POPULATION STRUCTURE OF *PUCCINIA GRAMINIS* F. SP. *TRITICI* IN THE UNITED STATES

A THESIS SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERISTY OF MINNESOTA
BY

SAM STOXEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Advisor: Les J Szabo

May, 2012

Acknowledgements

I would like to acknowledge Dr. Les Szabo, my advisor, for his support, assistance and advice throughout this process. I would also like to thank my committee members, Drs. Peter Tiffin and James Bradeen, for their advice and comments. I am also grateful for the assistance and thoughtful discussions with members of Cereal Disease Lab most notably Drs. Jerry Johnson, Matt Rouse, Yue Jin and Alan Roelfs. Lastly I would like to thank the faculty, staff and fellow graduate students in the Department of Plant Pathology.

Dedication

I would like to dedicate this thesis to my family, especially my mom and Angela Ridgeway for their unending support.

Table of Contents

List of Tables

Chapter 1

Chapter 2

List of Figures

Chapter 1

Chapter 2

General Introduction

Emerging Infectious Disease

Infectious diseases of the food supply have plagued man for centuries. Currently the estimated loss of food and cash crops due to diseases is approximately 42% of crop production, which is more than 300 billion dollars (Oerke et al. 1994). The majority of these losses are due to emerging infectious diseases (EIDs) and the epidemics that are commonly associated with a new pathogen. EIDs are defined by pathogens that: have increased in incidence, geographic or host range, have changed pathogenesis, have newly evolved or have been newly recognized (Lederberd et al 1992, Daszak et al 2000). Technological advances such as enzyme-linked immunosorbent assay (ELISA) and real time polymerase chain reaction (PCR) assays have given researchers the ability to detect and diagnose EIDs much quicker than in the past but EIDs are still a threat to global food security and social stability. In the past twenty years an assortment of EIDs on important food crops have surfaced, such as potato late blight, rice blast, karnal bunt, cassava mosaic virus, tomato yellow leaf curl and citrus canker. These diseases have the ability to wipe out upwards of 90% of the host plant (Vurro et al. 2010).

Puccinia gramins f. sp. *tritici* (*Pgt*) has historically been a major contributor to the loss of wheat crops worldwide but severe epidemics have not been seen in the United States since the 1950's. Unfortunately a new race of *Pgt* was identified in Uganda in 1999 (Ug99) that has the ability to overcome the majority of the resistance genes that have provided protection for the last 60 years (Singh et al 2011). Few resistance genes effective against the spread of the Ug99 have been identified but most will require extensive breeding programs and pyramiding of genes will be necessary for prolonged protection (Vurro et al. 2010). Since 1999, Ug99 and other phenotypically similar isolates termed the Ug99 race group have spread to neighboring countries causing epidemics resulting in as much as 70% loss to the total wheat produced (Vurro et al. 2010). In many of the affected countries wheat is a staple crop providing about 40% of

the caloric intake (Vurro et al. 2010) so major losses to production result in countries importing more food causing global demand and crop prices to increase.

The Ug99 race group has not made it to the United States but the likelihood of an introduction in the next decade is quite high based on the dispersal mechanisms of *Pgt* and the large amount of susceptible germplasm (Singh et al. 2006). In order to understand how an introduction of a new virulent race of *Pgt* will affect the population dynamics of the fungus in the United States an understanding of the biology of the pathogen, history of *Pgt* since its introduction and the current population structure of the organism in the United States are needed.

Stem rust fungus *Puccinia graminis*

Puccinia graminis is a basidiomycete rust fungus that has a broad host range, being able to infect 365 species of plants (Anikster 1984). *P. graminis* has been broken into subspecies and formae speciales based on spore morphology, fertility crosses and host range. Urban (1967) separated the species into two subspecies based on morphology. Subspecies *graminicola* contained stem rusts found on non-cereal grasses while subspecies *graminis* contained the stem rusts found on cereal crops. Subspecies *graminis* was then further divided into variety *stakmanii*, found on barley, oat and rye and variety *graminis*, which is found mainly on wheat. The subspecies proposed by Urban are in disagreement with formae speciales designations based on fertility crosses. Crosses between formae speciales *tritici* and *secalis* and between *avenae* and *poae* were found to produce viable offspring (Johnson 1949). Neighbor joining and parsimony analyses of the ITS region grouped *Puccinia graminis* ff. sp. *avenae*, *dacylis, lolii* and *poae* with high support as a clade and ff. sp. *secalis* and *tritici* with high support as a second clade, further supporting the formae speciales designations (Zambino and Szabo 1993). Zambino and Szabo (1993) found the closest relatives to *P. graminis* in their data set were *P. recondita* f. sp. *tritici*, *P. recondita* f. sp. *agropyri* and *P. brachypodii* whereas a subsequent analysis also using ITS sequence but different sampling of species found *P. coronata* f. sp. *avenae* and *P. coronata* f. sp. *lolii* to be the closest relatives (Jafary et al. 2006) .

Pgt infects cereal crops as urediniospores during the warm months. Urediniospores are dikaryotic (n+n) and spread via agitation of the host plant such as from rain or wind. These spores can land on other cereal hosts and germinate to cause new infection. This process is the asexual stage of the fungus and can result in rapid increase of the organism. Near late summer when temperatures begin to drop teliospores are produced. These teliospores undergo karyogomy before overwintering (Boehm et al 1992). In early spring the teliospores germinate producing 4 haploid basidiospores per spore cell (Roelfs 1985). Basidiospores are carried by agitation to *Berberis* species where they directly penetrate the leaf tissue forming pycnia. Pycnia produce receptive hyphae, haploid pycniospores belonging to one of two mating types and exudates (Johnson and Newton 1946). The pycniospores can be wind and water dispursed as well as transferred by insects that are attracted to the exudates (Leonard and Szabo 2005). When pycniospores of one mating type are transferred to receptive hyphae of a pycnium of the compatible mating type an aecium forms below the pycnium (Johnson and Newton 1946). The aecium produces diploid aeciospores, which infect cereal crops and form uredinea (Roelfs 1985).

In presence of water urideospores will germinate producing a germ tube (Leonard and Szabo 2005). The germ tube grows along the epidermis toward the stoma. If the germ tube comes in contact with a stoma it will form an appressorium. Out of the bottom of the appressorium a penetration peg will grow into the substomatal cavity and form a vesicle. The substomatal vesicle will then elongate and hyphae will grow inside the plant tissue. Once a hypha comes in contact with a cell wall a haustorial mother cell forms. The haustorial mother cell exudes enzymes in order to dissolve the plant cell wall (Harder and Chong 1984). Upon invaginating the cell a haustoria is produced. The haustoria are responsible for the exchange of nutrients and proteins with the plant and are required in sustaining the growth of the fungus.

The genome of *Pgt* was determined to be made of 18 chromosomes (Boehm et al 1992) and was estimated by genome sequence assembly to be 88.6 Mb (Duplessis et al 2005). No evidence for a whole genome duplication was found but the larger than

expected genome seems to be a result of a large number of transposable elements, which composed 45% of the genome (Duplessis et al 2005). 17,773 protein coding genes were predicted, of which only 35% showed significant homology to known proteins (Duplessis et al. 2005).

Race phenotypes

The races of *Pgt* are determined by phenotypic avirulence/virulence testing on a standard set of wheat differentials. Nomenclature used for designating *Pgt* races has evolved, starting with a chronological numbering system by Stakman (Stakman and Levine 1922) into the current letter system (Roelfs and Martens 1988). This system is based on the infection types different isolates of *Pgt* produce on lines of wheat. Infection types are classified as; 0, no infection, ;, flecking, 1, small uredinia often surrounded by necrosis, 2, small to medium size uredinia surrounded by necrosis or chlorosis, 3, medium uredia associated with chlorosis and 4, large uredinia without chlorosis or necrosis (Roelfs and Martens 1988). An infection type of 0,;,1 or 2 was designated to be a resistant reaction where as a 3 or 4 a susceptible reaction.

As our understanding of rust and wheat genetics increased the wheat lines used to determine rust races changed. In the 60's it became evident that single resistance gene wheat lines needed to be developed in order to have meaningful discussions between scientists in different countries. Currently there are at least 50 resistance genes that have been identified, located on a chromosomal arm and proven to be distinct from other genes (McIntosh et al. 1995, Singh et al. 2011). In 1988 Roelfs and Martens published the system of naming that has become the standard for race identification. This system uses single gene wheat lines in sets of four to assign letter codes based on resistance and susceptibility to rust. The standard set of *Pgt* wheat differentials originally contained 12 lines (three-letter code) was first expended to 16 lines (four-letter code) in 1988 (Roelfs and Martens 1988) and recently to 20 lines (five-letter code) (Jin et al 2008).

Race groups define sets of races with similar phenotypes by representing the variation between the different members of the race group with an underscore ("_"). For

example, the race group MCC_ includes the races MCCD and MCCF, which have different avirulence/virulence phenotypes on the wheat differential line CnsSrTmp containing the wheat stem rust resistance gene *SrTmp*. Line CnsSrTemp is the fourth differential in the fourth set and therefore the fourth letter in the race code is changed to an underscore. Since the 1930's only eleven phenotypic race groups have been common in the central United States. Based on 16 single gene lines these phenotypic race groups were defined as T__K, R__Q, H_LG, RCRS, RHTS, MCC_, _KCJ, _THJ, Q_CQ, QFCS and QCCJ (Roelfs et al. 1997). Even though some race groups contain a single race phenotype they are considered race groups for the ease of discussion.

Wheat stem rust disease

Wheat stem rust has been a problem for human civilizations for thousands of years. According to the Christian Bible the Israelites suffered from severe epidemics of rust as a punishment for their straying from the lord (Chester 1946) and rust infected wheat tissue has even been found in Israel dating back to the Bronze Age (3300-1200 BCE) (Kislev 1982). Roman history also shows fear of epidemics and can be seen in the festival of Robigalia where they gave praise to Robigo in the belief that their crops would prosper with the favor of the gods (Chester 1946). Both the hosts and pathogen were first reported in the Old World and wheat stem rust was not an important disease of the New World until both hosts became prevalent.

Common barberry was brought to North American in the seventeenth century by European settlers (Lind 1915). The fruit of the barberry bush was used for making jelly, juice, wine and vinegar, while the bush as a whole provided a natural barrier for wind and animals (Gregerson 1932). In 1865 de Bary proved that the fungus *Aecidium berberidis* that was commonly found on barberry was the same organism that caused stem rust on wheat (De Bary 1866). Although this finding was crucial to understanding the life cycle of *Pgt* the general public did not understand the implications and by 1900 the common barberry became well established in the Midwest (Bills 1968). Along with the establishment of barberry in the late 1800's came the widespread planting of wheat in northern United States and southern Canada. The United States had a growing reliance

on bread wheat and the need to control stem rust became apparent during severe epidemics in 1878, 1892, 1894 and 1904 (Hamilton 1939, Carleton 1899, Freeman and Johnson 1911).

In 1905 the U.S. Department of Agriculture (USDA) began funding research into wheat cultivars with resistance to stem rust (Stakman n.d.). Unfortunately when crossing resistant durum wheat with bread wheat it was found that many stem rust resistance genes were closely linked to durum characteristic genes and highly resistant bread wheat cultivars would be difficult to produce (Stakman n.d). Any resistant bread wheat cultivars that would have been produced would also have to contend with the changing populations of *Pgt* and resistance would likely be short lived. In 1912 a new cultivar of wheat called Marquis, that exhibited early maturation and moderate stem rust resistance was produced in Canada (Stakman n.d.). The following years cultivar Marquis was planted in large quantities and showed signs of being able to protect wheat crops from stem rust. Unfortunately the resistance provided by cultivar Marquis proved to be short lived and a severe stem rust epidemic in 1916 brought about a major loss to wheat production in the United States, where up to 60% of the crop was lost in some Midwestern states such as Minnesota and North Dakota (Stakman and Fletcher 1933). In 1917 the United States became involved in World War I and the concern of food shortages prompted the U.S. government to take action. In 1918 policy makers made the decision that steps needed to be taken to control stem rust and the Barberry Eradication Program was formed as a cooperative effort between state and federal agencies (Roelfs 1982).

At the onset of the eradication program, workers in 13 states strove to remove barberry bushes in towns and cities but soon realized the magnitude of bushes growing on farms (Roelfs 1982). In 1920 farm-to-farm surveys began to be conducted in an attempt to locate and monitor all barberry bushes growing between the Rocky and Appalachian Mountains. Barberry bushes were removed by hand digging, dislodging the plant with the use of a tractor or sometimes even destroyed using dynamite. Resurveys of sites where barberry bushes were removed often showed signs of new bushes growing from

any viable roots that were left in the ground. To restrict new barberry growth the eradication program started applying salt or kerosene to the area around the removed bushes. By the mid 1930's the majority of barberry bushes had been eliminated and an effort was being made to remove the remaining bushes that were growing on mostly uncultivated land. In addition to the removal of susceptible barberry bushes screening newly derived lines of barberry began to be conducted in order to identify hybridizaiton events between susceptible and resistant barberries (Roelfs 1982). The eradication program was continually funded until the late 70's and by 1980 all federal funding was terminated. Presently the majority of the responsibility of continued efforts in screening for susceptible barberry varieties is a burden of the state authorities with some assistance from the USDA Animal and Plant Health Inspection Service (APHIS).

The reduction in aeciospore production in the United States has had a large effect on stem rust epidemiology. The removal of barberry bushes directly adjacent to wheat fields resulted in delaying disease onset by about 10 days and has reduced the initial inoculum by removing the overwintering sites in the northern states (Roelfs 1982). The current inoculum source for the majority of the *Pgt* infections in the United States is from long distance transport of urediniospores from overwintering sites in the southern United States and Mexico (Roelfs 1982). The removal of the alternative host has also restricted stem rust to an asexual lifestyle. Without barberry *Pgt* can't undergo genetic recombination and has resulted in a reduction in the number and stabilization of the races present (Roelfs 1982, Peterson et al. 2005).

The eradication program mainly focused on removing barberry bushes between the Rocky and Appalachian Mountain Ranges because that is where the majority of the wheat is grown in the United States. The presence of barberry bushes on the coastal sides of these mountain ranges still have the ability to allow sexual reproduction in *Pgt* and collections along the Pacific Northwest have had a wider variety of races and races not found in the intermountain range (IMR) region of the United States (Roelfs et al. 1991).

Fluctuations in the prevalence of phenotypic race groups

From 1926-1928 race groups 36 and 21 comprised the majority of the stem rust races collected (Kolmer 2001). The prevalence of these race groups was due to the continued planting of Marquis (Kolmer 2001). The cross between cultivars Marquis and Kota called Ceres was developed in 1926 and showed a high level of resistance to the races found at that time (Kolmer 2001). The removal of cultivar Marquis from the field has resulted in the elimination of races 36 and 21 since the late 1920's. Cultivar Ceres provided moderate protection from stem rust until epidemics in 1935 and 1937 caused by race 56, now know as race group MCC_ (Kolmer 2001). The epidemic of 1935 was so severe that North Dakota, Minnesota and South Dakota lost 56%, 52% and 29% of their wheat crops, respectively (Roelfs 1978). To overcome the MCC_ race group, cultivar Thatcher was developed in the mid 1930's and provided immunity to stem rust untill1950 when race group 15B or T__K became prevalent (Leonard 2001). In 1939 the first isolates of T__K were collected near Fort Dodge, IA (Kolmer 2001). From 1939 until 1949 isolates of race group T__K were found at low levels but by the 1950's this race group caused severe epidemic due to its virulence to *Sr12* found in cultivar Thatcher and its derivatives (Kolmer 2001). The epidemic of 1954 was not as severe as the 1935 epidemic but caused major economic losses to farmers by wiping out 43% of the wheat in North Dakota, 21% in South Dakota, 18% in Minnesota and 15% in Iowa (Roelfs 1978). Great efforts were put into breeding new lines of wheat to overcome T__K and in 1956 the resistant cultivar Selkirk became the dominant wheat planted (Kolmer 2001). Cultivar Selkirk was the solution to the T__K epidemics but allowed race group MCC_ to increase in frequency and again cause losses of up to 15% in some states in 1965 (Roelfs 1978). In 1964 the hard red winter wheat cultivar Triumph was released. Cultivar Triumph contained the gene *SrTmp*, which provides resistance to race group MCC_ but not T__K and allowed race group T__K to again rise in frequency (Roelfs 1978). T__K became the dominant race group of stem rust collected from 1964 until 1998, with the exception of 1990 and 1991 but *Pgt* didn't cause severe epidemics after the 1950's (Kolmer et al. 2007, Leonard 2001).

Since 1955 there has not been a severe epidemic of stem rust except in a few southern states (Leonard 2001). In the southern United States lines derived from the cultivar Arthur were commonly grown because of the durable resistance provided by *Sr*36. In the early 1970's Arthur derived lines Blueboy and McNair 701 were heavily planted because of their high yield. In 1974 areas that were planted with Blueboy and McNair 701 suffered heavy losses because these two new lines were also very susceptible to stem rust. This high susceptibility allowed *Pgt* to overwinter, increase in frequency and infect the new crop much earlier in the growing season than under conditions with a resistant variety of wheat.

In 1989 two new race groups of *Pgt*, QCCJ and QFCS were detected in the central United States (Roelfs et al. 1991). Isolates of race group QCCJ were primarily found on barley and were thought to be specific to *Hordeum* species due to virulence on the widely used resistance gene *Rpg1*. In 1990 race group QCCJ was also found on winter wheat plots in Kansas and eastern Colorado and established itself as a significant member of the rust population until 1996 (Leonard 2001). The sharp decline in the frequency of race group QCCJ between 1996 and 1997 is thought to be a result of a decrease in planting of the susceptible winter wheat in Kansas, which resulted in a reduction of overwintering for this race group (Dill-Macky and Roelfs 1998). Race group QFCS is now the most common race found in North America due to its virulence on several hard red winter wheat and soft red winter wheat cultivars (Kolmer et al. 2007). Isolates of both race groups QCCJ and QFCS had previously been collected in the Pacific Northwest where it is thought that stem rust is still undergoing sexual reproduction (Roelfs et al. 1991). In the IMR region both races were first identified in the northern states of Minnesota, North Dakota and South Dakota indicating that they may have crossed the Rocky Mountains and have now become established in the IMR asexual population instead of their presence being due to convergent evolution of phenotypic markers in mating events or parasexuality (Roelfs et al. 1991, Roelfs et al. 1997).

Recently there has been a renewed interest in *Pgt* due to the discovery of isolates in Uganda that showed virulence to *Sr31* (Pretorius et al. 2000). The resistance gene

Sr31 is on a translocation from rye that became very prevalent in wheat germplasm beginning in the mid 1980s. *Sr31* was not included in the differential set published by Roelfs et al. in 1993 so to accommodate this new race group wheat lines with resistance genes *Sr24*, *Sr31*, *Sr38* and *SrMcN* were added to the standard set of differentials (Jin et al. 2008). Based on the expanded set of differentials the strain of *Pgt* found in Uganda in 1998 (Ug99) was determined to be TTKSK (Jin et al. 2008). Collections made in Kenya in 2006 were found to contain isolates of the Ug99 race group with virulence to *Sr31* and *Sr38* (TTKSK) as well as *Sr31*, *Sr38* and *Sr24* (TTKST) (Jin et al 2008). Field collections from 2007 identified a third variant that contained virulence to *Sr31*, *Sr38* and *Sr36* (TTTSK) (Jin et al. 2009). Other variants have been identified in South Africa that are virulent on *Sr38* but not on *Sr24* or *Sr31* (TTKSF, Uv55) (Boshoff et al. 2002), virulent on *Sr24* and *Sr38* but not *Sr31* (TTKSP, Uv59) (Terefe et al. 2010), virulent on *Sr24*, *Sr31* and *Sr38* but not *Sr21* (PTKST, Uv60) (Pretorius 2010) and virulent on *Sr*31 but not *Sr*21, *Sr*24 or *Sr*36 (PTKSK) (Singh et al. 2011).

The Ug99 race group has rapidly dispersed across eastern Africa and now into the Middle East. It was first identified in Uganda in 1999 (Pretorius et al. 2000) then found in Kenya in 1999, Ethiopia in 2003, Yemen in 2006 (Singh et al. 2006) and Iran in 2007 (Nazari et al. 2009). This trend is particularly alarming due to the high production of wheat in India and China and it is only a matter of time before this race migrates to these countries as well as the United States. The migration pattern seen in the Ug99 race group is similar to what was also seen with a race of the *P. stiiformis* identified in 1986 (Singh et al. 2004). This race of stripe rust was found to be virulent on a commonly used resistant gene *Yr9* and swept through eastern Africa, the Middle East and southern Asia in an 11-year time span causing severe epidemics (Singh et al. 2004).

Genetic analyses

The first examination of the relatedness of the common races of *Pgt* in the US was conducted in 1985 (Burdon and Roelfs 1985a). This study used 12 differentials to separate isolates into nine race groups; RCRS, RHTS, T__K, H_LG, _KCJ, MCC_, R__Q, Q_CQ and _THJ. An analysis of phenotypic similarity showed the RCRS and

RHTS groups to be the most similar and T__K to have the most variation within the group. Thirteen isozyme markers were then screened on the nine races and only six of the isozyme markers were found to be polymorphic and useful to distinguish the race groups. By analyzing the mean isozyme dissimilarity they found the highest similarity between the T__K, H_LG and R__Q groups and the _KCJ, _THJ and RHTS groups. The six isozyme markers were also found to have no variation within a race group.

In an attempt to further examine the relationship between the race groups Kim et al. (1992) showed that size variation in the IGR region of the ribosomal repeat DNA can be used to identify *Pgt* race differences. The authors used one primer pair and nine phenotypic races of *Pgt* and found the region was heterozygous and useful in separating some of the phenotypic race groups. Unfortunately some phenotypic race groups were found to be heterozygous within the group as well as some race groups contained similar sized amplicons. These findings showed that this genetic marker is not sufficient to distinguish all the race groups.

Fox et al. (1995) used the markers designed by Kim et al. (1992) to examine the variation within *Pgt* isolates within the race group QCC. At the time of the Fox et al. (1995) study only 12 differential corresponding to a three letter code were commonly used but there were obvious distinctions between isolates with the race code QCC. In the late 1980's QCC isolates virulent to barley lines containing the gene *Rpg1* began to be collected in the Midwest. Fox et al. found that using IGR repeats they could show the QCC isolates from barley and wheat had different genotypes, with isolates collected off barley having an additional 1.1 kb amplicon. With the expanded set of differentials published in 1993 (Roelfs et al. 1993) the isolates collected off of wheat were determined to be from the Q__Q race group and the isolates collected off of barley became known as phenotypic race QCCJ.

Roelfs et al. (1997) conducted a similar analysis as conducted by Burdon and Roelfs (1985a) but included *Pgt* isolates of the newly described QCCJ and QFCS races in an attempt to determine if these races were a result of a parasexual cross or an introduction/migration event. Roelfs et al. (1997) used 5 polymorphic isozyme markers

and 31 differential host lines to examine relatedness between the 11 race group. Genotypically the QFCS isolates were found to have a high dissimilarity to the other races but QCCJ only had 1 isozyme difference from the commonly found Q__Q race. Although the genotype data were collected, no hypothesis was given as to why QCCJ was so closely related the Q__Q with respect to genotype and instead the authors focused on phenotypic data. The authors found that QCCJ has at least 7 virulence gene differences and QFCS has at least 6 virulence gene differences from any other common NA race. These phenotypic differences values are approximately equal to the number of differences found when the other races were compared to one another; however, no combination of virulence genes found in the other races would yield the virulence pattern found in QCCJ.

To examine the genetic and phenotypic relatedness as well as the possibility of migration between isolates collected in North America, South America and Europe, McCallum et al. (1999) used RAPD markers, isozyme markers and phenotypic reactions to known virulence genes. The authors use phylogenetic trees constructed from each type of data to test the origins and look for migrants in each of the populations. Analysis of the phylogenetic trees could yield two possible outcomes; (i) if isolates from one country were found to have high bootstrap support with isolates from a different country it would suggest that there have been recent introductions from one country to another (ii) if isolates from a country were found to group with isolates of the same country it would suggest that there has been a radiation of isolates within a country. The 1 supported branch in the phylogenetic tree constructed from isozyme data and 10 of the 11 supported branches in the phylogenetic tree constructed from RAPD data are showing support between isolate within the same country, which could indicate recent radiations of the isolates. As a whole the phylogenetic trees had very few supported branches indicating either that these isolates are very closely related or the markers used are not phylogenetically informative enough for the analysis. In order to get a better understanding of migration patterns and possible introduction events additional isolates

need to be collected and additional markers need to be developed in an attempt to gain better phylogenetic support to test these hypotheses.

Effects of genetic recombination on population structure

In an attempt to describe the evolutionary processes acting on the phenotypic virulence markers of *Pgt*, Roelfs and Groth (1980) compared isolates from a sexual population to isolates from asexual populations in North America. The authors were testing the hypothesis that sexual populations would have an accelerated evolutionary adaptation and would produce more variable offspring, some having a higher fitness level than what is found in an asexual population (Smith 1968). By examining 16 phenotypic virulence gene reactions the sexual population was found to have more distinct phenotypes and to be more diverse than the asexual. However the asexual population was found to have a higher mean number of loci expressing virulence per isolate and none of the loci were found to deviate strongly from expected mean frequencies therefore providing no evidence for selection. The higher mean number of virulent loci in the asexual population could be a result of the differentials used in the study. The differential set used was chosen based on the ability of the *Sr* genes in the wheat lines to differentiate the different groups within the asexual population so a bias toward asexual virulence may have been present.

To examine the genotypic difference between isolates of the different modes of reproduction Burdon and Roelfs (1985b) used a subset of the isozyme markers that were developed in the asexual genotyping analysis (Burdon and Roelfs 1985a) to describe the diversity found between asexual and sexual populations of *Pgt* isolates collected in 1972, 1975 (the isolates included in the Roelfs and Groth 1980 paper) and 1976. The authors found that isolates in the asexual population had an average of 2.73 heterozygous loci, none of the isolates were completely heterozygous or homozygous and fixation index calculations significantly deviated from 0. In contrast the sexual population was normally distributed and included isolates that were completely heterozygous and homozygous, the fixation index calculations never deviated from 0 and was found to have greater diversity.

Chapter 1

Genetic Structure of the North American Asexual Populations of the Wheat Stem Rust Pathogen

Introduction

Wheat stem rust caused by *Pgt* historically has been the most devastating pathogen of wheat (Leonard 2001) and caused severe epidemics in the United States from the late 1880s through the 1950s. The occurrence of *Pgt* has been drastically reduced in the last half of the 20th century because of a sustained effort on the deployment of wheat varieties with multiple effective stem rust resistance genes and early maturing dwarf wheat varieties, combined with the earlier eradication of the alternate host, common barberry, between the Rocky Mountains and the Appalachian Mountains. *Pgt* is a macrocyclic, heteroecious rust fungus with the sexual stage occurring on *Berberis* spp. (barberry) and *Mahonia* spp. and the asexual stage on wheat and barely as well as, wild grasses (Leonard and Szabo, 2005). The sexual part of the life cycle is the primary stage in which genetic recombination occurs. As a result of the eradication of common barberry, race changes in the population of *Pgt* in the IMR region are most likely driven by human mediated selection due to changes in wheat varieties grown or new introductions of the pathogen.

The United States IMR populations of *Pgt* have been divided into 11 phenotypic race groups (Roelfs et al. 1997). Although two of the 11 phenotypic race groups currently contain only a single race they are referred to as a race group for simplicity and represent distinct races from the other nine race groups. The predominant groups found in the IMR region of the United States since the 1930's are RCRS, RHTS, Q_CQ, QCCJ, QFCS, MCC_, _THJ, _KCJ, H_LG, R_ _Q, and T_ _K (Roelfs et al. 1997). Race groups MCC_ and T_ _K have historically been of major interest due to the severe epidemics they caused in the early 1930's and 1950's, respectively, with upwards of 56% wheat loss in the effected states (Roelfs 1978). Race groups QCCJ and QFCS have only recently been found within the IMR populations of *Pgt*, having their first identification in 1990 (Roelfs et al. 1993). Race group QFCS has become the dominant race within the IMR populations (Jin 2005, Long et al. 2005, Long et al. 2007, Long et al. 2008) over the last decade. The remaining seven race groups described in this study were found at lower frequencies over the past 80 years (Figure 1.1).

Pgt isolates within a phenotypic race group could be a result of clonal propagation yielding populations with very similar genetic and phenotypic makeup or convergent evolution of phenotype due to the cultural practice of planting varieties of wheat that contain particular virulence genes. Studies by Burdon and Roelfs (1985a) and Roelfs et al. (1997) used five to six polymorphic isozyme markers to investigate the phenotypic relationship between isolates within race groups. Isolates within a race or race group contained identical isozyme pattern consistent with asexual clonal population. The number of isozyme allelic differences between race groups ranged from one to five and therefore was unable to resolve relationships between race groups.

To examine population structure within *Pgt* other studies have used random amplified polymorphic DNA (RAPD) markers (McCallum et al. 1999), amplified fragment length polymorphism (AFLP) markers (McCallum et al. 1999, Keiper et al. 2003) and length variation within the repeat units of the intergenic spacer region of the ribosomal DNA (Kim et al 1992, Fox et al. 1995). The use of dominant markers, such as RAPDs and AFLPs have been shown to be less effective in identifying population structure in diploid or dikarytic organisms than co-dominant markers (Selkoe and Toonen 2006). Co-dominant markers such as simple sequence repeat markers (SSR) have been shown to occur at high frequency (Munkacsi et al. 2006, Selkoe K A and Toonen 2006) and to be useful in identifying within species population structure in multiple fungi (Munkacsi et al 2008, Barnes et al 2001). Di- and tri-nucleotide repeat SSR markers have been developed for *Pgt* and have been shown to be useful in distinguishing between race groups (Szabo et al. 2007, Jin et al. 2009, Zhong et al. 2009).

The objectives of this study are; (i) use SSR markers to verify earlier findings that the race groups are clonal lineages, (ii) determine the validity of the current race group boundaries and (iii) examine the relationships between the common race groups of *Pgt* that have been found in the United States since barberry eradication. An understanding of how the common races are related may lead to insight into how virulence genes in the populations evolved in response to human mediated selection via resistant wheat

varieties, may assist further research into avirulence gene identification and sets up a baseline for future monitoring of *Pgt* race dynamics in the United States.

Materials and methods

Pgt isolates and DNA extractions.

A sample set of 62 *Pgt* isolates was selected to represent the eleven common phenotypic races or race groups that have made up the intermountain range populations over a 55-year time span as described by Burdon and Roelfs (1985a) and Roelfs et al. (1997). A summary of the isolates used in this study is listed in Table 1.1. Wherever possible, multiple isolates of each race were included to give a representative sample of isolates across both geographic and temporal scales. The *Pgt* isolates were originally collected and race-typed according to the International System of Nomenclature (Roelfs and Martens 1988) during annual stem rust surveys conducted by researchers associated with the Cereal Disease Laboratory in St. Paul, MN between 1956-2006, and held in long-term storage in liquid nitrogen at the Cereal Disease Laboratory, USDA-ARS, St. Paul, MN. This set of isolates was not evaluated against the resistance genes *Sr24, Sr31, Sr38* and *SrMcN* making up the fifth set used in current North American wheat stem rust differentials (Jin et al. 2008). Genomic DNA was extracted from roughly 40 mg of tissue from germinated or ungerminated urediniospores. Spore material was pulverized by shaking samples in tubes with a a mixture of lysing matrix C beads (MP Biomedicals, Solon OH) and roughly 5 mg of Diatomaceous Earth (Sigma, St. Louis MO) in a FastPrep shaker (MP Biomedicals Solon OH) for 20 sec on setting 4.0 m/s. This step was repeated once. DNA was isolated from the pulverized tissue using the OmniPrep DNA Extraction kit (GenoTech, St. Louis MO) according to the manufacturer's protocol.

Phenotypic analysis.

In order to verify the race designations for this sample set, at least one isolate from each race or race group was removed from long-term storage to verify the race codes assigned at the time of original storage in liquid nitrogen. Isolates were revived by heat shocking at 45^{o}C for 45s, then re-hydrated in an 80% relative humidity chamber for

at least 1h (Roelfs et al. 1997). Revived urediniospores were spray inoculated in a suspension of Soltrol oil onto a standard set of 16 wheat differentials (Roelfs et al. 1993) to assess virulence phenotype, and onto either the wheat cultivar McNair 701 or Line E to increase spore quantities for DNA extractions. Inoculations were performed on approximately 8-day-old wheat seedlings, allowed to dry on plants for 30min, then incubated overnight in a misting chamber. After drying, plants were transferred to a greenhouse and approximately two weeks after inoculation, race phenotypes were determined (Roelfs and Martens 1998). Dr. Yue Jin performed all phenotyping of isolates.

Genotypic analysis.

Ninety-nine trinucleotide SSR markers were screened to determine a suitable set for genotypic analysis. Eleven *Pgt* isolates were chosen for the screen test, which represented seven of the 11 race groups and three additional reference isolates. Twenty of the markers were selected for analysis (Table 1.2), based on; (i) polymorphisms between isolates, (ii) the presence of no more than 2 amplicons per isolate, (iii) verification that each set of primers amplified only one region in the genome sequence assembly of *Pgt* isolate CRL 75-36-700-3 using a BLAST analysis, (iv) a repeat region could be indentified within the amplified region of the reference genome. The 20 selected SSR markers are expected to be unlinked from one another, as they are distributed across 18 different supercontigs in the reference genome, and the two SSR loci found on the same supercontig are separated by approximately 260 kb. Twelve of the SSR markers were originally designed from expression sequence tags (Zhong et al. 2009) and eight markers were developed from an SSR enriched genomic library constructed by Vizon SciTech (Vancouver, British Columbia) (Jin et al. 2009).

PCRs were performed in a PTC-200 (MJ Research, Waltham MA) thermocycler as follows: 10 µL reactions containing 1X Phusion HF buffer with 1.5 mM $MgCl_2$ (New England BioLabs, Ipswich MA), 0.2 mM dNTPs (Roche, Indianapolis IN), 0.005% casein (Sigma, St. Louis MO), 1 µM IRD700/800 labeled forward primer (LI-COR, Lincoln NE), 1 µM reverse primer, 0.1 U Phusion Taq polymerase (New England

BioLabs, Ipswich MA), and 20ng DNA, with a cycle of: 98 °C for 5 min, 30 cycles at 98 °C for 30 s, 62 °C for 30 s, 72 °C for 30 s and a final extension at 72 °C for 10 min. PCR products were diluted 1:10 in ddH_2O, 5 µL of Stop Buffer (31.5 ml formamide, 1.32 ml 0.5 M EDTA, 330 µl dH_2O and 48 mg bromophenol blue) was added, and samples were denatured for 3 min prior to being analyzed on a LI-COR 4300 DNA sequencer (LI-COR, Lincoln NE) using a 7% polyacrylamide gel. Amplicon sizes were determined using a DNA size standard ranging from 50 to 350 bp (LI-COR, Lincoln NE) and the SAGAGT software (LI-COR, Lincoln NE). When examining gels, amplicons within a lane were scored if they exhibited similar amplification intensity as another amplicon in the same lane; amplicons of less intensity were considered bleed over from adjacent lanes, stutter bands or amplification from a secondary locus and ambiguous results were tested at least twice to increase accuracy. *Pgt* isolate CRL 75-36-700-3 was included on each gel as a standard and amplicons were scored as number of bases, then converted to a repeat number using the predicted allele size based on the *Pgt* genome sequence as a size standard. Reactions that did not produce any products or produced more than two amplicons were recorded as missing data.

Data analyses.

For each SSR locus, the total number of alleles, number of alleles private to a race group, number of rare alleles per race group (alleles found at a frequency of 5% or less), the number of effective alleles, observed and expected heterozygosity and the fixation index were calculated using GenAlEx (Peakall and Smouse 2006). For phenotypic virulence data, a binary matrix was constructed by scoring resistant interactions as 0 and susceptible interactions as 1. Three-dimensional principal coordinate analyses (PCA) were conducted using GenAlEx 6, three-dimensional plots were constructed using Delta Graph (Red Rock Sofware, Salt Lake City UT) to assess clustering of individuals within race groups for both phenotypic and genotypic data. Bootstrap support was assessed by constructing 1000 neighbor-joining (NJ) trees from a frequency matrix based on shared alleles in the program PowerMarker V3.25 (Liu and Muse 2005). Consensus trees were constructed from the 1000 trees using Phylip 3.67 (Felsenstein 1989) and used to

calculate the bootstrap support; trees were viewed using FigTree (http://tree .bio.ed.ac.uk/software/figtree/) and are displayed as midpoint rooted. In the above programs missing data were not coded as an additional state for analyses.

The number of phenotypic and genotypic differences within a race group and between race groups was calculated by counting the number of virulent phenotypic infection types or the number of alleles found in only one of two race groups in the pair wise comparisons. Calculations of total numbers of genotypes and allele differences were conducted assuming missing data points contained identical amplicons as the other isolates in the race group. The level of genotypic (Rst) and phenotypic differentiation (ΦPT) was calculated over 999 permutations of the data set using AMOVA in GenAlEx6. To test data congruence, the genotypic Rst data was compared with the phenotypic ΦPT analyses using a Mantel test was performed in GenAlex6.

Results

Genetic variation and cluster analyses.

A summary of genetic variability for each of the SSR loci is presented in Table 1.3. From the 20 SSR loci examined, a total of 76 alleles were amplified, with 2-7 alleles per locus and a low occurrence of rare alleles, with an average of 0.6 per locus. In spite of repeated attempts to produce amplicons for all *Pgt* isolates, on average, 1.8% of the data were scored as missing for each locus. Locus Pgest341 and PgtCAA39 had the highest amount of missing data with 6% and 8%, respectively. No significant correlation was found between level of missing data and race group.

Neighbor-joining (NJ) analysis of the SSR data matrix assigned the 62 *Pgt* isolates into nine well support genetic groups (Figure 1.2) and is consistent with historic phenotypic race groups with the exception of combining four previously separated race groups into two genetic groups. Isolates from the _THJ and _KCJ race groups were genetically very similar and formed a single, well-supported (99.8%) clade and were placed in a single genetic group (NA-006). Similarly, isolates from the race groups RCRS and RHTS were found to have high genetic similarity and formed a single, well-

supported (100%) clade and was placed in a single genetic group (NA-008). The phylogenetic analysis further identified three genetic clusters (GC) containing seven of the nine genetic groups. All isolates from the genetic groups NA-001 (T_ _K), NA-002 (H_LG) and NA-003 (R_ _Q) formed GC I (100% support), isolates from genetic groups NA-004 (QCCJ) and NA-005 (QFCS) formed GC II (85.5% support), and isolates from genetic groups NA-006 (_THJ, _KCJ) and NA-007 (MCC_) formed a GC III (79.8% support). The frequency matrix used to generate the genotypic neighbor-joining tree contained 76 characters of which 70 were informative. Terminal structure within the genetic groups has low support and is mainly due to missing data. Principal coordinate (PC) analysis was conducted to validate clustering patterns found in the NJ analysis (Figure 1.3). The first three dimensions accounted for 73.23% of the variation in the data. Clustering patterns were consistent between these two analyses. In addition, PC analysis supported the clustering NA-008 and NA-009, which was not supported in NJ analysis.

Estimates of genetic variability within genetic groups and genetic clusters are summarized in Table 1.4. Sixteen unique genotypes were observed from the 62 *Pgt* isolates, with five genetic groups comprised of multiple genotypes. Genetic groups NA-001, NA-006 and NA-009 each contained an isolate with a single allele difference, NA-003 contains two isolates that have a single allele difference at different loci, and race group NA-008 contains an isolate with two allele differences and two other isolates (one of race RCRS and the other RHTS) with an identical allele difference at a third locus. The average observed heterozygosity over all 20 loci for each genetic group was significantly greater than expected, with fixation index estimates indicating an excess of heterozygosity (-0.89- -1.00). Private alleles were found in genetic groups NA-004, NA-007, and each of the genetic clusters.

Pair wise comparison was used to determine the number of allelic differences within and between nine genetic groups (Table 1.5). Genetic groups NA-004 and NA-005 have the lowest number of allele differences with 14, race groups NA-001, NA-002 and NA-003 have 16 or 17 differences and NA-006 has 16 differences from NA-007. To examine the level of genetic differentiation between race groups Rst values were

calculated and significant levels of genetic differentiation were found between all the race groups except when comparing NA-001 to NA-002, NA-004 to NA-005, NA-006 and NA-007 and NA-007 to NA-005 and NA-009 (Table 1.6).

Race phenotypes and cluster analyses.

A subset of the isolates used in the study weas reevaluated on wheat differentials to validate the avirulence/virulence phenotypes assigned. All assigned phenotypes were found to be consistent with identifications made at the time the isolates were originally placed in long-term storage. A total of 21 different phenotypes were observed from the 62 *Pgt* isolates, representing all of the 11 race groups identified by Roelfs et al (1997). Seven of the race groups exhibited multiple phenotypes (Table 1.7).

A NJ analysis of the avirulence/ virulence phenotypes of the 62 *Pgt* isolates was performed using a binary matrix. The analysis could not separate the races into supported race groups due to low bootstrap support (Figure 1.4). The frequency matrix used to build the trees yielded 31 characters, 30 of which were informative. High levels of support were found for all isolates in the QTHJ isolates (93%), the 2 QCCQ isolates (92.6%), the 2 RHTS isolates (90.5%), the 2 TMLK isolates (88.3%), all isolates of MCC_ race group (83.9%) and the 3 TPMK isolates (78.1%), while the rest of the tree was unresolved. The low level of support for the majority of the branches on the tree is likely due to the low number of characters in the study. A three-dimensional PC plot visualizing the avirulence phenotypic data is shown in Figure 1.5. The first three dimensions in the PC analysis accounted for 71.63% of the variation.

To test for differentiation between race groups a ΦPT analysis was performed on phenotypic avirulence/virulence data and significant levels of differentiation were found between all race groups except when comparing QCCJ and QFCS (Table 1.8). A mantel test comparing Rst and ΦPT values for each race group yielded an R^2 value of 0.0096, showing the genotypic and phenotypic data were not congruent.

Discussion

Over the last 80 years, the number and diversity of phenotypic races of *Pgt* in the IMR region North America has declined. Currently the IMR region is dominated by a single race (QFCS) (Jin, 05). This dramatic change in the population of *Pgt* is primarily due: i) to the elimination of the sexual reproduction by eradication of the common barberry, the alternate host, ii) the use of wheat cultivars with multiple effective stem rust resistance genes, and (iii) a genetic bottleneck due to survival of the asexual urediniospores in a small region of the United States along the Gulf Coast. Analysis of the population structure of *Pgt* is currently based on phenotypic race data and very little is known about the genetic structure of this population. In order to better understand the genetic population structure, SSR analysis of 62 *Pgt* isolates representing the eleven most predominant race groups found in the United States over the last half century was performed. Genetic analysis determined that the 11 race groups are composed of nine well-supported genetic groups and confirmed that the population structure is due to asexual clonal propagation of race groups rather than convergent evolution.

The 9 supported genetic groups identified in this study are R_ _S, Q_CQ, QCCJ, QFCS, MCC_, _THJ/_KCJ, H_LG, R_ _Q, and T_ _K. The difference between the 11 phenotypic race groups previously defined (Roelfs et al, 1997) and the nine SSR genetic groups is the placement of the four race groups RCRS, RHTS, _THJ and _KCJ. In the original race nomenclature system, RCRS and RHTS were part of race 11 (Stakman et al, 1962). Based on differences in avirulence/virulence phenotypes on *Sr6* and *Sr30* and four isozyme markers, Burdon and Roelfs (1985a) separated RCRS and RHTS (Table 1.9). However, the SSR analysis indicates that these two race groups belong to a single genetic group NA-008 (Figure 1.2). The five isolates examined in this study representing RCRS and RHTS races have very similar genotypes with only three allele differences, two of the allele differences are in one isolate and the third is not race-specific (Table 1.5). These isolates were collected over a 22-year span (1956-1977) and across central U.S. from Louisiana to North Dakota. Genetic group NA-008 has a high level of support

(100%) in phylogenetic analysis (Figure 1.2) and formed a single cluster in PCA (Figure 1.3). Therefore, we propose that races RCRS and RHTS be placed in a single phenotypic race group designated R__S (Table 1.7). The nine *Pgt* isolates representing race groups _THJ and _KCJ used in this study are genetically very similar with only a single allelic difference across the 20 SSR loci examined (Table 1.5). These isolates belong to a single, well supported (99.8%), genetic group (NA-006) in the phylogenetic analysis (Figure 1.2) and a single cluster in the PCA analysis (Figure 1.3). The nine representative isolates were collected over a span of 18 years (1959-1976) and from six states spanning the US. Race groups _KCJ and _THJ represent races 29 and 48, respectively, in the original Stakman race designation (Stakman et al., 1962). As the wheat differential set evolved - race group 29 became 29-32 (Burdon and Roelfs, 1985a) and 48 became 151-32 (Burdon and Roelfs, 1985a). In combining race groups _THJ and _KCJ, convention dictates to keep only the common letters and use an "underscore" for registers where the letters vary. In this case only the last letter "J" is in common. However, a race group code with only a single letter code can be misleading, therefore this race group is represented as - _THJ/_KCJ.

Overall, the SSR data are consistent with the isozyme data of Burdon and Roelfs (1985a). Both datasets indicate that the genetic structure of the *Pgt* population in the intermountain range region of North America represents asexual reproduction within the dominant race groups. The resolving power of the isozyme dataset was limited, due to only five alleles being polymorphic. No variation in isozyme pattern was observed within the race groups and a range of one to five allele differences were found between race groups. Isozyme data suggested that race groups R_ _Q, H_LG and T_ _K are genetically closely related (differ by a single allele) and this is confirmed with SSR data showing that NA001, NA-002 and NA-003 are genetically related and make up GC I. Similarly, only a single isozyme allelic difference is observed between isolates in the race groups MCC_ and _KCJ, which corresponds to NA-006 and NA-007 of GC III.

In order to further examine the relationships between the genetic race groups Burdon and Roelfs (1985a) calculated the number of allele differences between the

groups. The lowest mean isozyme allele dissimilarity values were found between race groups _THJ and _KCJ, _THJ and RHTS, _KCJ and MCC_, T_ _K and H_LG, T_ _K and R_ _Q and R_ _Q and H_LG corresponding roughly to GC I and GC III+RHTS. The examination of these race groups using SSR data show low allele difference values for the all pair wise comparisons in Burdon and Roelfs except between race groups _THJ/_KCJ (NA-006) and RCRS/RHTS (NA-008). Further, the NJ analysis shows high phylogenetic support for GC I and GC III race group cluster but the RHTS isolates are not found to be within the GC III group. In the phylogenetic analysis NA-008 has lower bootstrap support than the 75% cutoff so was not assigned to a group cluster with any other race group but the principal component analysis performed on the SSR data indicates an association between the NA-008 and the NA-009. This finding is in contrast to the findings in the isozyme analysis, which found the RCRS and RHTS race groups were the most dissimilar to the Q_CQ race group.

Differentiation measures based on Rst calculations identify six pair wise comparisons that did not show significant differentiation (Table 1.6). The lack of differentiation when comparing NA-004 with NA-005 and NA-001 with NA-002 is also found in the NJ analysis and reinforces the idea of recent divergences within GC I and GC II. The relationships found by the Rst analysis when comparing NA-007 to NA-004, NA-005 and NA-009 and NA-004 to NA-008 are not found with any other analysis and may be due to the low number of private alleles in the NA-007 and NA-009 populations.

Further illustration of the recent divergence of GC I can bee seen when examining the number of private alleles per group. Many of the GCs contain more private alleles than the sum of the groups that are contained within. For example NA-006 and NA-007 have a total of three private alleles but the number of private alleles increase to four when analyzed as GC III. The two race groups in GC II have a combined total of four alleles not found with other groups but as GC II they have six. The three race groups found in GC I don't contain any alleles that are private to each of the race groups; but GC I contains nine alleles that are not found in any other isolates. The large increase in the number of private alleles in GC I suggests a more recent divergence from a common

ancestor than the members of the other race clusters. The more recent divergence of these three groups could be due to repeated introductions from an outside source, sexual recombination or both.

The fixation index for each of the nine genetic groups is close to -1 indicating higher than expected levels of heterozygosity within each race group. This high level of heterozygosity may be due to neutral forces or a heterozygote advantage. Asexual isolates of *P. graminis* proliferate as urediniospores and are dikayrotic, the two haploid nuclei would be expected to be under independent neutral drift and may have drifted toward different alleles. Alternatively there may be a selective force acting on these isolates that gives heterozygous individuals an advantage. Asexual populations are thought to have a disadvantage to sexual populations in that they have a diminished ability to deal with environmental variability due to the lack of mixing of genes during cross over events. Peck and Waxman (2000) showed that if there is a heterozygosity advantage in a population, asexual isolates have a higher fitness than sexual isolates because of the inability of sexual isolates to maintain high levels of heterozygosity due to crossing over in sexual propagation. The eradication of barberry in the IMR region has eliminated sexual reproduction in the *Pgt* populations and resulted in an asexual clonal population structure. It is unclear if the current asexual populations were derived from the sexual population in the United States, co-existed with the sexual populations in the early 1900's or a combination of the two.

Race groups QCCJ and QFCS were proposed to have originated in the Pacific Northwest due to both race groups being collected in the region prior to being found in the Midwest (Roelfs et al. 1990, Roelfs et al. 1991) and the first occurrences of these races was in the northern plain states of the U.S. (Roelfs et al. 1991). In 1997 Roelfs et al. used a set of five isozyme markers and 27 resistance genes to examine the relationship between the QCCJ, QFCS and the previously collected race groups in the Midwest. The isozyme markers showed the isolates of the QCCJ and Q_CQ race group to be closely related with only one allele difference while the QCCJ and QFCS race groups had three differences. The association between QCCJ and Q_CQ is not found in this study. The

genotypic data show high similarity between the NA-004 (QCCJ) and NA-005 (QFCS) race groups as indicated by high support on a phylogenetic tree, clustering on a PC plot, low number of allele differences and non-significant Rst values. Race group NA-009 (Q_CQ) is a distinct genetic group not closely related to the NA-004 (QCCJ) or NA-005 (QFCS) race groups. The highly supported relationship between the NA-004 (QCCJ) and NA-005 (QFCS) race groups found in this study indicate that these two race groups came from a common source. It is possible that the common source is located in the Pacific Northwest of the United States but further studies need to be conducted to test this theory.

The 16 avirulence/virulence markers examined in this study produced six supported branches on a phylogenetic tree all of which were either within a race group or supported the differentiation of a particular race group but no associations between race groups could be identified (Figure 1.4). Clustering of isolates in a principal component analysis (Figure 1.5) was found to be similar to a previously published avirulence/virulence analyses (Roelfs et al. 1997). However, in the current study isolates that made up the genotypic GC I and GC II appear to cluster within their GCs in the phenotypic analyses but a statistical test could not be preformed on this observation. A ΦPT calculation found QCCJ and QFCS to not be significantly different while the remaining phenotypic race groups were significantly different from each other (Table 1.6). A mantel test comparing the genotypic and phenotypic data found the two types of data to be incongruent, demonstrating the importance of understanding the genetic relationship between these race groups and showing that care needs to be given when assuming ancestral relationships based on phenotypic markers.

This study further illustrates the usefulness of SSR markers for intra species examination of *Pgt* isolates. The focus of this paper was the relationships between the isolates and race groups that have made up the intermountain range populations of *Pgt* found in North America over the past 80 years and the findings of this study illustrate the need for further examination into some of race group's relationships. Isolates of race groups QCCJ and QFCS were first detected in the Midwest in the late 1980's and were speculated at being introductions from the Pacific Northwest (Roelfs et al. 1991, Roelfs

et al. 1997). The SSR markers used in this study show that isolates of the phenotype QCCJ and QFCS form the sister clades NA-004 and NA-005, which have a high level of support as being different from the other race groups. A comparison of the genetic make up of isolates collected in the Pacific Northwest and the isolates in this study would shed light on the possibility of the NA-004 (QCCJ) and NA-005 (QFCS) race groups being recent migrants that crossed the Rocky Mountains. The NA-001 (R_ _Q), NA-002 (H_LG) and NA-003 (T_ _K) isolates also form an unexpectedly strongly supported group of isolates phenotypically as well as genotypically, with the only well supported clade made up of more than two race groups. These three race groups also shared a high number of shared private alleles and low level of allele differentiation suggesting a more recent radiation than the other races. The races that make up the T_ _K race group have been important agronomic races of *Pgt* due to their severity in the 1950's and further investigations into the relationship between these three race groups should be undertaken. With the sudden threat to global wheat production from the Ug99 race group an understanding of the population structure of *Pgt* populations around the world will be important in developing monitoring programs and controlling this pathogen in the future (Singh et al. 2006).

Table 1.1 Summary of *Puccinia graminis* f. sp. *tritici* isolates used in the intermountain range study.

Race group[a]	Race	Isolate	State	Year	Virulence (*Sr* genes)	Avirulence (*Sr* genes)	Genotype
H_LG	HFLG	60-SD-19A	SD	1960	7b, 8a, 9d, 9g, 21, 36	5, 6, 9a, 9b, 9e, 10, 11, 17, 30, Tmp	NA-002a
	HFLG	62-TX-13A	TX	1962			NA-002a
	HFLG	72-MN-1173-3	MN	1972			NA-002a
	HFLG	80-SD-512C	SD	1980			NA-002a
	HPLG	68-TX-73A	TX	1968	7b, 8a, 9d, 9g, 11, 21, 36	5, 6, 9a, 9b, 9e, 10, 17, 30, Tmp	NA-002a
MCC_	MCCD	74-PA-751	PA	1974	5, 7b, 9g, 10, 17	6, 8a, 9a, 9b, 9d, 9e, 11, 17, 21, 30, 36, Tmp	NA-007a
	MCCD	75-WA-974B	WA	1975			NA-007a
	MCCD	75-WV-510A	WV	1975			NA-007a
	MCCF	59-KS-19	KS	1959	5, 7b, 9g, 10, 17, Tmp	6, 8a, 9a, 9b, 9d, 9e, 11, 17, 21, 30, 36	NA-007a
	MCCF	60-ND-55C	ND	1960			NA-007a
	MCCF	76-MN-568B	MN	1976			NA-007a
	MCCF	76-OK-513A	OK	1976			NA-007a
Q_CQ	QCCQ	59-OH-5B	OH	1959	5, 9a, 9d, 9g, 17, 21	6, 7b, 8a, 9b, 9e, 10, 11, 30, 36, Tmp	NA-009a
	QCCQ	03-LA-94C	LA	2003			NA-009a
	QFCQ	75-MN-28C	MN	1975	5, 8a 9a, 9d, 9g, 17, 21	6, 7b, 9b, 9e, 10, 11, 30, 36, Tmp	NA-009b
	QFCQ	81-SD-744B	SD	1981			NA-009a
	QFCQ	93-WI-225C	WI	1993			NA-009a
QCCJ	QCCJ	89-SD-374B	SD	1989	5, 9d, 9g, 10, 17, 21	6, 7b, 8a, 9a, 9b, 9e, 11, 30, 36, Tmp	NA-004a
	QCCJ	91-IA-154-1	IA	1991			NA-004a
	QCCJ	91-LA-164	LA	1991			NA-004a
	QCCJ	91-SD-191-3	SD	1991			NA-004a
	QCCJ	96-UT-116-2	UT	1996			NA-004a
	QCCJ	98-OR-76	OR	1998			NA-004a
QFCS	QFCS	89-OR-436B	OR	1989	5, 8a, 9a, 9d, 9g, 10, 17, 21	6, 7b, 9b, 9e, 11, 30, 36, Tmp	NA-005a
	QFCS	91-KS-60A	KS	1991			NA-005a
	QFCS	91-MT-506C	MT	1991			NA-005a
	QFCS	93-NY-224B	NY	1993			NA-005a
	QFCS	98-MN-313C	MN	1998			NA-005a
	QFCS	06-ND-76C	ND	2006			NA-005a

[a]Race groups as defined by Roelfs et al., 1997.

Table 1.1 continued.

Race group[a]	Race	Isolate	State	Year	Virulence (*Sr* genes)	Avirulence (*Sr* genes)	Genotype
_THJ	QTHJ	68-TX-64E	TX	1968	5, 6, 8a, 9b, 9d, 9g, 10, 11, 17, 21	7b, 9a, 9e, 30, 36, Tmp	NA-006a
	QTHJ	69-MN-399	MN	1969			NA-006a
	QTHJ	72-VA-703C	VA	1972			NA-006a
	QTHJ	75-MN-717C	ND	1975			NA-006a
	QTHJ	76-TX-10B	TX	1976			NA-006a
	QTHJ	76-WA-1201	WA	1976			NA-006a
_KCJ	HKCJ	59-TX-45B	TX	1959	6, 7b, 8a, 9d, 9g, 10, 17, 21	5, 9a, 9b, 9e, 11,30. 36. Tmp	NA-006b
	RKCJ	70-MN-523C	MN	1970	5, 6, 7b, 8a, 9d, 9g, 10, 17, 21	9a, 9b, 9e, 11, 30, 36, Tmp	NA-006a
	RKCJ	74-KS-479C	KS	1974			NA-006a
RCRS	RCRS	56-SD-37B	SD	1956	5, 7b, 9a, 9b, 9d, 9g, 10, 17, 21, 36	6, 8a, 9e, 11, 30, Tmp	NA-008a
	RCRS	74-LA-69C	LA	1974			NA-008b
	RCRS	77-ND-82A	ND	1977			NA-008c
RHTS	RHTS	68-MO-192-1	MO	1968	5, 6, 7b, 9a, 9b, 9d, 9g, 10, 17, 21, 30, 36	8a, 9e, 11, Tmp	NA-008a
	RHTS	69-SD-657C	SD	1969			NA-008b
R_ _Q	RFQQ	61-PA-80A	PA	1961	5, 7b, 8a, 9a, 9b, 9d, 9g, 21, 36	6, 9e, 10, 11, 17, 30, Tmp	NA-001a
	RKQQ	75-MN-10C	MN	1975	5, 6, 7b, 8a, 9a, 9b, 9d, 9g, 21, 36	9e, 10, 11, 17, 30, Tmp	NA-001a
	RKQQ	99-KS-76A-1	KS	1999			NA-001a
	RKRQ	56-IN-702C	IN	1956	5, 6, 7b, 8a, 9a, 9b, 9d, 9g, 17, 21, 36	9e, 10, 11, 30, Tmp	NA-001a
	RKRQ	76-ND-744C	ND	1976			NA-001a
	RKRQ	80-MN-518-3	MN	1980			NA-001a
	RPQQ	61-IN-262C	IN	1961	5, 7b, 8a, 9a, 9b, 9d, 9g, 11, 21, 36	6, 9e, 10, 17, 30, Tmp	NA-001a
	RPQQ	69-CA-495B	CA	1969			NA-001b
	RPQQ	72-KS-504C	KS	1972			NA-001a
T_ _K	TCLK	62-ND-27C	ND	1962	5, 7b, 9d, 9e, 9g, 10, 21, 36, Tmp	6, 8a, 9a, 9b, 11, 17, 30	NA-003a
	TCLK	73-TX-187A	TX	1973			NA-003a
	TCLK	74-KS-264A	KS	1974			NA-003a
	TCLK	74-MN-705C	MN	1974			NA-003b
	TFLK	75-OK-259A	OK	1975	5, 7b, 8a, 9d, 9e, 9g, 10, 21, 36, Tmp	6, 9a, 9b, 11, 17, 30	NA-003c
	TMLK	61-IA-21C	IA	1961	5, 7b, 9d, 9e, 9g, 10, 11, 21, 36, Tmp	6, 8a, 9a ,9b, 30, 17	NA-003a
	TMLK	76-MN-1391	MN	1976			NA-003a
	TPMK	74-MN-1409A	MN	1974	5, 7b, 8a, 9d, 9e, 9g, 10, 11, 17, 21, 36, Tmp	6, 9a, 9b, 30	NA-003d
	TPMK	80-MN-633B	MN	1980			NA-003d
	TPMK	89-GA-96C	GA	1989			NA-003a

[a]Race groups as defined by Roelfs et al., 1997.

Table 1.2 Primer sequences and characteristics of the 20 SSR makers used in the intermountain range study.

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Super contig[a]	Size range (bp)	No. of repeats	Motif	Reference
PgtCAA39	CGTCGTCCCTCCATAGTCTTA	CTCTCAAGCACCCTCAACATC	2.11	197-245	13-29	CAA	This publication
PgtCAA49	TCGTCTGATCGTGAGAAACG	GACGATTGCTGAGGATTGCT	2.21	131-152	13-20	CAA	This publication
PgtCAA53	AGGCTCAACACCACCCATAC	AGGAGGAGGTGAAGGGGATA	2.14	215-239	8-16	CAA	Jin et al. 2009
PgtCAA80	GCCTCCAGACGAATGGTTTA	TTGGTGATGATGATGGTTGG	2.3	239-248	10-13	CAA	This publication
PgtCAA93	CGCCTGTGATGGTTGTATTG	CACTCTCGCCAAACCTCATT	2.43	155-182	5-13	CAA	Jin et al. 2009
PgtCAA98	ATTCGGATGGTCCGTTACTG	CCATCCCACTCAAATCATCC	2.10	153-180	8-17	CAA	Jin et al. 2009
PgtCAT4.2	CCGTGTCGATCCCAATAATC	AGCAAGGTGAGAATCGGAAA	2.16	123-144	8-15	CAT	This publication
PgtGAA8.1	TGTCTGCCTGTCTGTCGAAC	GGATGATCGGTCAGTTGGTT	2.33	188-209	5-12	GAA	This publication
Pgest21	CCGAATGCAGATTACCCTTG	GTTTGCCTGATGATGGATGA	2.2	246-255	6-9	AAG	This publication
Pgest24	TCATCGACCAAGAGCATCAG	TTCGGGAGTGAGTCTCTGCT	2.3	157-160	7-8	CAT	This publication
Pgest59	AGGTTGATGATGAGGATGC	ATTATGCGGGACAAATCGAG	2.5	215-221	5-7	ACG	This publication
Pgest109	CCATCCGATCATTTCTTCGT	CCGACCTTCTCTTGCTTCT	2.13	170-176	5-7	CCT	Zhong et al. 2009
Pgest142	GATGGTGAAGTCCGGTATGG	CCACCAACAAACCAACAAGA	2.17	198-213	9-14	TTG	This publication
Pgest173	TCCATTGAGTTCCATCGTGA	TCCCTTGACCATTCTCAACG	2.22	181-190	5-8	ATC	Zhong et al. 2009
Pgest227	CACACGTCTCGAGGAACAGA	CTCGTGGGATGAAGTCCATT	2.33	193-214	8-15	AAG	Zhong et al. 2009
Pgest293	GAACCTTGGCCTGAGTGCTA	GCAGCCTACAGCAAGAATCC	2.54	240-246	4-6	GGT	Zhong et al. 2009
Pgest318	GATGTCGGTCTTGGTCCACT	ACAGACACTCCCGAGCTCAT	2.64	256-274	5-11	ATG	This publication
Pgest325	TTGGGTGAGTCAGAGTTTGAGA	CCCACCCACTCTCAGTCAAT	2.70	171-180	4-8	AAG	This publication
Pgest341	GGCCTTGGTACCCAATTTCT	GATGTCGCACTCGGTTTCTT	2.83	214-217	4-5	TGG	Zhong et al. 2009
Pgest353	ACGTCTTGGGTTTCTGTGGA	TCGAATCCCAAGGAACAGAG	2.92	235-241	4-6	AGG	Zhong et al. 2009

[a]Information based on *Puccinia graminis* f. sp. *tritici* genome assembly (http://www.broadinstitute.org/annotation/genome/puccinia_group.1/MultiHome.html).

Table 1.3 Genotype statistics for 20 SSR loci across 62 *Puccinia graminis* f. sp. *tritici* isolates collected in the intermountain range region of the United States.

	Parameters[a]						
Locus	No. of genotypes	No. of alleles	Rare alleles	N_e	H_o	H_e	Missing data (%)
PgtCAA39	8	7	0	5.2	0.912	0.806	8.1
PgtCAA49	5	3	0	2.5	0.836	0.594	1.6
PgtCAA53	6	4	0	3.8	0.590	0.734	1.6
PgtCAA80	6	4	1	2.9	0.823	0.658	0.0
PgtCAA93	6	5	2	2.4	0.705	0.577	1.6
PgtCAA98	8	6	1	4.3	0.803	0.769	1.6
PgtCAT4.2	6	4	0	2.9	0.645	0.657	0.0
PgtGAA8.1	7	5	1	3.5	0.700	0.712	3.2
Pgest21	6	4	1	2.9	0.790	0.657	0.0
Pgest24	2	2	1	1.0	0.048	0.047	0.0
Pgest59	3	3	1	1.7	0.565	0.428	0.0
Pgest109	3	3	1	1.6	0.468	0.374	0.0
Pgest142	2	3	0	2.3	1.000	0.568	0.0
Pgest173	2	2	0	2.0	0.721	0.497	1.6
Pgest227	5	4	0	2.9	0.820	0.658	1.6
Pgest293	3	3	0	2.0	0.661	0.488	0.0
Pgest318	7	6	1	4.4	0.903	0.772	0.0
Pgest325	5	4	1	2.9	0.915	0.660	4.8
Pgest341	3	2	0	2.0	0.207	0.499	6.5
Pgest353	2	2	0	1.5	0.450	0.349	3.2
Mean	4.8	3.8	0.6	2.7	0.678	0.575	1.8

[a] N_e = Effective population size; H_o = Observed Heterozygosity; H_e = Expected Heterozygosity

Table 1.4 Population statistics for genotypic SSR markers for the nine genotypic groups and five genotypic clusters in intermountain range region of the United States[a].

Genotypic cluster and group	Parameters[b]							
	No. of samples	Average No. of alleles	N_e	H_o	H_e	F	No. of private alleles	No. of genotypes
GC I	24	2.60(0.197)	2.24(0.114)	0.90(0.056)	0.53(0.033)	-0.72(0.054)	9	6
NA-001 (R__Q)	9	1.95(0.050)	1.91(0.065)	0.91(0.065)	0.46(0.031)	-0.95(0.048)	0	2
NA-002 (H_LG)	5	1.80(0.092)	1.80(0.092)	0.80(0.092)	0.40(0.046)	-1.00(0.000)	0	1
NA-003 (T__K)	10	2.0(0.073)	1.97(0.054)	0.94(0.050)	0.48(0.026)	-0.97(0.018)	0	4
GC II	12	1.80(0.156)	1.63(0.131)	0.45(0.095)	0.31(0.055)	-0.45(0.115)	6	2
NA-004 (QCCJ)	6	1.50(0.115)	1.50(0.115)	0.50(0.115)	0.25(0.057)	-1.00(0.000)	3	1
NA-005 (QFCS)	6	1.40(0.112)	1.40(0.112)	0.40(0.112)	0.20(0.056)	-1.00(0.000)	1	1
GC III	16	1.95(0.211)	1.84(0.201)	0.55(0.101)	0.34(0.062)	-0.64(0.071)	2	4
NA-006 (_THJ/_KCJ)	9	1.60(0.112)	1.60(0.112)	0.59(0.111)	0.30(0.056)	-0.98(0.014)	0	2
NA-007 (MCC_)	7	1.5(0.115)	1.50(0.115)	0.50(0.115)	0.25(0.057)	-1.00(0.000)	1	1
-								
NA-008 (R__S)	5	1.80(0.117)	1.72(0.105)	0.69(0.100)	0.36(0.050)	-0.89(0.058)	9	3
-								
NA-009 (Q_CQ)	5	1.60(0.112)	1.56(0.112)	0.56(0.112)	0.28(0.055)	-0.93(0.057)	0	2

[a] Genotypic population statistic data is shown as an average and standard error across all 20 loci.
[b] N_e = Effective population size; H_o = Observed Heterozygosity; H_e = Expected Heterozygosity; F = Fixation index

Table 1.5 Number of genotypic SSR allelic differences for pair wise comparisons of race groups in the intermountain range region of the United States.

	Genotypic groups								
Genotypic groups	NA-001	NA-002	NA-003	NA-004	NA-005	NA-006	NA-007	NA-008	NA-009
NA-001	31								
NA-002	17	0							
NA-003	17	16	3						
NA-004	31	32	32	0					
NA-005	33	34	34	14	0				
NA-006	31	26	32	20	22	1			
NA-007	37	36	22	26	24	16	0		
NA-008	45	46	40	36	32	36	30	4	
NA-009	27	40	34	22	24	22	22	24	1

Table 1.6 R_{st} values of genetic differentiation between race groups of *Puccinia graminis* f. sp. *tritici* in the intermountain range region of the United States based on SSR markers.

	Genotypic groups[a]								
Genotypic groups	NA-001	NA-002	NA-003	NA-004	NA-005	NA-006	NA-007	NA-008	NA-009
NA-001									
NA-002	0.066								
NA-003	0.276*	0.243*							
NA-004	0.141*	0.158*	0.282*						
NA-005	0.162*	0.138*	0.215*	0.098					
NA-006	0.227*	0.251*	0.388*	0.072	0.188*				
NA-007	0.300*	0.231*	0.162*	0.064	0.043	0.168*			
NA-008	0.385*	0.556*	0.607*	0.325*	0.449*	0.469*	0.431*		
NA-009	0.258*	0.459*	0.372*	0.170*	0.126*	0.185*	0.111	0.500*	

[a] Asterick (*) = significant difference at $P < 0.05$.

Table 1.7 Number of phenotypic virulence differences for pair wise comparisons of the race groups in the intermountain range region of the United States.

	Phenotypic groups								
Phenotypic groups	R__Q	H_LG	T__K	QCCJ	QFCS	_THJ/_KCJ	MCC_	R__S	Q_CQ
R__Q	3								
H_LG	5	1							
T__K	6	4	3						
QCCJ	8	7	6	0					
QFCS	6	6	6	2	0				
_THJ/_KCJ	3	6	5	4	5	4			
MCC_	10	9	6	4	6	7	1		
R__S	3	9	6	6	6	5	8	2	
Q_CQ	5	6	6	3	2	6	7	5	1

Table 1.8 Φ_{PT} values of differentiation between race groups of *Puccinia graminis* f. sp. *tritici* based on virulence phenotypes in the intermountain range region of the United States.

	Phenotypic groups[a]								
Phenotypic groups	R__Q	H_LG	T__K	QCCJ	QFCS	_THJ/_KCJ	MCC_	R__S	Q_CQ
R__Q									
H_LG	0.736*								
T__K	0.781*	0.784*							
QCCJ	0.875*	0.971*	0.827*						
QFCS	0.832*	0.971*	0.856*	0.000					
_THJ/_KCJ	0.716*	0.816*	0.774*	0.730*	0.730*				
MCC_	0.877*	0.935*	0.810*	0.913*	0.944*	0.818*			
R__S	0.636*	0.886*	0.776*	0.888*	0.888*	0.735*	0.870*		
Q_CQ	0.756*	0.911*	0.839*	0.897*	0.806*	0.746*	0.906*	0.833*	

[a] Asterisk (*) = significant difference at $P < 0.05$.

Table 1.9. Race group designation changes of *Puccinia graminis* f. sp. *tritici* in the intermountain range region of the United States over the past 50 years.

Stakman et al. 1962	Roelfs and McVey 1976	Roelfs and Groth 1980	Burdon and Roelfs 1985	Roelfs et al. 1997	This study
11	11-32-113 (-RCR, -RKQ, -RPQ, -RTQ)	11(-RCR, -RHR)	11-RCR	RCRS	R__S
11	11-32-113 (-RCR, -RKQ, -RPQ, -RTQ)	11(-RCR, -RHR)	11-RHR	RHTS	R__S
17			17-H_L	H_LG	H_LG
38	151(-QSH, -QCB, -QFB)	151(-QCB, -QFB)	151-Q_B	Q_CQ	Q_CQ
113	11-32-113 (-RCR, -RKQ, -RPQ, -RTQ)	113(-RKQ, -RPQ, -RTQ)	113-R_Q	R__Q	R__Q
29			29-32_JC	_KCJ	_THJ/_KCJ
48	151(-QSH, -QCB, -QFB)	151(-QSH), 32(-RSH)	151-32_SH	_THJ	_THJ/_KCJ
56	56(-MBC)	56(-MBC)	56-MBC	MCC_	MCC_
15	15(-TLM, -TDM, -TNM)	15(-TLM, -TDM, -TNM)	15-T_M	T__K	T__K
				QFCS	QFCS
				QCCJ	QCCJ

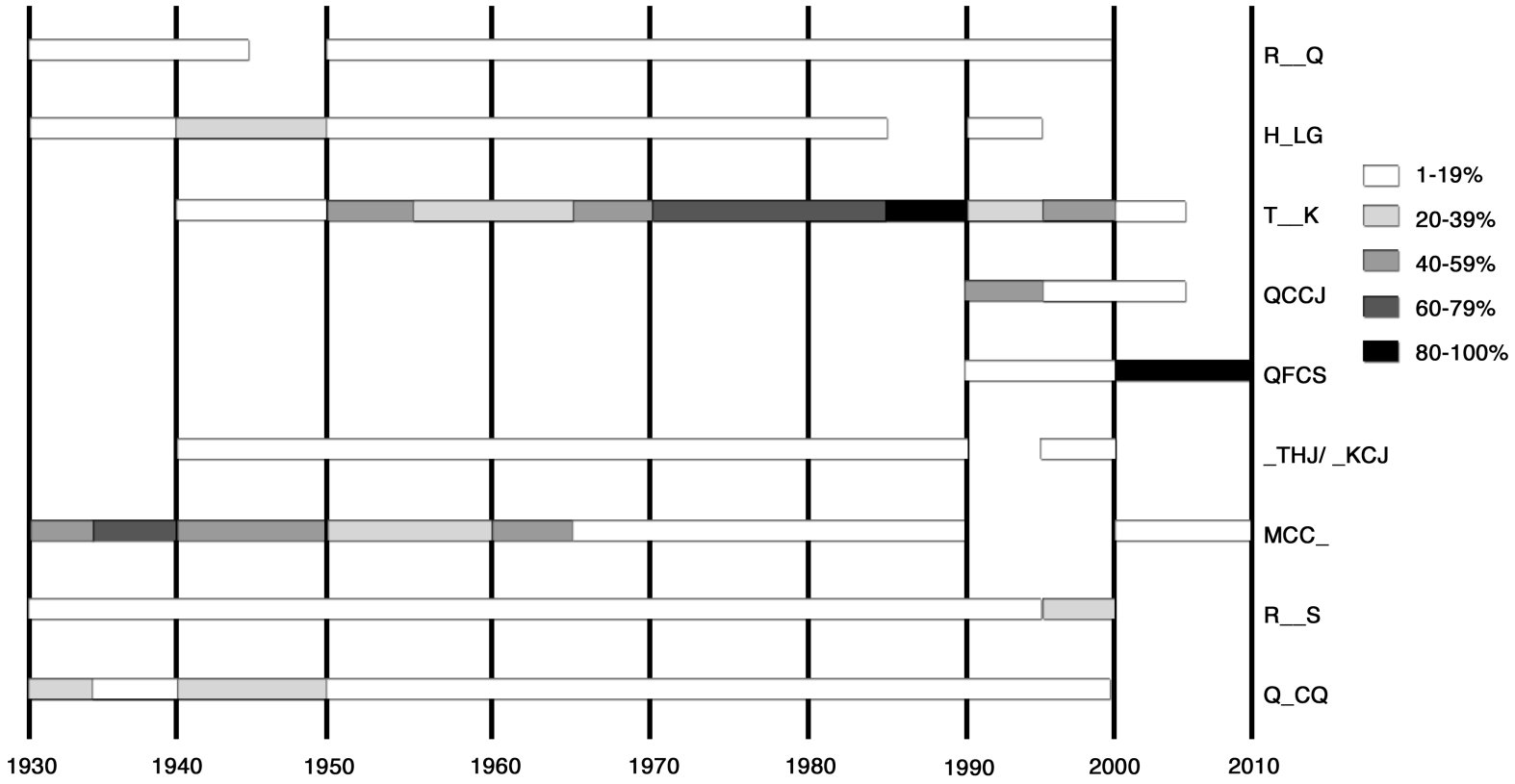

Figure 1.1 Proportion of *Puccinia graminis* f. sp. *tritici* race groups collected over the past 80 years in the intermountain range region region of the United States.

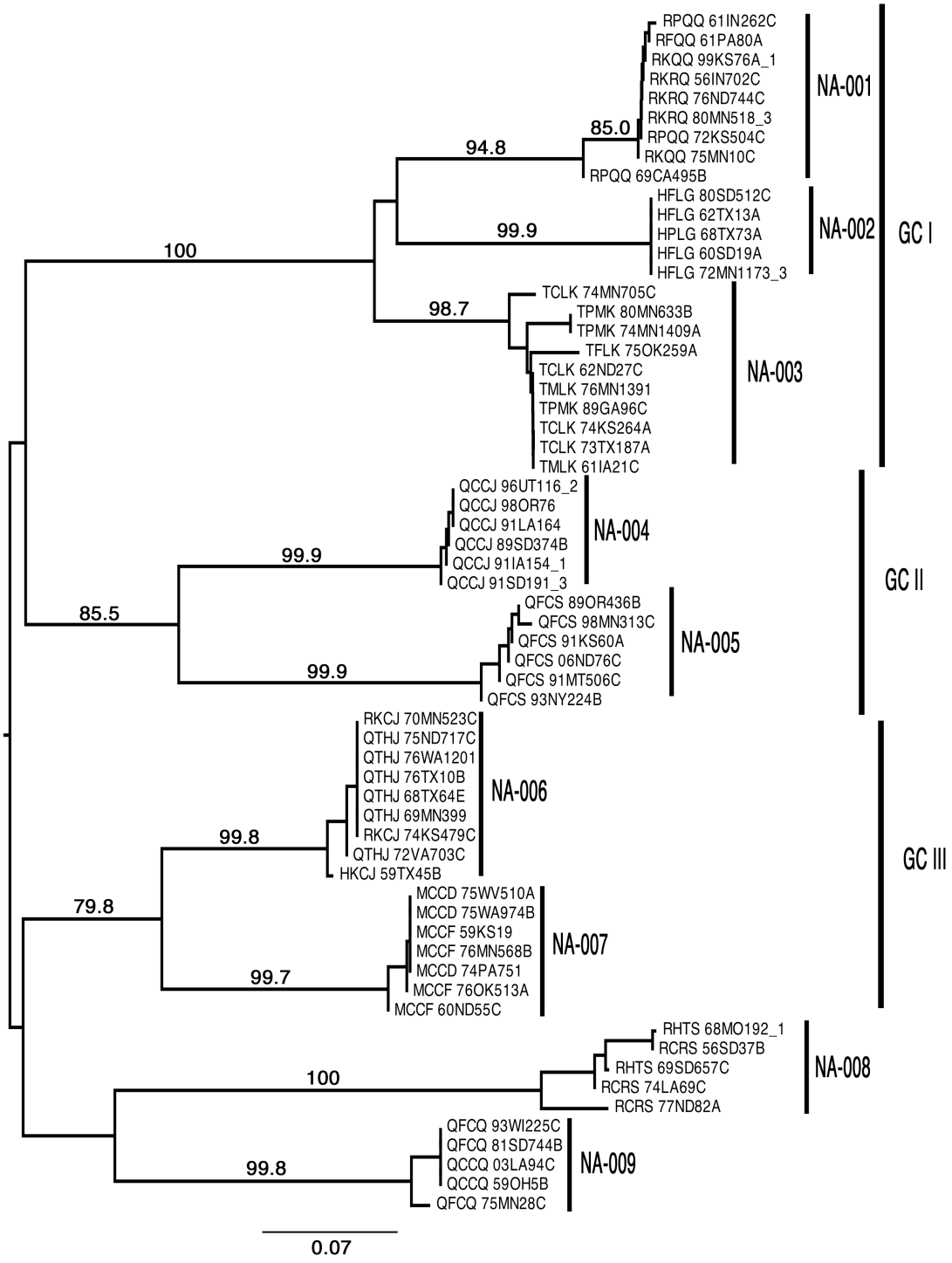

Figure 1.2 Midpoint rooted neighbor joining tree of *Puccinia graminis* f. sp. *tritici* isolates collected in the intermountain range region of the United States constructed from 20 simple sequence repeat markers. Only branches with bootstrap support of greater than 75% are labeled.

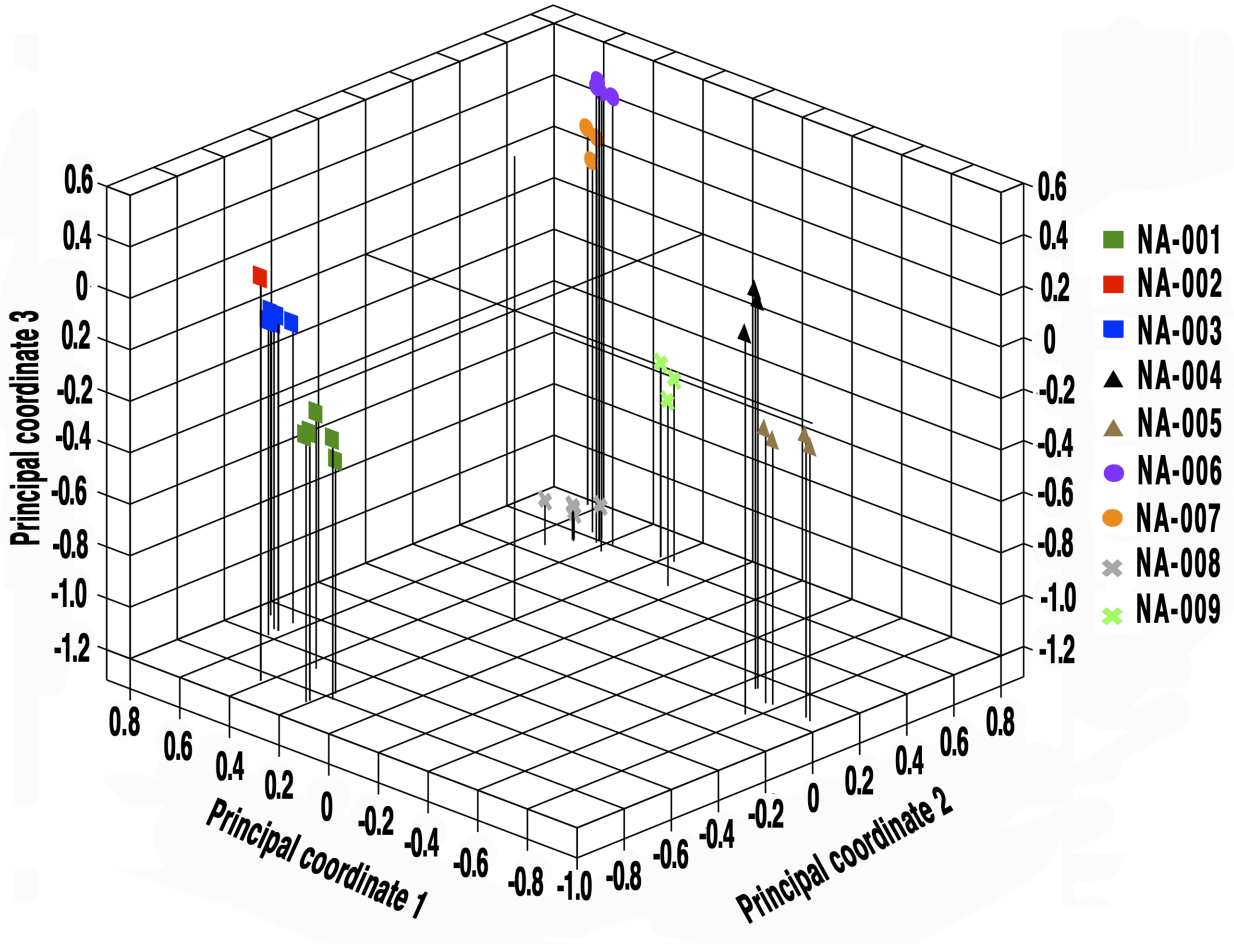

Figure 1.3 Principal coordinate analysis of *Puccinia graminis* f. sp. *tritici* isolates collected in the intermountain range region of the United States constructed of 20 simple sequence repeat markers.

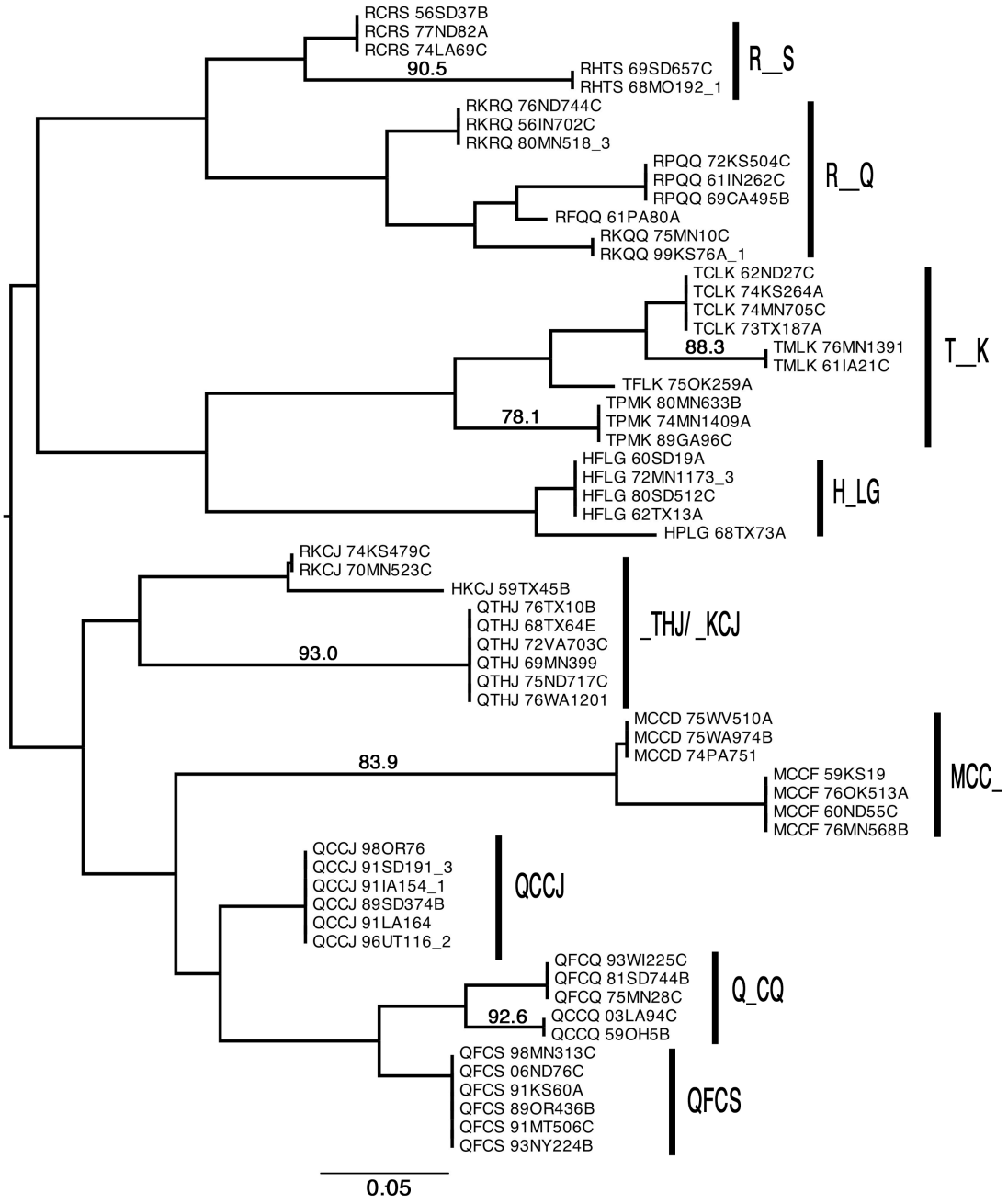

Figure 1.4 Midpoint rooted neighbor joining tree of *Puccinia graminis* f. sp. *tritici* isolates collected in the intermountain range region of the United States constructed from 16 virulence phenotype markers. Only branches with bootstrap support of greater than 75% are labeled.

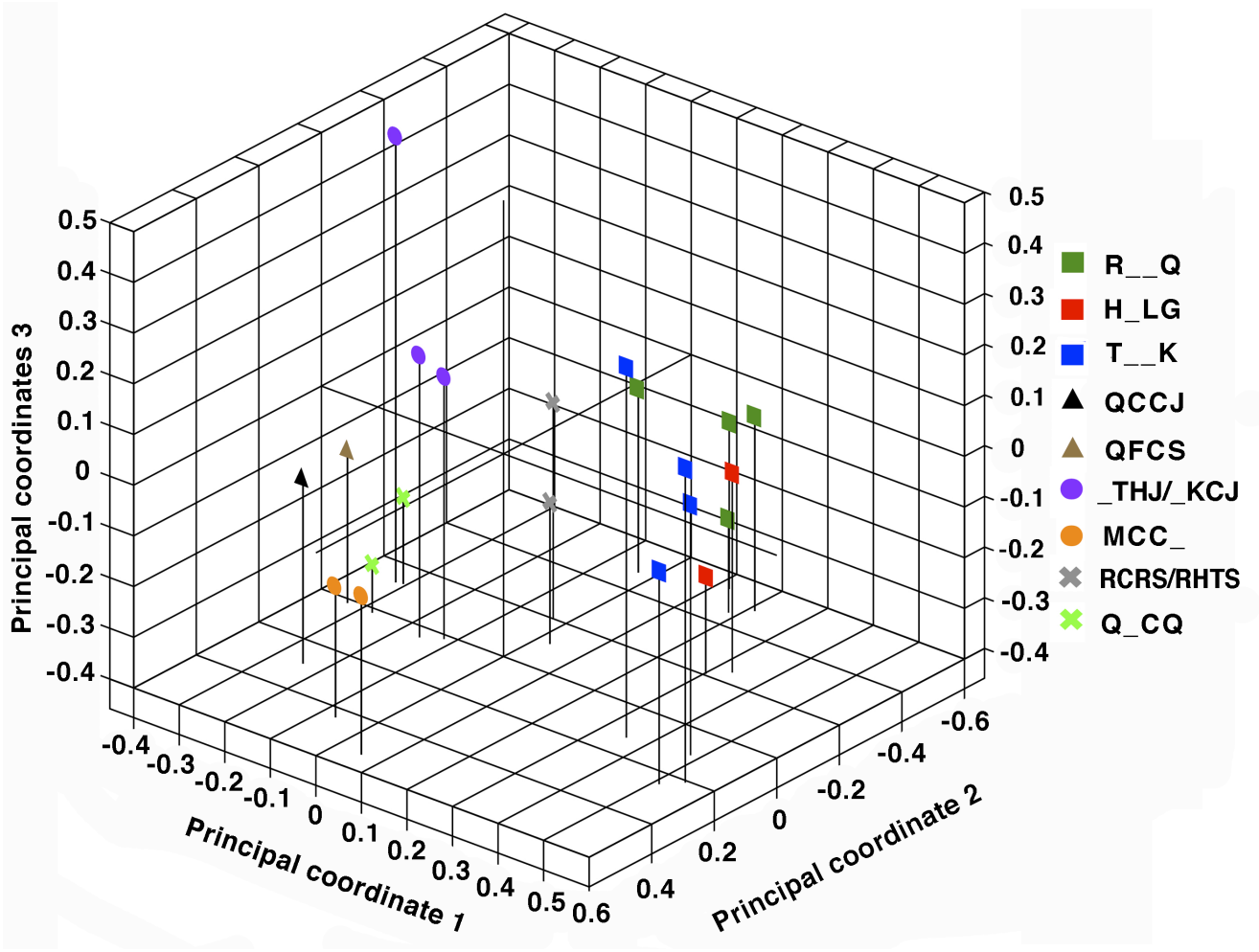

Figure 1.5 Principal coordinate analysis of *Puccinia graminis* f. sp. *tritici* isolates collected in the intermountain range region of the United States constructed from 16 virulence phenotype markers.

Chapter 2

Genetic Structure of the Wheat Stem Rust Pathogen in a Field in Washington State

Introduction

Severe epidemics of stem rust on wheat in the early and mid part of the 20th century caused a major initiative to take steps in reducing the impact of this fungus. With state and federal funding research was conducted to develop wheat lines with multiple resistance genes and a push to eradicate common barberry (*Berberis vulgaris*), the alternate host of *Pgt,* was undertaken (Roelfs 1982). By the mid 1920s the majority of the barberry bushes directly adjacent to wheat fields had been removed and surveys were conducted to monitor areas where bushes had been known to grow. The loss of a secondary host for *Pgt* stopped sexual reproduction, limited and caused the stabilization of the number of races found in the central United States, which has resulted in long-term effectiveness of many resistance genes. The barberry eradication program focused on removing bushes in the major wheat growing section of the United States that lies between the Rocky Mountain and Appalachian Mountain Ranges (IMR region). Barberry bushes outside this area did not succumb to the efforts of eradication leaving the possibility of sexual reproduction in *Pgt* in these areas.

Race surveys from the 70's show that collections made west of the Rocky Mountains have a larger number of stem rust races present and a variety of races not normally seen in the IMR region of the United States (Burdon and Roelfs 1985a). This pattern has lead to hypotheses that there is a sexual population east of the Rocky Mountains but infected barberry plants have not been found in the area so the source of the sexual reproduction has not been identified.

Until their detection in the central United States in 1989 (Roelfs et al 1991) races QCCJ and QFCS had previously only been collected in the Pacific Northwest. When QCCJ was first identified east of the Rocky Mountains it was primarily a pathogen of barley but in 1990 QCCJ was found on winter wheat plots in Kansas and Colorado (Roelfs et al. 1993) and established itself as a significant member of the *Pgt* population until 1996. Race QFCS is now the most prevalent race found in North America due to its virulence on several hard red winter wheat and soft red winter wheat cultivars (Kolmer et

al 2007). When first identified both races were found in the northern states of Minnesota, North Dakota, South Dakota and Idaho indicating the possibility that they crossed the Rocky Mountains and have now become established in the North American asexual population (Roelfs et al. 1991). Roelfs et al. (1997) used avirulence phenotypic reactions and isozyme markers in an attempt to find evidence of race QCCJ and QFCS's origins. Unfortunately the molecular techniques of the time could not shed light on this hypothesis and strong evidence for this migration event is still lacking.

The objectives of this study are; (i) genotypically evaluate field samples of *Pgt* collected in the Pacific Northwest in order to examine the population structure (ii) describe the relationship these isolates have with isolates of the asexual populations of the intermountain range region of the United States. If sexual reproduction is taking place in the Pacific Northwest and isolates are found to be crossing the Rocky Mountains there is a significant risk of races of *Pgt* that can overcome commonly used resistance gene combinations entering the wheat growing areas of the United States and producing epidemics not seen since the beginning of the 20th century.

Materials and methods

Pgt isolates and DNA extractions.

In 2008 bulk samples of barley tissue collected from a single field in Arden WA were sent to the Cereal Disease Lab in Saint Paul MN where samples of *Puccinia graminis* urediniospores were collected. Urediniospore samples were spray inoculated in a solution of soltrol oil onto Line E of wheat, Hiproly barley and Prolific rye. Inoculations were performed on ~8-d old wheat seedlings, allowed to dry on plants for 30m, then incubated overnight in a misting chamber. After drying, plants were transferred to a greenhouse and approximately two weeks after inoculation 61, 14 and 4 single pustule isolations were collected from each respective host. A summary of the isolates used in this study is listed in Table 2.1. Dr. Matt Rouse performed inoculation and collection of isolates. Genomic DNA was extracted from pulverized tissue from germinated or ungerminated urediniospores using the OmniPrep DNA Extraction kit (GenoTech, St. Louis MO) according to the manufacturer's protocol as described in Chapter 1.

Genotypic analysis.

SSR analyses were performed using the 20 SSR primers developed and shown to be useful in previous analyses using *Pgt* in the United States. PCRs were performed in a PTC-200 (MJ Research, Waltham MA) thermocycler as follows: 10 µL reactions containing 1X Phusion HF buffer with 1.5 mM $MgCl_2$ (New England BioLabs, Ipswich MA), 0.2 mM dNTPs (Roche, Indianapolis IN), 0.005% casein (Sigma, St. Louis MO), 1 µM IRD700/800 labeled forward primer (LI-COR, Lincoln NE), 1 µM reverse primer, 0.1 U Phusion Taq polymerase (New England BioLabs, Ipswich MA), and 20ng DNA, with a cycle of: 98 °C for 5 min, 30 cycles at 98 °C for 30 s, 62 °C for 30 s, 72 °C for 30 s and a final extension at 72 °C for 10 min. PCR products were diluted 1:10 in ddH_2O, 5 µL of Stop Buffer (31.5 ml formamide, 1.32 ml 0.5 M EDTA, 330 µl dH_2O and 48 mg bromophenol blue) was added, and samples were denatured for 3 min prior to analysis on a LI-COR 4300 DNA sequencer (LI-COR, Lincoln NE) using a 7% polyacrylamide gel. Amplicon sizes were determined using a DNA size standard ranging from 50 to 350 bp (LI-COR, Lincoln NE) and the SAGAGT software (LI-COR, Lincoln NE). When examining gels, amplicons within a lane were scored if they exhibited similar amplification intensity as another amplicon in the same lane; amplicons of less intensity were considered bleed over from adjacent lanes or amplification from a secondary locus and ambiguous results were tested at least twice to increase accuracy. *Pgt* isolate CRL 75-36-700-3 was included on each gel as a standard and amplicons were scored as number of bases, then converted to a repeat number using the predicted allele size based on the *Pgt* genome sequence as a size standard. Reactions that did not amplify or amplified more than 2 products were recorded as missing data.

Phenotypic analysis.

The 79 single pustule urediniospores samples were spray inoculated in a solution of soltrol oil on a standard set of 16 wheat differentials (Roelfs and Martens 1988). Inoculations were performed as described and approximately two weeks after inoculation, race phenotypes were determined as outlined by Roelfs and Martens (1988). Race phenotype tests were repeated three times to insure accuracy of race designations. Dr. Matt Rouse conducted all phenotyping of Washington isolates.

Data analyses.

For each SSR locus, the total number of alleles, number of alleles private to a race or race group, number of rare alleles per race group (alleles found at a frequency of 5% or less), the number of effective alleles, observed and expected heterozygosity, the fixation index and population assignment analyses were calculated using GenAlEx (Peakall and Smouse 2006). Three-dimensional principal coordinate analyses (PCA) were conducted using GenAlEx6, three-dimensional plots were constructed using Delta Graph (Red Rock Sofware, Salt Lake City UT) to assess clustering of individuals within race groups for genotypic data. Bootstrap support was assessed by constructing 1000 neighbor-joining (NJ) trees from a frequency matrix based on shared alleles in the program PowerMarker V3.25 (Liu and Muse 2005). Consensus trees were constructed from the 1000 trees using Phylip 3.67 (Felsenstein 1989) and used to calculate the bootstrap support; trees were viewed using FigTree (http://tree .bio.ed.ac.uk/software/figtree/) and are displayed as midpoint rooted. In the above programs missing data were ignored and not coded as an additional state for analyses.

The number of genotypic differences within a race group and between race groups was calculated by counting the number of alleles found in only one of two race groups in the pair wise comparisons. Calculations of total numbers of genotypes were conducted assuming missing data points contained identical amplicons as the other isolates in the race group. The level of genotypic (Rst) differentiation was calculated over 999 permutations of the data set using AMOVA in GenAlEx6.

To compare the isolates collected in Washington to the intermountain range populations in Chapter 1 an analysis was performed including both data sets. For ease of discussion isolates outlined in Chapter 2 will be referred to as WA (Washington) isolates and isolates from Chapter 1 will be referred to as IMR (Intermountain Range) isolates. The previous IMR study included one SSR locus (Pgest59) that was not included in this study due to a high failure rate. This marker was removed for the comparative analyses.

Results

Genetic variation and cluster analyses.

A summary of genetic variability for the 19 SSR markers is presented in Table 2.2. From the 19 SSR loci examined, 93 alleles were amplified, with 2-11 alleles per marker and an average of 1.9% missing data per locus.

Neighbor-joining (NJ) analysis of the SSR data matrix assigned the 79 isolates of *Pgt* into genetic groups (Figure 2.1). Nineteen branches on the NJ tree were found to have higher than 75% bootstrap support. Of these 19 well-supported branches only 4 of the clades contained 4 or more isolates (WA-001, WA-002, WA-003 and WA-004) and were considered separate groups from the rest of the isolates in the study (WA-005). The frequency matrix used to construct the phylogenetic tree contained 93 characters, 84 of which were informative. Principal coordinate (PC) analysis was conducted to validate the clustering pattern found in the NJ analysis (Figure 2.2). The first three dimensions show 66.30% of the variation in the data. Clustering patterns were consistent between analyses.

Pairwise comparison was used to determine the number of allelic differences within and between groups to further validate the groupings (Table 2.3). WA-001 was found to have only seven allele differences between isolates within the group and a range of 27-59 differences from the other groups. WA-002 group was the only supported clade composed of genotypically identical isolates. The NJ analysis grouped five isolates into WA-003 with high support (97.7%) and a sixth isolate (isolate 180) with much lower support (67.7%). Isolate 180 was found to have five alleles not found in the five isolates in WA-003 but was found to have 20-60 allele differences when compared to the other 4 groups. A population assignment test also showed that isolate 180 was more likely a member if WA-003 than any of the other groups so was included in WA-003 for further analyses. WA-004 contained the lowest number of isolates and had the second highest number of allele differences (eight, five of which were due to locus CAA39) but the high level of support, high number of allele differences between the other groups (24-56) and high probability as a separate group in population assignment tests led us to continue to consider WA-004 as a separate group. Within WA-004 there is further support for two

groupings each containing two isolates collected from different hosts with one and two allele differences within the two groups and 7 between. To test for differentiation between groups an Rst analysis was performed on the genotypic data. Significant levels of differentiation were found between all the groups except when comparing WA-004 and WA-005 (Table 2.4).

Estimates of genetic variability partitioned by race are summarized in Table 2.5. The average observed heterozygosity over all 19 markers was found to be higher than what was expected for the WA-001, WA-002, WA-003 and WA-004 groups and resulted in a negative fixation index with a range between -1 and -0.54. WA-005 had an average observed heterozygosity similar to what is expected under Hardy-Weinberg Equilibrium and had a fixation index value of 0.11 indicating a sexual population.

Race phenotypes and cluster analyses.

When examining the phenotypic data with the restrictions of the five groups obtained from the genotype analysis, groupings of phenotype are observed. WA-001 is made up of nine isolates, seven with race code JCBB and one isolate each of the race codes GCBB and SCBB corresponding to two virulence differences within the group (Table 2.1). WA-002 group was the only group containing phenotypically identical isolates (QHMJ). WA-003 contained five isolates of QCMB and an isolate of QCBB, which corresponds to two virulence differences. WA-004 contained four isolates, three of which were BCBG and one that was BCBB, which corresponded to one virulence difference within the group. The remaining 54 samples comprise WA-005 and consist of 18 different phenotypes. WA-005 also contains isolates with phenotypes identical to isolates in WA-001 and WA-004.

Genetic variation and cluster analyses (combined).

A NJ analysis of the SSR data from the IMR populations in Chapter 1 and the WA populations in Chapter 2 identified many but not all of the previously identified associations (Figure 2.3). All of the basal groupings or race groups for both the WA and IMR groups were identified but the only race group cluster identified from the IMR isolates was RGC 1, which contained isolates of phenotypic race codes T__K, H_LG and

R__Q. A PC analysis was conducted to identify any further groupings of isolates (Figure 2.4). The first three dimensions account for 63.18% of the variation. The plot shows separation between the majority of the IMR isolates and the WA isolates. The isolates of the NA-004 (race QCCJ) and NA-005 (race QFCS) groups were found to be more closely related to the WA isolates than to the other IMR isolates.

To examine the level of genetic differentiation between race groups Rst values were calculated and significant levels of genetic differentiation were found between all but 11 pair wise comparisons (Table 2.6). Six of the significant comparisons within the IMR race groups and one within the WA race groups were detected in previous Rst analyses in Chapters 1 and 2. The remaining significant relationships are the comparisons of NA-004 (race QCCJ) to WA-004 and NA-005 (race QFCS) to WA-003, WA-004, and WA-005.

Discussion

Race surveys conducted since at least the 1970's indicate the presence of a sexual population of *Pgt* in the Pacific Northwest (Roelfs et al. 1991). The majority of past analyses on isolates collected from this region focused on phenotypic race data with very little information collected on the genetic structure of the populations. To get a better understanding of the genetic make up of *Pgt* in this region 79 isolates collected from a single field and comprising 23 virulence phenotypes, were examined with the use of 19 SSR loci. Genetic analyses determined the WA isolates in this field can be broken into four well supported genetic groups.

The four supported genetic groups (WA-001 – WA-004) identified by the SSR analysis have few differences between isolates and one large group that is of mixed genotype (WA-005). The WA-005 contains 54 isolates, 18 phenotypes, 53 genotypes and the SSR loci do not deviate from Hardy Weinberg equilibrium indicating that this group is a sexual population. The four smaller groups of isolates have between 4-9 isolates, few phenotypic (0-2) and genotypic (0-9) differences within the groups, exhibit a deviation from Hardy Weinberg equilibrium with an excess of heterozygosity and are considered to be clonal race groups. Although there was a skewed sample size toward isolates collected

off of wheat, the isolates collected from rye and barley were phenotypically and genotypically similar to isolates collected on wheat and there doesn't appear to be an association toward host.

WA-001 was the largest of the clonal groups with nine members and contains three phenotypes and five genotypes. The predominant race code found in this group is JCBB with the remaining isolates differing by one or two resistant genes. These phenotype changes also corresponded to 3 of the differences in genotype in the group with the fourth being an allele change within the JCBB isolates. The isolate SCBB is the only isolate in this group that was increased on barley and contains four alleles that are not present in the other eight members of this group but phylogenetic analyses place it within WA-001 (Figure 2.1). WA-002 contains six isolates, one isolate increased on barley and five on wheat, and is the only race group that contained isolates of completely identical phenotype and genotype. WA-003 contains six isolates, all increased of wheat, with two phenotypes and three genotypes. Five of the members of this group have a QCMB phenotype and the other member a QCBB phenotype. The QCBB isolate contains five alleles not found in the other isolates from this group but has between 20-60 allele differences when compared to the other four groups and a population assignment test included the isolate in the WA-003 group. WA-004 is the smallest and most varied group containing four isolates with two phenotypes and each with a unique genotype. Within WA-004 isolates collected on different hosts have high support as being more closely related to an isolate found on the same host. Although each isolate is genetically unique there are only four alleles and one virulence phenotype not shared by all member of this group. The dissimilarity between the WA-004 isolates and other isolates can be seen in high bootstrap support on the genotypic NJ tree (Figure 2.1) and has a higher probability as a separate group than as members of the large sexual population WA-005. WA-004 is also the only group that contained further phylogenetic support within the group. This phylogenetic relationship corresponds with which host the isolates were increased, with two isolates from barley and the other two from wheat.

The four groups found in this study appear to be reproducing asexually as is seen in the IMR populations. The phylogenetic analysis identified 10 other branches with high

bootstrap support (87.8-100%) but the supported clades had fewer than 4 isolates and although they may also represent asexual populations these isolates were not considered separate groups for analyses in this study. The results of molecular analyses performed in this study support the idea of long-term stability of these asexual populations and makes the possibility of a bias due to sampling technique unlikely. The fixation indices of the four clonal groups were found to be close to -1. This is due to the high level of heterozygosity between the loci studied. The likelihood of a single or a few cycles of clonal propagation yielding this high a level of heterozygosity is unlikely and it is more probable that this is a result of a long-term fitness advantage. If these asexual populations do in fact persist over multiple years further research into these areas could lead to a better understand of what the population structure of *Pgt* may have looked like prior to the eradication of barberry in the IMR region.

The combined analyses of the isolates collected from the field in Washington and the IMR populations illustrate the effects the different sexual modes of reproduction have on phenotypic and genotypic characters. The high number of race phenotypes present in the sexually reproducing WA population is a stark contrast to the few races that are normally found in the IMR populations per year or even over the last 80 years. In addition, a lower percentage of avirulence reactions for each phenotypic marker was found in the WA isolates compared to the IMR isolates but the WA isolates had combinations of avirulence genes that are not commonly found in the IMR populations. These findings are in agreement with an earlier study where isolates assumed to be sexually reproducing collected in Washington and Idaho were tested on 16 wheat differential lines and were found to have a larger number of phenotypes than an asexual population (Roelfs and Groth 1980). Genetically the sexually reproducing WA isolates have more genotypes and private alleles than the IMR asexual collections as well as a fixation index value that didn't significantly deviate from zero as opposed to significant deviations found in the asexual collections. Again this is in agreement with a previous study that used isozyme markers to examine populations of sexual and asexual isolates of *Pgt* (Burdon and Roelfs 1985b).

The most intriguing findings of the combined analysis are the findings of similarity between the NA-004 (QCCJ) and the NA-005 (QFCS) isolates with the WA isolates. Isolates of the NA-004 and NA-005 race groups were not found in the IMR region until the late 80's but previous race surveys of the Pacific Northwest show these races to be common prior to this date. The association between NA-004, NA-005 and isolates from Washington was first proposed in 1991 (Roelfs et al. 1991) and further examined in 1997 (Roelfs et al. 1997) but experimentation was based on phenotypic data and lacked genotypic support. Despite the close phenotypic similarity to other race groups found in IMR region since the early 30's, namely NA-009 (Q_CQ), genotypic differentiation shows that NA-004 and NA-005 race groups have a different common ancestor. SSR analyses in Chapter 1 also showed that the NA-004 and NA-005 race groups have high support as sister races and are genetically separate from the other IMR race groups of *Pgt*. In this study a PC analysis of the combined genotypic data of the IMR and WA isolates places the NA-004 and NA-005 isolates within the cluster of WA isolates. An Rst analysis also shows that the NA-005 group is genetically similar to WA-003, WA-004 and WA-005. However, the phylogenetic analysis of the IMR and WA isolates did not show an association between the NA-004 and NA-005 race groups and the WA isolates with a bootstrap value of 42.4%. The lack of phylogenetic support could be due to the low number of alleles found in many of the markers and the high level of allele shuffling in the sexual population. The inclusion of the WA isolates into the phylogenetic analysis also removed many of the GCs identified in the IMR study. Overall the analyses presented in this paper show genetic evidence that supports the idea that the NA-004 and NA-005 race groups of the IMR population were likely due to founder events from the Pacific Northwest and additional introductions are possible.

Table 2.1 Summary of *Puccinia graminis* f. sp. *tritici* isolates used in the Washington study.

Genotype	Race	ID	Increase Host	Virulence (*Sr* genes)	Avirulence (*Sr* genes)
WA-001a	JCBB	155	Wheat	9e, 9g, 21	5, 6, 7b, 8a, 9a, 9b, 9d, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-001a	JCBB	163	Wheat		
WA-001a	JCBB	168	Wheat		
WA-001b	JCBB	169	Wheat		
WA-001b	JCBB	173	Wheat		
WA-001c	JCBB	182	Wheat		
WA-001a	JCBB	183	Wheat		
WA-001d	GCBB	175	Wheat	9g, 21	5, 6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-001e	SCBB	542	Barley	5, 9e, 9g, 21	6, 7b, 8a, 9a, 9b, 9d, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-002a	QHMJ	130	Wheat	5, 6, 9d, 9g, 10, 17, 21, 36	7b, 8a, 9a, 9b, 9e, 11, 24, 30, 31, 38, Tmp
WA-002a	QHMJ	135	Wheat		
WA-002a	QHMJ	137	Wheat		
WA-002a	QHMJ	142	Wheat		
WA-002a	QHMJ	147	Wheat		
WA-002a	QHMJ	515	Barley		
WA-003a	QCBB	180	Wheat	5, 9g, 21,	6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-003b	QCMB	114	Wheat	5, 9g, 17, 21, 36,	6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 24, 30, 31, 38, Tmp
WA-003b	QCMB	156	Wheat		
WA-003c	QCMB	161	Wheat		
WA-003b	QCMB	170	Wheat		
WA-003b	QCMB	172	Wheat		

Table 2.1 continued.

Genotype	Race	ID	Increase Host	Virulence (*Sr* genes)	Avirulence (*Sr* genes)
WA-004a	BCBB	526	Barley	9g	5, 6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-004b	BCBG	179	Wheat	9d, 9g	5, 6, 7b, 8a, 9a, 9b, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-004c	BCBG	181	Wheat		
WA-004d	BCBG	541	Barley		
WA-005	BBBB	102	Wheat	-	5, 6, 7b, 8a, 9a, 9b, 9d, 9e, 9g, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BBBB	109	Wheat		
WA-005	BBBB	110	Wheat		
WA-005	BBBB	115	Wheat		
WA-005	BBBB	121	Wheat		
WA-005	BBBB	124	Wheat		
WA-005	BBBB	126	Wheat		
WA-005	BBBB	128	Wheat		
WA-005	BBBB	154	Wheat		
WA-005	BBBB	158	Wheat		
WA-005	BBBB	166	Wheat		
WA-005	BBBB	504	Barley		
WA-005	BBBB	510	Barley		
WA-005	BBBB	519	Barley		
WA-005	BBBB	532	Barley		
WA-005	BBBB	533	Barley		
WA-005	BBBB	902	Rye		
WA-005	BBBB	905	Rye		

Table 2.1 continued.

Genotype	Race	ID	Increase Host	Virulence (*Sr* genes)	Avirulence (*Sr* genes)
WA-005	BBBG	159	Wheat	9d	5, 6, 7b, 8a, 9a, 9b, 9e, 9g, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BCBB	108	Wheat	9g	5, 6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BCBB	123	Wheat		
WA-005	BCBB	152	Wheat		
WA-005	BCBB	153	Wheat		
WA-005	BCBB	178	Wheat		
WA-005	BCBB	505	Barley		
WA-005	BCBB	536	Barley		
WA-005	BCBG	118	Wheat	9d, 9g	5, 6, 7b, 8a, 9a, 9b, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BCBG	131	Wheat		
WA-005	BCBG	149	Wheat		
WA-005	BCBQ	116	Wheat	9a, 9d, 9g	5, 6, 7b, 8a, 9b, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BCGG	167	Wheat	9b, 9d, 9g	5, 6, 7b, 8a, 9a, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BMBB	146	Wheat	9g, 11	5, 6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	BMBB	151	Wheat		
WA-005	GCBB	134	Wheat	9g, 21	5, 6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-005	GCBB	136	Wheat		
WA-005	GCBB	143	Wheat		
WA-005	GCBB	145	Wheat		
WA-005	GCBG	105	Wheat	9d, 9g, 21	5, 6, 7b, 8a, 9a, 9b, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-005	GCBG	150	Wheat		
WA-005	GCBG	171	Wheat		
WA-005	GCBG	903	Rye		

Table 2.1 continued.

Genotype	Race	ID	Increase Host	Virulence (*Sr* genes)	Avirulence (*Sr* genes)
WA-005	GHBB	144	Wheat	6, 9g, 21	5, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	JCBB	141	Wheat	9e, 9g, 21	5, 6, 7b, 8a, 9a, 9b, 9d, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-005	JCBB	177	Wheat		
WA-005	LCBN	139	Wheat	5, 9a, 9g, 10	6, 7b, 8a, 9b, 9d, 9e, 11, 17, 21, 24, 30, 31, 36, 38, Tmp
WA-005	QCBB	107	Wheat	5, 9g, 21	6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-005	QCBG	112	Wheat	5, 9d, 9g, 21	6, 7b, 8a, 9a, 9b, 9d, 9e, 10, 11, 17, 30, 31, 36, 38, Tmp
WA-005	QCBG	125	Wheat		
WA-005	QCBG	157	Wheat		
WA-005	QCBQ	140	Wheat	5, 9a, 9d, 9g, 21	6, 7b, 8a, 9b, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-005	QCBQ	517	Barley		
WA-005	QCCN	527	Barley	5, 9a, 9g, 10, 17, 21	6, 7b, 8a, 9b, 9d, 9e, 11, 24, 30, 31, 36, 38, Tmp
WA-005	QHBG	901	Rye	5, 6, 9d, 9g, 21	7b, 8a, 9a, 9b, 9e, 10, 11, 17, 24, 30, 31, 36, 38, Tmp
WA-005	SFBD	531	Barley	5, 8a, 9e, 9g, 10, 21	6, 7b, 9a, 9b, 9d, 11, 17, 24, 30, 31, 36, 38, Tmp

Table 2.2 Genotype statistics for 19 simple sequence repeat loci across 79 *Puccinia graminis* f. sp. *tritici* isolates collected in Washington.

Locus	Parameters[a]						
	No. of genotypes	No. of alleles	No. of rare alleles	N_e	H_o	H_e	Missing data (%)
PgtCAA39	32	11	4	7.2	0.82	0.86	0
PgtCAA49	11	5	1	2.9	0.56	0.66	0
PgtCAA53	12	5	1	3.5	0.59	0.71	1.2
PgtCAA80	8	5	1	2.0	0.44	0.49	0
PgtCAA93	10	7	4	2.2	0.56	0.54	1.2
PgtCAA98	13	7	2	3.8	0.73	0.74	0
PgtCAT4.2	15	7	3	3.8	0.68	0.74	5.1
PgtGAA8.1	14	5	0	4.2	0.80	0.76	6.3
Pgest21	8	4	1	2.8	0.70	0.64	0
Pgest24	7	4	0	2.4	0.70	0.58	3.8
Pgest59	4	3.	1	1.7	0.12	0.42	15
Pgest109	3	3	1	1.4	0.32	0.28	1.2
Pgest142	9	5	2	2.7	0.50	0.62	1.2
Pgest173	5	3	1	2.0	0.30	0.51	0
Pgest227	9	5	2	2.6	0.43	0.62	6.4
Pgest293	10	5	1	3.0	0.68	0.67	0
Pgest318	10	4	0	3.7	0.81	0.73	5.0
Pgest325	9	4	0	2.8	0.46	0.64	3.8
Pgest341	3	2	0	1.9	0.15	0.47	0
Pgest353	3	2	0	1.1	0.03	0.12	1.3

[a] N_e = Effective population size; H_o = Observed Heterozygosity; H_e = Expected Heterozygosity

Table 2.3 Number of genotypic simple sequence repeat allelic differences for pair wise comparisons of *Puccinia graminis* f. sp. *tritici* race groups in Washington.

	Genotypic groups				
Genotypic groups	WA-001	WA-002	WA-003	WA-004	WA-005
WA-001	9				
WA-002	33	0			
WA-003	27	22	9		
WA-004	37	35	24	8	
WA-005	59	64	60	56	92

Table 2.4 R_{st} values of genetic differentiation between race groups of *Puccinia graminis* f. sp. *tritici* based on simple sequence repeat markers in Washington.

	Genotypic groups[a]				
Genotypic groups	WA-001	WA-002	WA-003	WA-004	WA-005
WA-001					
WA-002	0.480*				
WA-003	0.344*	0.380*			
WA-004	0.140*	0.387*	0.229*		
WA-005	0.271*	0.279*	0.102*	0.063	

[a] Asterisk (*) = significant difference at P <0.05.

Table 2.5 Population statistics for genotypic simple sequence repeat markers for the five genotypic groups of *Puccinia graminis* f. sp. *tritici* in Washington[a].

	Parameters[b]							
Genotypic group	No. of samples	No. of alleles (SE)	N_e	H_o	H_e	F	No. of private alleles	No. of genotypes
WA-001	9	1.74(0.168)	1.54(0.128)	0.47(0.114)	0.26(0.062)	-0.64(0.143)	0	5
WA-002	6	1.58(0.116)	1.58(0.116)	0.58(0.116)	0.29(0.063)	-1.00(0.000)	1	1
WA-003	6	1.68(0.154)	1.40(0.117)	0.36(0.104)	0.23(0.058)	-0.54(0.101)	0	3
WA-004	4	1.90(0.215)	1.82(0.190)	0.66(0.109)	0.41(0.068)	-0.87(0.061)	0	4
WA-005	54	4.84(0.467)	2.96(0.305)	0.56(0.045)	0.61(0.038)	0.11(0.054)	39	53

[a] Genotypic population statistic data is shown as an average and standard error across all 20 loci.
[b] N_e = Effective population size; H_o = Observed Heterozygosity; H_e = Expected Heterozygosity; F = Fixation index

Table 2.6 R_{st} values of genetic differentiation between race groups of *Puccinia graminis* f. sp. *tritici* based on simple sequence repeat markers in Washington and the intermountain range region of the United States.

	Genotypic group[a]													
Genotypic group	WA-001	WA-002	WA-003	WA-004	WA-005	NA-001	NA-002	NA-003	NA-004	NA-005	NA-006	NA-007	NA-008	NA-009
WA-001														
WA-002	0.480*													
WA-003	0.344*	0.380*												
WA-004	0.140*	0.387*	0.229*											
WA-005	0.271*	0.279*	0.102*	0.063										
NA-001	0.390*	0.385*	0.333*	0.225*	0.173*									
NA-002	0.420*	0.548*	0.281*	0.303*	0.082*	0.067								
NA-003	0.415*	0.606*	0.366*	0.343*	0.214*	0.276*	0.243*							
NA-004	0.393*	0.362*	0.205*	0.202*	0.154*	0.141*	0.158*	0.282*						
NA-005	0.246*	0.339*	0.077	0.007	0.021	0.161*	0.138*	0.214*	0.095					
NA-006	0.445*	0.466*	0.319*	0.269*	0.221*	0.226*	0.250*	0.388*	0.069	0.188*				
NA-007	0.348*	0.405*	0.129*	0.148*	0.172*	0.299*	0.231*	0.162*	0.061	0.043	0.168*			
NA-008	0.617*	0.453*	0.585*	0.462*	0.438*	0.385*	0.557*	0.608*	0.325*	0.449*	0.469*	0.431*		
NA-009	0.353*	0.576*	0.420*	0.158	0.156*	0.258*	0.459*	0.371*	0.167*	0.126*	0.185*	0.111	0.500*	

[a] Asterisk (*) = significant difference at $P < 0.05$.

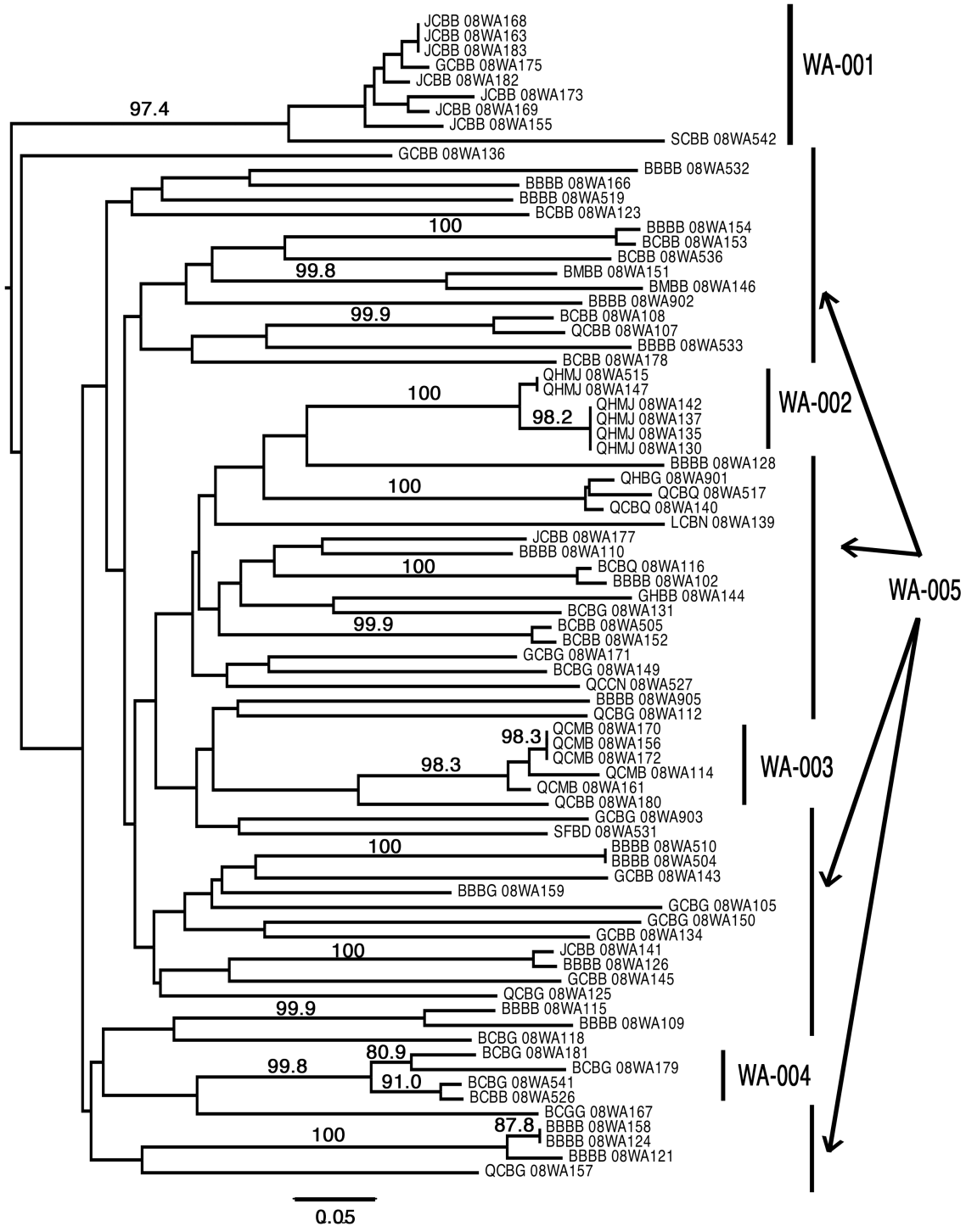

Figure 2.1 Midpoint rooted neighbor joining tree of *Puccinia graminis* f. sp. *tritici* isolates collected in Washington constructed from 19 simple sequence repeat markers. Only branches with bootstrap support greater than 75% are labeled.

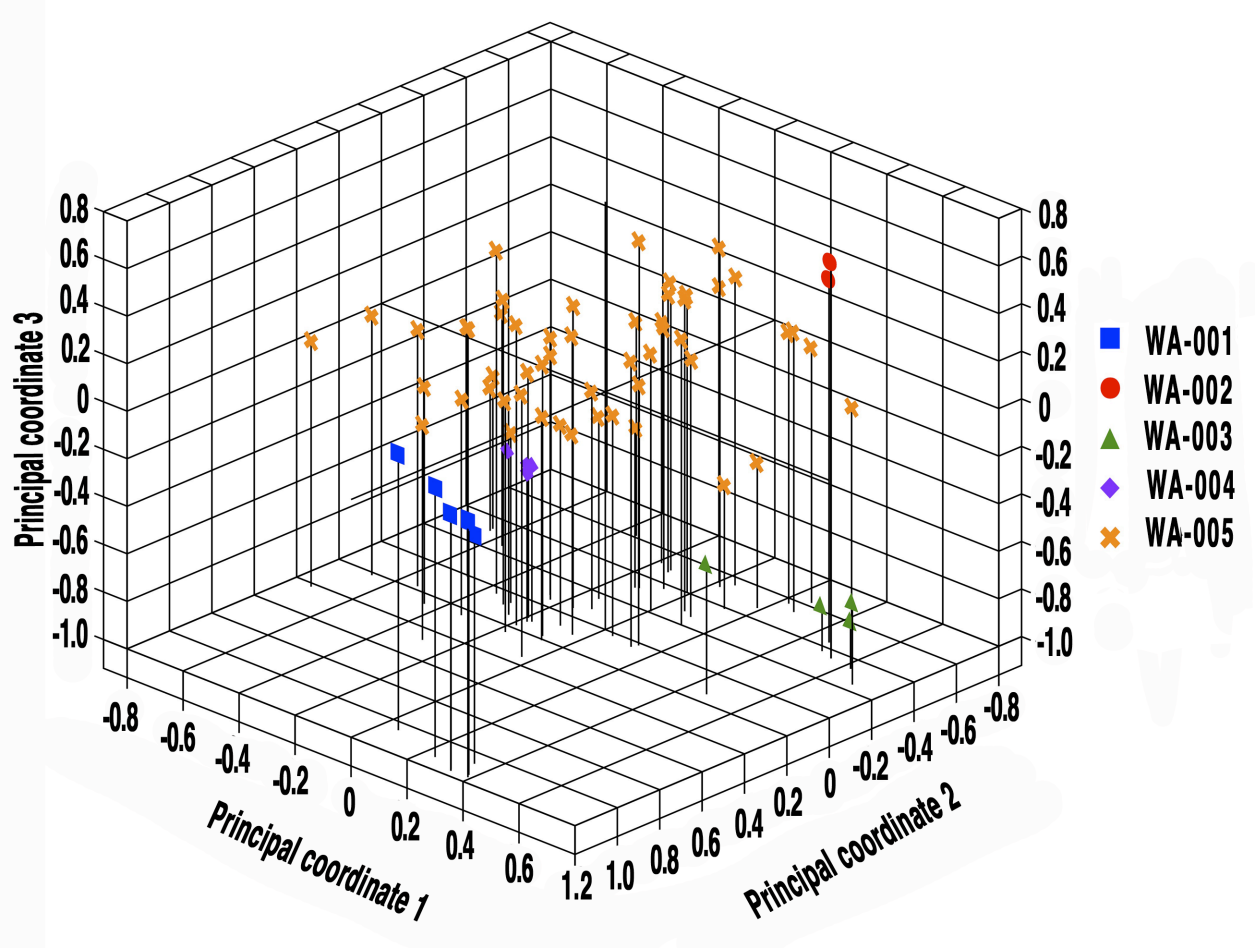

Figure 2.2 Principal coordinate analysis of *Puccinia graminis* f. sp. *tritici* isolates collected in Washington constructed from 19 simple sequence repeat markers.

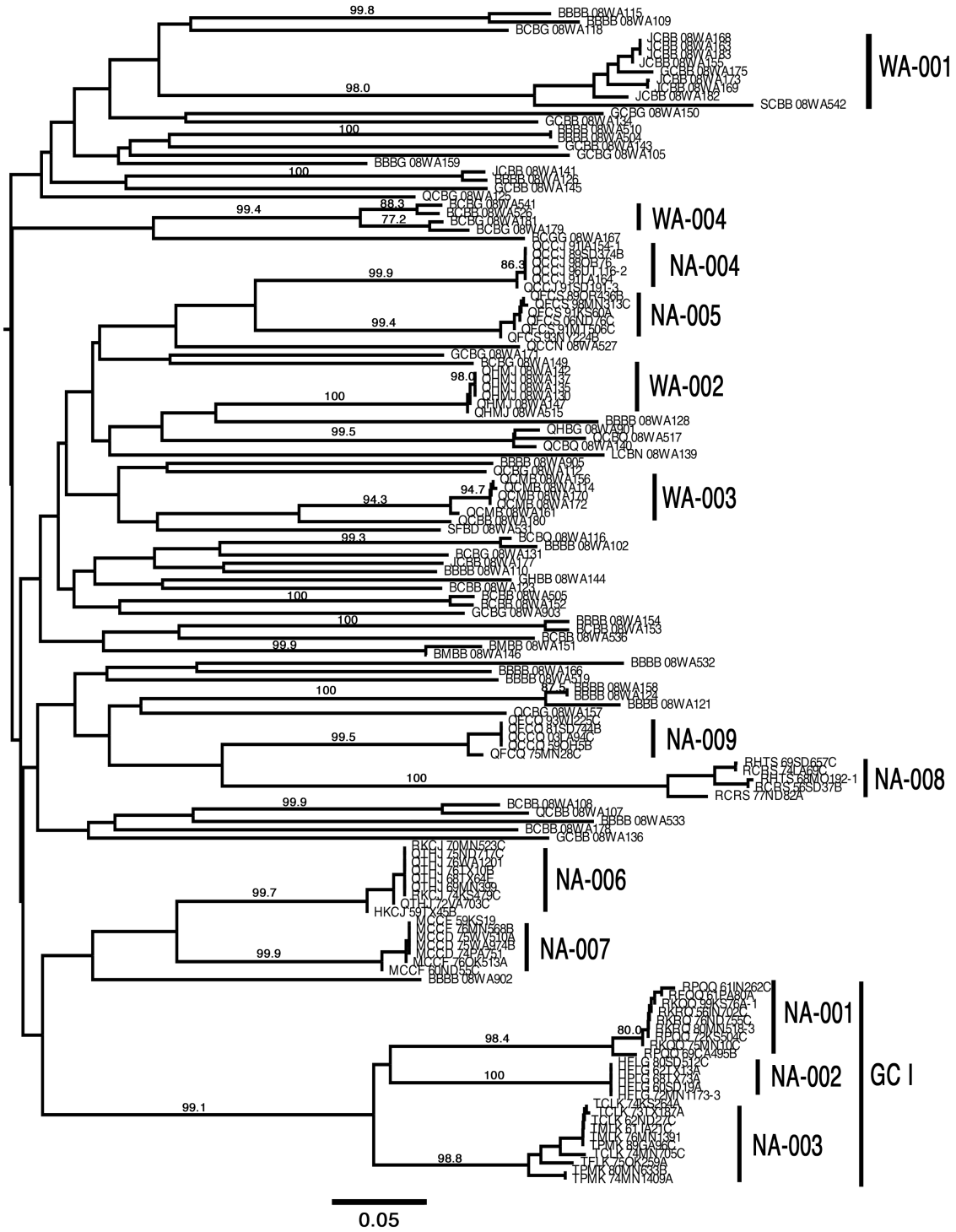

Figure 2.3 Midpoint rooted neighbor joining tree of *Puccinia graminis* f. sp. *tritici* isolates collected in Washington and the intermountain range region of the United States constructed from 19 simple sequence repeat markers. Only branches with bootstrap support greater than 75% are labeled.

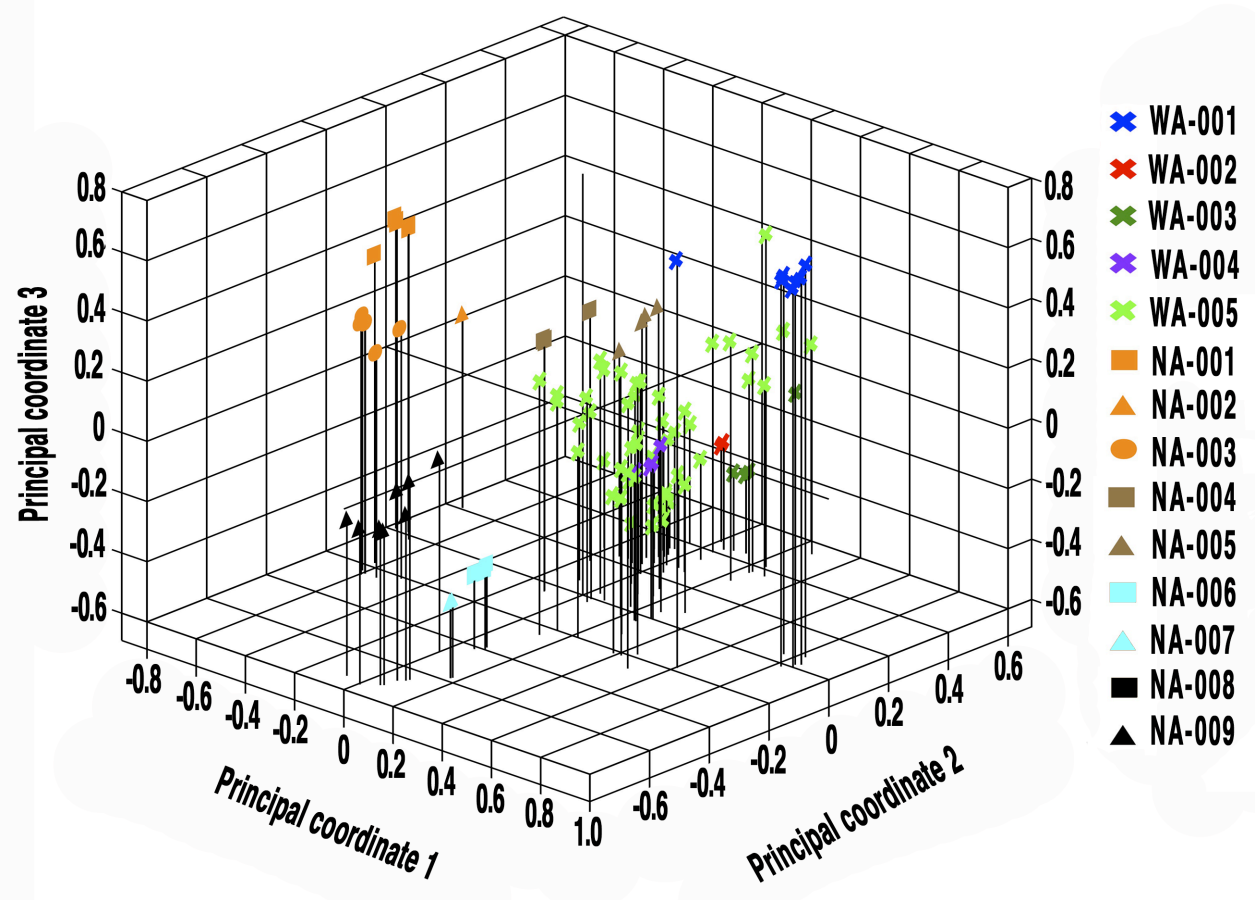

Figure 2.4 Principal coordinate analysis of *Puccinia graminis* f. sp. *tritici* isolates collected in Washington and the intermountain range region of the United States constructed from 19 simple sequence repeat markers.

Bibliography

Alexander H M, Roelfs A P and Groth J V. (1984) Pathogenicity Association in *Puccinia graminis* f. sp. *tritici* in the United States. Phytopathology 74:1161-1166.

Anikster Y. (1984) The formae speciales. In Cereal Rusts, Vol 1. (Bushnell, W R and Roelfs AP eds). Orlando, USA:Academic Press, pp. 115-130.

Anikster Y, Szabo L J, Eilam J, Manisterski J, Koike S T and Bushnell W R. (2004) Morphology, life cycle biology and DNA sequence analysis of rust fungi on garlic and chives in California. Phytopathology 94:569-577.

Barnes I, Gaur A, Burgess T, Roux J, Wingfield B D and Wingfield M J. (2001) Microsatellite markers reflect intra-specific relationships between isolates of the vascular wilt pathogen *Ceratocystis fimbriata*. Molecular Plant Pathology 2:319-325.

Bills R W. (1968) One half century of service: The first 50 Years of barberry eradication. Unpublished Report. USDA ARS Cereal Disease Laboratory, St. Paul, MN.

Boehm E W A, Wenstrom J C, McLaughlin D J, Szabo L J, Roelfs A P and Bushnell W R. (1992) An ultrastructural pachytene karyotype for *Puccinia graminis* f. sp. *tritici*. Canadian Journal of Botany 70:410-413.

Boshoff W H P, Pretorius Z A, van Niekerk B D and Komen J S. (2002) First report of virulence in *Puccinia graminis* f. sp. *tritici* to wheat stem rust resistance genes *Sr8b* and *Sr38* in South Africa. Plant Disease 86:922.

Burdon J J and Roelfs A P. (1985a) Isozyme and virulence variation in asexually reproducing populations of *Puccinia graminis* and *P. recondita* on wheat. Phytopathology 75:907-913.

Burdon J J and Roelfs A P. (1985b) The effect of sexual and asexual reproduction on the isozyme structure of populations of *Puccinia graminis*. Phytopathology 75:1068-1073.

Carleton M A. (1899) *Cereal Rusts of the United States: A Physiological Investigation*. USDA Division of Vegetable Physiology and Pathology Bulletin 16. Washington, D.C.: GPO.

Chester K S. (1946) The nature and prevention of the cereal rusts as exemplified in the leaf rust of wheat. Waltham, MA: Chronica Botanica.

Daszak P, Cunningham A A and Hyatt A D. (2000) Emerging infectious diseases of wildlife – threats to biodiversity and human health. Science 287:443-449.

De Bary A. 1865-1866. Neue Untersuchungen über Uredineen, inbesondere die Entwicklung der *Puccinia graminis* und den Zusammenhang derselbe mit *Aecidium berberidis*. In two parts, 15-49, 205-215. Berlin: Monatsberichte der Berlinder.

Dill-Macky R and Roelfs A P. (1998) The effect of race QCCJ of *Puccinia graminis* f.sp. *tritici* on yield and Quality in Barley. Plant Disease 82:674-678.

Duplessis S, Cuomo C A, Lin Y, Aerts A, Tisserant E, Veneault-Fourrey C, Joly D L, Hacquard S, Amselem J, Cantarel B L, Chiu R, Coutinho P M, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr M G, Kodira C D, Kohler A, Kues U, Lindquist E A, Lucas S M, Mago R, Mauceli E, Morin E, Murat C. Pangilinan J L, Park R, Pearson M, Quesneville H, Roucheir N, Sakthikumar S, Salamov A A, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan G A, Henrissat B, Van de Peer Y, Rouze P, Ellis J G, Dodds P N, Schein J E, Zhong S, Hamelin R C, Grigoriev I V, Szabo L J and Martin F. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. Proceedings of the National Academy of Science 108:9166-9171.

Felsenstein J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Flor H H. (1955) Host-parasite interaction in flax rust- its genetics and other implications. Phytopathology 45:680-685.

Fox S L, Harder D E and Kim W K. (1995) Use of virulence and length variability within the rDNA repeat unit to distinguish isolates of *Puccinia graminis* f. sp. *tritici* race QCC. Canadian Journal of Plant Pathology 17:197-204.

Freeman E M and Johnson E C. (1911) *The Rusts of grains in the United States*. USDA Bureau of Plant Industry Bulletin 216. Washington, D.C.: GPO

Gregerson R O. (1932) History of barberry. Unpublished paper. Barberry Eradication Records, University of Minnesota Archives.

Hamilton L M. (1939) Stem rust in spring wheat area in 1878. *Minnesota History* 20:156-164. Harder D E and Chong J. (1984) Structure and physiology of haustoria. In Cereal Rusts, Vol 1. (Bushnell, W R and Roelfs AP eds). Orlando, USA:Academic Press, pp.431-476.

Jafary H, Szabo L J and Niks R E. (2006) Innate nonhost immunity in barley to different heterologous rust fungi is controlled by sets of resistance genes with different overlapping specificities. Molecular Plant-Microbe Interactions 19:1270-1279.

Jin Y. (2005) Races of *Puccinia graminis* in the United States during 2003. Plant Disease 89:1125-1127.

Jin Y, Szabo L J, Pretorius Z A, Singh R P, Ward R and Fetch T Jr. (2008) Detection of virulence to resistance gene *Sr24* within race TTKS of *Puccinia graminis* f. sp. *tritici*. Plant Disease 92:923-926.

Jin Y, Szabo L J, Rouse M N, Fetch T, Pretorious Z A, Wanyera R and Njau P. (2009) Detection of virulence to resistance gene *Sr36* within the TTKS race lineage of *Puccinia graminis* f. sp. *tritici*. Plant Disease 93:367-370.

Johnson T. (1949) Intervarietal crosses in *Puccinia graminis*. Canadian Journal of Research, Section C. 27:45-65.

Johnson T and Newton M. (1946) Specialization, hybridization, and mutation in the cereal rust. Botanical Review 12:337-392.

Keiper F J, Hayden M J, Park R F and Wellings C R. (2003) Molecular genetic variability of Australian isolates of five cereal rust pathogens. Mycological Research 107:545-556.

Kim W K, Zerucha T and Klassen G S. (1992) A region of heterogeneity adjacent to the 5S ribosomal RNA gene of cereal rusts. Current Genetics 22:101-105.

Kislev M E. (1982) Stem rust of wheat 3300 years old found in Israel. Science, 216:993-994.

Kolmer J. (2001) Early research on the genetics of *Puccinia graminis* and stem rust resistance in wheat in Canada and the United States. In stem rust of wheat: From ancient enemy to modern foe (ED. Peterson P D) pp. 51-82 (APS Press : St. Paul, MN).

Kolmer J A, Jin Y and Long D L. (2007) Wheat leaf and stem rust in the United States. Australian Journal of Agricultural Research 58:31-638.

Lederberg J, Shope R E, Oaks S C. (1992) Emerging infections: Microbial threats to health in the United States, Institute of Medicine, National Academy Press.

Leonard K. (2001) Stem rust- future enemy? In stem rust of wheat: From ancient enemy to modern foe (ED. Peterson P D) pp. 119-146 (APS Press : St. Paul, MN).

Leonard K and Szabo L J. (2005) Stem rust of small grains and grasses caused by *Puccinia graminis*. Plant Molecular Pathology 6:99-111.

Lind J. (1915) *Berberisbusken og Berberislovan* (The Barberry Bush and Its Law). Kobenhaven: Tryke Hos Nielson & Lydiche.

Liu K and Muse S V. (2005) PowerMarker: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128-2129.

Long D L, Kolmer J A, Jin Y and Hughes M E. (2007) Wheat rusts in the United States in 2006. Annual Wheat Newsletter 53:130-141.

Long D L, Kolmer J A, Jin Y, Hughes M E and Wanschura L A. (2005) Wheat rusts in the United States in 2004. Annual Wheat Newsletter. 51:211-220.

Long D L, Kolmer J A, Jin Y, Hughes M E and Wanschura L A. (2008) Wheat rusts in the United States in 2007. Wheat Newsletter 54:182-192.

McCallum B D, Roelfs A P, Szabo L J and Groth J V. (1999) Comparison of *Puccinia graminis* f.sp. *tritici* from South America and Europe. Plant Pathology 48:574-581.

McIntosh R A, Wellings C R and Park R F. (1995) Wheat rusts: An atlas of resistance genes. London: Kluwer Academic Publishers.

Munkacsi A, Kawakami S, Pan J, Lee K, Stoxen S, May G. (2006) Genome-wide assessment of tandem repeat markers for biogeographic analyses of the corn smut fungus, *Ustilago maydis*. Molecular Ecology Notes 6:221-223.

Munkacsi A, Stoxen S and May G. (2008) *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. Proceedings of the Royal Society 275:1037-1046.

Nazari K, Mafi M, Yahyaoui A, Singh RP and Park R F. (2009) Detection of wheat stem rust (*Puccinia graminis* f. sp. *tritici*) race TTKSK (Ug99) in Iran. Plant Disease 93:317.

Oerke E C, Dehne H W, Schonbeck F and Weber A. (1994) Crop production and crop protection: Estimated losses in Major Food and Cash Crops, Elsevier.

Peakall R and Smouse P E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 6:288-295.

Peck J R and Waxman D. (2000) What's wrong with a little sex? Journal of Evolutionary Biology 13:66-69.

Peterson P D, Leonard K J, Roelfs A P and Sutton T B. (2005) Effect of barberry eradication on changes in populations of *Puccinia graminis* in Minnesota. Plant Disease 89:935-940.

Pretorius Z A, Bender C M, Visser B and Terefe T. (2010) First report of a *Puccinia graminis* f. sp. *tritici* race virulent to the *Sr24* and *Sr31* wheat stem rust resistance genes in South Aftica. Plant Disease 94:784.

Pretorius Z A, Singh R P, Wagoire W W and Payne T S. (2000) Detection of virulence to wheat stem rust reeistane gene Sr31 in *Puccinia graminis* f. sp. *tritici* in Uganda. Plant Diseases 84:203.

Roelfs A P. (1978) Estimated losses caused by rust in small grain cereals in the Untied States- 1918-1976. U.S. Department of Agriculture, Agricultural Research Service, Miscellaneous Publication 1363.

Roelfs A P. (1982) Effects of barberry eradication on stem rust in the United States. Plant Disease 66:177-181.

Roelfs A P. (1985) Wheat and rye stem rust. In The Cereal Rusts, Vol. 2 (Roelfs, A.P. and Bushnell, W.R., eds). Orlando, FL: Academic Press, pp 3-37.

Roelfs A P, Casper D H, Long D L and Roberts J J. (1990) Races of *Puccinia graminis* f. sp. *tritici* in the United States and Mexico during 1988. Plant Disease 74:555-557.

Roelfs A P, Casper D H, Long D L and Roberts J J. (1991) Races of *Puccinia graminis* in the United States in 1989. Plant Disease 75:1127-1130.

Roelfs A P and Groth J V. (1980) A comparison of virulence phenotypes in wheat stem rust populations reproducing sexually and asexually. Phytopathology 70:855-862.

Roelfs A P, Long D L and Roberts J J. (1993) Races of *Puccinia graminis* in the United States during 1990. Plant Disease 77:125-128.

Roelfs A P and Martens J W. (1988) An international system of nomenclature for *Puccinia graminis* f.sp. *tritici*. Phytopathology 78:526-533.

Roelfs A P, McCallum B, McVey D and Groth J V. (1997) Comparison of virulence and isozyme phenotypes of *Pgt*-QCCJ and Great Plains races of *Puccinia graminis* f.sp. *tritici*. Phytopathology 87:910-914.

Roelfs A P and McVey D V. (1976) Races of *Puccinia graminis* f. sp. *tritici* in the USA during 1975. Plant Disease Reporter 60:656-660.

Selkoe K A and Toonen R J. (2006) Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615-629.

Singh R P, Hodson D P, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh P K, Singh S, Govindan V. (2011) The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. Annual Review of Phytopathology 49:465-481.

Singh R P, Hodson D P, Jin Y, Huerta-Espino J, Kinyau M. Wanyera R, Njau P, Ward W. (2006) Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. In CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 1(54):1-13.

Singh R P, William H M, Huerta-Espino J, Rosewarne G. (2004) Wheat rust in Asia: meeting the challenges with old and new technologies. In: New Directions for a Diverse Planet: Proceedings of the 4th International Crop Science Congress, Brisbane, Australia, 26 September–1 October 2004. 2004. ISBN 1 920842 20 9. Available from: URL: http:// www.cropscience.org.au./icsc2004/symposia/3/7/ 141_singhrp.htm (verified 1 June 2006)

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

Smith M. (1968) Evolution in asexual and sexual populations. American Naturalist 102:469-473.

Stakman E C. n.d. The war of attrition against wheat stem rust in North America. Unpublished paper. American Phytopathological Society Records, Iowa State University Special Collections, Ames.

Stakman E C and Fletcher D G. (1930) The Common Barberry and Black Stem Rust. USDA Farmers' Bulletin 1544. Washington, D.C.:GPO.

Stakman E C, Stewart D M and Loegering W Q. (1962) Identification of physiologic races of *Puccinia graminis* var. *tritici*. U.S. Department of Agriculture, Agriculture Research Service. E-617 53 pp.

Stakman E C and Levine M N. (1922) The determination of biologic forms of *Puccinia graminis* on *Triticum* spp. Minnesota Agriculture Experimental Station Technical Bulletin 8 pp.

Szabo L J. (2007) Development of simple sequence repeat markers for the plant pathogenic rust fungus *Puccinia graminis*. Molecular Ecology Notes 7:92-94.

Terefe T, Pretorius Z A, Paul I, Mebalo J, Meyer L and Naicker K. (2010) Occurrence and pathogenicity of *Puccinia graminis* f. sp. *tritici* on wheat in South Africa during 2007 and 2008. South African Journal of Plant Soil 27:163-167.

Urban Z. (1967) The taxonomy of some European graminicolous rusts. Ceska Mykologia 21:12-16.

Visser B, Herselman L and Pertorius Z A. (2009) Genetic comparison of Ug99 with selected South African races of *Puccinia graminis* f. sp. *tritici*. Molecular Plant Pathology 10:213-222.

Vurro M, Bonciani B and Vannacci G. (2010) Emerging infectious diseases of crop plants in developing countries: impacts on agriculture and socio-economic consequences. Food Security 2:113-132.

Wanyera R, Kinyua M G, Jin Y and Singh R P. (2006) The spread of stem rust caused by *Puccinia gramins* f. sp. *tritici,* with virulence on Sr31 in wheat in Eastern Africa. Plant Disease 90:113.

Zambino P J and Szabo L J. (1993) Phylogenetic relationships of selected cereal and grass rusts based on rDNA sequence analysis. Mycologia 85:401-14.

Zhong S, Leng Y, Friesen T I, Faris J D and Szabo L J. (2009) Development and characterization of expressed sequence tag-derived microsatellite markers for the wheat stem rust fungus *Puccinia graminis* f.sp. *tritici*. Phytopathology 99:282-289.