672	0.95	2	1.10	0.09	0.19
635	12	651	593	0.94	2	1.12	0.11	0.22
625	N/A	624	624	0.01	2	1.02	0.02	0.06
600	12	644	563	0.95	2	1.10	0.09	0.19
585	N/A	587	587	0.01	2	1.02	0.02	0.06
540	11	557	503	0.99	2	1.02	0.02	0.06
535	N/A	535	535	0.01	2	1.02	0.02	0.06
490	N/A	491	491	0.01	2	1.02	0.02	0.06
455	10	466	447	0.03	2	1.07	0.07	0.15
450	7	461	429	0.99	2	1.02	0.02	0.06
415	6	426	398	0.94	2	1.12	0.11	0.22
380	14	393	357	0.06	2	1.12	0.11	0.22
355	6	365	331	0.98	2	1.05	0.04	0.11
335	8	351	306	0.95	2	1.10	0.09	0.19
290	7	309	278	0.21	2	1.49	0.33	0.51
275	1	276	273	0.05	2	1.10	0.09	0.19
275	5	281	264	0.95	2	1.10	0.09	0.19
250	4	262	243	0.78	2	1.52	0.34	0.52
215	5	222	205	0.99	2	1.02	0.02	0.06
205	N/A	204	204	0.01	2	1.02	0.02	0.06
175	7	188	170	0.07	2	1.15	0.13	0.25

***Rounded to nearest 5 base pairs**

Table 11. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer CCA

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
2310	17	2345	2270	0.25	2	1.61	0.38	0.57
2225	41	2280	2152	0.10	2	1.23	0.19	0.33
2005	24	2036	1964	0.10	2	1.23	0.19	0.33
1830	28	1877	1779	0.24	2	1.58	0.37	0.55
1720	41	1755	1662	0.05	2	1.10	0.09	0.19
1610	17	1657	1573	0.52	2	2.00	0.50	0.69
1540	24	1603	1486	0.40	2	1.93	0.48	0.67
1480	37	1589	1431	0.18	2	1.43	0.30	0.48
1345	34	1406	1298	0.16	2	1.37	0.27	0.44
1340	34	1421	1291	0.36	2	1.85	0.46	0.65
1290	46	1380	1221	0.38	2	1.89	0.47	0.66
1215	25	1268	1167	0.49	2	2.00	0.50	0.69
1155	12	1208	1148	0.48	2	2.00	0.50	0.69
1120	15	1178	1094	0.48	2	2.00	0.50	0.69
1120	15	1140	1083	0.25	2	1.61	0.38	0.57
1080	31	1117	1031	0.25	2	1.61	0.38	0.57
1025	12	1049	992	0.56	2	1.97	0.49	0.69
1005	14	1029	964	0.62	2	1.89	0.47	0.66
985	20	1006	943	0.46	2	1.99	0.50	0.69
955	13	982	941	0.13	2	1.28	0.22	0.38
925	8	940	895	0.80	2	1.46	0.31	0.49
890	10	910	877	0.14	2	1.31	0.24	0.40
860	20	904	825	0.32	2	1.77	0.44	0.63
810	15	858	766	0.54	2	1.99	0.50	0.69
805	22	862	777	0.16	2	1.37	0.27	0.44
770	27	806	737	0.06	2	1.12	0.11	0.22
730	9	755	705	0.98	2	1.05	0.04	0.11
710	14	729	692	0.07	2	1.15	0.13	0.25
660	15	703	629	0.78	2	1.52	0.34	0.52
595	9	615	579	0.99	2	1.02	0.02	0.06
565	10	585	536	0.36	2	1.85	0.46	0.65
550	8	565	527	0.20	2	1.46	0.31	0.49

***Rounded to nearest 5 base pairs**

Table 11 (con't). Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer CCA

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
505	11	528	486	0.89	2	1.26	0.20	0.36
475	19	493	444	0.07	2	1.15	0.13	0.25
440	10	455	425	1.00	1	1.00	0.00	0.00
380	9	394	366	1.00	1	1.00	0.00	0.00
330	10	357	301	0.83	2	1.40	0.29	0.46
295	8	315	276	0.71	2	1.69	0.41	0.60
270	8	291	262	0.20	2	1.46	0.31	0.49
250	6	267	238	0.54	2	1.99	0.50	0.69
230	7	241	218	0.24	2	1.58	0.37	0.55
210	2	211	206	0.05	2	1.10	0.09	0.19
195	5	201	186	0.09	2	1.20	0.17	0.31
175	5	180	174	0.02	2	1.05	0.04	0.11
170	3	177	161	0.52	2	2.00	0.50	0.69
150	3	161	145	0.36	2	1.85	0.46	0.65

***Rounded to nearest 5 base pairs**

Table 12. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer DAAG

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
2365	64	2483	2243	0.39	2	1.91	0.48	0.67
2215	24	2254	2173	0.25	2	1.61	0.38	0.57
2130	27	2182	2065	0.28	2	1.67	0.40	0.59
1960	24	2014	1913	0.48	2	2.00	0.50	0.69
1895	35	1932	1798	0.16	2	1.37	0.27	0.44
1780	38	1854	1702	0.85	2	1.34	0.25	0.42
1705	27	1774	1642	0.57	2	1.96	0.49	0.68
1665	17	1701	1618	0.92	2	1.17	0.15	0.28
1605	27	1658	1546	0.69	2	1.75	0.43	0.62
1565	19	1605	1526	0.30	2	1.72	0.42	0.61
1540	23	1600	1494	0.77	2	1.55	0.35	0.54
1485	18	1543	1428	0.85	2	1.34	0.25	0.42
1430	15	1465	1387	0.83	2	1.40	0.29	0.46
1375	15	1405	1347	0.63	2	1.87	0.47	0.66
1315	14	1336	1286	0.26	2	1.64	0.39	0.58
1280	16	1331	1243	0.72	2	1.67	0.40	0.59
1215	14	1249	1187	0.47	2	1.99	0.50	0.69
1160	19	1191	1128	1.00	1	1.00	0.00	0.00
1105	8	1119	1092	0.13	2	1.28	0.22	0.38
1080	6	1090	1069	0.29	2	1.69	0.41	0.60
1045	14	1070	1013	0.62	2	1.89	0.47	0.66
990	22	1019	935	0.55	2	1.98	0.49	0.69
945	13	984	927	0.31	2	1.75	0.43	0.62
900	20	956	851	0.69	2	1.75	0.43	0.62
870	11	886	857	0.10	2	1.23	0.19	0.33
845	N/A	846	846	0.01	2	1.02	0.02	0.06
815	16	825	785	0.07	2	1.15	0.13	0.25
770	15	793	740	0.95	2	1.10	0.09	0.19
745	7	752	735	0.06	2	1.12	0.11	0.22
710	9	722	684	0.30	2	1.72	0.42	0.61
675	3	676	672	0.02	2	1.05	0.04	0.11
660	13	679	628	0.93	2	1.15	0.13	0.25

***Rounded to nearest 5 base pairs**

Table 12 (cont'). Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer DAAG

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
640	2	639	636	0.03	2	1.07	0.07	0.15
615	13	629	592	0.11	2	1.26	0.20	0.36
610	11	615	600	0.02	2	1.05	0.04	0.11
595	13	616	565	0.98	2	1.05	0.04	0.11
570	16	601	518	0.92	2	1.17	0.15	0.28
550	3	550	546	0.02	2	1.05	0.04	0.11
540	9	559	526	0.10	2	1.23	0.19	0.33
500	14	532	471	1.00	1	1.00	0.00	0.00
450	26	516	401	0.85	2	1.34	0.25	0.42
435	4	444	432	0.16	2	1.37	0.27	0.44
400	10	421	391	0.08	2	1.17	0.15	0.28
390	10	406	367	0.95	2	1.10	0.09	0.19
380	1	380	378	0.02	2	1.05	0.04	0.11
365	4	366	361	0.02	2	1.05	0.04	0.11
335	5	343	323	0.24	2	1.58	0.37	0.55
300	3	310	294	0.32	2	1.77	0.44	0.63
260	7	268	251	0.07	2	1.15	0.13	0.25

***Rounded to nearest 5 base pairs**

Table 13. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer DAGC

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
1705	23	1741	1646	0.25	2	1.61	0.38	0.57
1580	42	1649	1473	0.68	2	1.77	0.44	0.63
1495	26	1553	1449	0.52	2	2.00	0.50	0.69
1290	11	1305	1268	0.10	2	1.23	0.19	0.33
1240	19	1274	1179	0.86	2	1.31	0.24	0.40
1175	13	1198	1146	0.13	2	1.28	0.22	0.38
1060	16	1091	1015	0.59	2	1.94	0.49	0.68
1015	15	1034	993	0.11	2	1.26	0.20	0.36
920	17	952	880	0.98	2	1.05	0.04	0.11
860	20	913	808	1.00	1	1.00	0.00	0.00
835	16	874	802	0.97	2	1.07	0.07	0.15
780	15	816	744	0.94	2	1.12	0.11	0.22
720	17	764	705	0.09	2	1.20	0.17	0.31
675	10	697	664	0.09	2	1.20	0.17	0.31
635	17	669	581	0.92	2	1.17	0.15	0.28
600	13	618	572	0.16	2	1.37	0.27	0.44
540	14	570	505	1.00	1	1.00	0.00	0.00
490	13	520	452	0.76	2	1.58	0.37	0.55
445	11	472	413	1.00	1	1.00	0.00	0.00
415	13	446	382	0.95	2	1.10	0.09	0.19
375	11	393	351	0.72	2	1.67	0.40	0.59
345	7	358	329	0.14	2	1.31	0.24	0.40
335	10	364	309	0.98	2	1.05	0.04	0.11
310	9	333	294	0.70	2	1.72	0.42	0.61
295	7	313	283	0.23	2	1.55	0.35	0.54
275	8	295	259	0.60	2	1.93	0.48	0.67
250	6	257	242	0.07	2	1.15	0.13	0.25
245	7	263	231	0.82	2	1.43	0.30	0.48
220	8	238	215	0.08	2	1.17	0.15	0.28
220	6	230	205	0.69	2	1.75	0.43	0.62
210	2	212	205	0.06	2	1.12	0.11	0.22
185	5	191	179	0.10	2	1.23	0.19	0.33
170	5	182	160	0.54	2	1.99	0.50	0.69

***Rounded to nearest 5 base pairs**

Table 14. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer DAGG

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
2400	108	2598	2240	0.47	2	1.99	0.50	0.69
1720	36	1789	1650	0.66	2	1.82	0.45	0.64
1610	15	1620	1599	0.01	2	1.02	0.02	0.06
1530	27	1584	1469	0.71	2	1.69	0.41	0.60
1475	59	1525	1317	0.32	2	1.77	0.44	0.63
1365	12	1390	1349	0.21	2	1.49	0.33	0.51
1350	23	1402	1281	0.72	2	1.67	0.40	0.59
1285	21	1346	1248	0.95	2	1.10	0.09	0.19
1230	31	1295	1201	0.07	2	1.15	0.13	0.25
1180	18	1229	1134	0.95	2	1.10	0.09	0.19
1120	31	1139	1096	0.01	2	1.02	0.02	0.06
1060	16	1103	1026	0.99	2	1.02	0.02	0.06
995	17	1038	965	0.94	2	1.12	0.11	0.22
980	9	993	956	0.37	2	1.87	0.47	0.66
950	N/A	950	950	0.01	2	1.02	0.02	0.06
910	17	930	882	0.06	2	1.12	0.11	0.22
890	13	916	853	0.80	2	1.46	0.31	0.49
875	15	903	854	0.15	2	1.34	0.25	0.42
870	32	940	816	0.84	2	1.37	0.27	0.44
845	20	895	808	0.85	2	1.34	0.25	0.42
820	17	863	806	0.09	2	1.20	0.17	0.31
800	12	835	787	0.08	2	1.17	0.15	0.28
755	13	806	723	0.49	2	2.00	0.50	0.69
740	17	782	714	0.99	2	1.02	0.02	0.06
720	4	724	717	0.02	2	1.05	0.04	0.11
690	8	700	674	0.09	2	1.20	0.17	0.31
670	25	718	638	0.16	2	1.37	0.27	0.44
635	14	670	602	0.57	2	1.96	0.49	0.68
630	18	673	598	0.97	2	1.07	0.07	0.15
590	12	625	554	0.30	2	1.72	0.42	0.61
575	17	622	553	0.95	2	1.10	0.09	0.19
555	11	568	533	0.03	2	1.07	0.07	0.15

***Rounded to nearest 5 base pairs**

Table 14 (con't). Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer DAGG

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
530	15	564	503	0.98	2	1.05	0.04	0.11
515	2	517	514	0.01	2	1.02	0.02	0.06
495	10	515	482	0.21	2	1.49	0.33	0.51
480	12	498	463	0.07	2	1.15	0.13	0.25
455	15	489	429	0.95	2	1.10	0.09	0.19
435	10	449	412	0.25	2	1.61	0.38	0.57
425	14	461	404	0.67	2	1.80	0.44	0.64
410	6	415	403	0.03	2	1.07	0.07	0.15
395	9	412	379	0.16	2	1.37	0.27	0.44
375	8	389	356	0.85	2	1.34	0.25	0.42
365	14	398	343	0.93	2	1.15	0.13	0.25
350	12	373	330	0.08	2	1.17	0.15	0.28
315	3	317	310	0.03	2	1.07	0.07	0.15
305	11	332	284	1.00	1	1.00	0.00	0.00
275	9	299	260	0.25	2	1.61	0.38	0.57
260	7	282	240	0.80	2	1.46	0.31	0.49
250	7	264	240	0.44	2	1.97	0.49	0.69
245	6	259	231	0.49	2	2.00	0.50	0.69
220	8	243	206	0.98	2	1.05	0.04	0.11
190	7	205	174	0.91	2	1.20	0.17	0.31

***Rounded to nearest 5 base pairs**

Table 15. Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer GT

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
1995	9	1999	1987	0.02	2	1.05	0.04	0.11
1800	14	1812	1766	0.09	2	1.20	0.17	0.31
1610	29	1644	1543	0.17	2	1.40	0.29	0.46
1540	33	1611	1464	0.70	2	1.72	0.42	0.61
1395	23	1454	1368	0.13	2	1.28	0.22	0.38
1285	24	1330	1254	0.16	2	1.37	0.27	0.44
1235	19	1274	1185	0.54	2	1.99	0.50	0.69
1205	2	1205	1201	0.03	2	1.07	0.07	0.15
1165	20	1201	1129	0.32	2	1.77	0.44	0.63
1105	13	1133	1069	0.54	2	1.99	0.50	0.69
1075	18	1117	1042	0.29	2	1.69	0.41	0.60
1030	18	1060	980	0.90	2	1.23	0.19	0.33
970	7	978	960	0.09	2	1.20	0.17	0.31
930	16	962	883	0.87	2	1.28	0.22	0.38
895	3	899	894	0.02	2	1.05	0.04	0.11
870	16	907	832	0.92	2	1.17	0.15	0.28
850	4	854	844	0.06	2	1.12	0.11	0.22
825	10	850	808	0.13	2	1.28	0.22	0.38
800	14	835	763	0.90	2	1.23	0.19	0.33
745	15	799	712	0.97	2	1.07	0.07	0.15
735	5	744	731	0.05	2	1.10	0.09	0.19
720	11	743	691	0.75	2	1.61	0.38	0.57
700	8	712	685	0.17	2	1.40	0.29	0.46
675	8	691	665	0.10	2	1.23	0.19	0.33
655	12	697	620	0.97	2	1.07	0.07	0.15
645	10	656	633	0.05	2	1.10	0.09	0.19
615	11	634	577	0.92	2	1.17	0.15	0.28
590	13	613	579	0.06	2	1.12	0.11	0.22
545	15	589	521	0.51	2	2.00	0.50	0.69
535	8	552	509	0.47	2	1.99	0.50	0.69
505	12	526	472	0.77	2	1.55	0.35	0.54
480	21	514	429	1.00	1	1.00	0.00	0.00

***Rounded to nearest 5 base pairs**

Table 15 (con't). Character scoring data indicating mean character base pair (bp) size with standard deviation, maximum and minimum base pair size for a given character, proportion of isolates showing presence for a given character, observed number of alleles (n_a), effective number of alleles (n_e), Nei's gene diversity (H_e) and Shannon's information index (I) for each character scored for combined groups of isolates (87) for primer GT

Character bp size*	Standard deviation in bp size	Max. bp size	Min. bp size	Proportion of isolates with presence	Observed number of alleles (n_a)	Effective number of alleles (n_e)	Nei's gene diversity (H_e)	Shannon's information index (I)
460	17	473	436	0.03	2	1.07	0.07	0.15
435	9	457	418	0.97	2	1.07	0.07	0.15
390	10	418	371	0.51	2	2.00	0.50	0.69
370	10	398	340	0.98	2	1.05	0.04	0.11
345	8	361	330	0.25	2	1.61	0.38	0.57
320	7	342	303	0.72	2	1.67	0.40	0.59
300	4	306	295	0.09	2	1.20	0.17	0.31
280	7	299	266	1.00	1	1.00	0.00	0.00
265	1	268	266	0.02	2	1.05	0.04	0.11
255	7	273	239	1.00	1	1.00	0.00	0.00
230	7	250	211	0.93	2	1.15	0.13	0.25
195	6	209	181	0.97	2	1.07	0.07	0.15
175	8	184	167	0.01	2	1.02	0.02	0.06
170	8	179	157	0.02	2	1.05	0.04	0.11

***Rounded to nearest 5 base pairs**