Evaluating the presence and introgression of the hybrid forest pathogen *Cronartium x flexili*

by

Kiah Allen

B. Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2019

©Kiah Allen, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Evaluating the presence and introgression of the hybrid forest pathogen *Cronartium x flexili*

submitted by Kiah Allen in partial fulfillment of the requirements for the degree of Master of Science in Forestry

Examinining Committee:

Richard Hamelin, Forest and Conservation Sciences
Supervisor

Nicholas Feau, Forest and Conservation Sciences
Supervisory committee member

Yousry El Kassaby, Forest and Conservation Sciences
Supervisory committee member

Chris Chanway, Land and Food Systems
Additional Examiner
Abstract

The recent discovery of a hybrid forest fungal pathogen, *Cronartium x flexili*, suggested to have arisen from sexual hybridization between the introduced *C. ribicola* and the native *C. comandrae*, was surprising because the parental species do not share hosts. Although the pathosystems of both parental species are well described, the impact of their hybridization is yet unknown. The purpose of this study is to determine the occurrence, level of hybridization and introgression of the hybrid pine stem rust *Cronartium x flexili*. A total of 831 samples from dikaryotic aecia of *C. ribicola* and *C. comandrae* were collected from 13 sites across British Columbia and Alberta and analysed both genetically and morphologically over two sampling seasons. Microscopic and genetic methods, including PCR, qPCR, and genotyping by sequencing (GBS), were used to identify hybrid samples. The results of these analyses indicate that *C. x flexili* is either no longer prevalent in areas where it was previously found, or else prevalent at such low frequencies that it has evaded detection in the sampling effort of this study. Two previously collected and extracted *C. x flexili* DNA samples were examined using a fixed loci analysis and did not demonstrate evidence of introgression, indicating that this hybrid does not facilitate gene flow between the introduced *C. ribicola* and the native *C. comandrae*. This suggests that when a hybridization event occurs between the parental *Cronartium* spp., first generation (F1) hybrids result but are not fertile and do not occur perennially. Some evidence suggests that this could be due to a number of factors including: low relative fitness of *C. x flexili* compared to the parental species if ‘hybrid breakdown’ occurs as explained by the Bateson-Dobzhansky-Muller incompatibility (BDMI) model; variations in local climate factors influencing life cycle parameters; or sexual incompatibility with the parental species. This work adds to the limited
literature on the genetics of hybrid forest fungal pathogens and improves our understanding of the evolutionary mechanisms occurring when allopatrically evolved forest fungal species hybridize.
Lay Summary

The goal of this study is to determine the significance of a hybridization event between two fungal pathogens of forests, the introduced white pine blister rust (WPBR) and the native comandra blister rust (CBR). WPBR is one of the most destructive forest pathogens worldwide, causing mortality of white pine species, while CBR is an important forest pathogen of lodgepole pine both economically and ecologically in the Pacific Northwest. Areas where the hybrid pathogen was previously known to occur were sampled, as well as control sites and additional regions of interest. Samples collected were analysed both genetically and microscopically to identify hybrid specimens and to evaluate genetic parameters of interest. Studying the hybridization between a native and non-native tree rust enhances our understanding of the evolutionary significance of such an event and helps to illuminate one of the potential consequences of introducing foreign pathogens into novel landscapes.
Preface

This dissertation is the original intellectual property of the author, K. R. Allen. The sample collection was the collaborative work of K. Allen, A. Woods, R. Hamelin and R. Reich. DNA extraction, PCR, normalization and preparation for GBS were done by myself in the Forest Pathology Laboratory at the University of British Columbia, Point Grey campus. A. Capron developed the LNA probe, and along with N. Herpin, contributed to the qPCR analysis. The sequences generated by GBS were analyzed by myself, with the help of N. Feau. The sampling maps and figures were generated by myself. R.C. Hamelin was the supervisory author on this project and was involved throughout the project in concept formation and manuscript edits.
# Table of Contents

Abstract .................................................................................................................. iii

Lay Summary .......................................................................................................... v

Preface .................................................................................................................... vi

Table of Contents .................................................................................................... vii

List of Tables .......................................................................................................... ix

List of Figures ......................................................................................................... x

Acknowledgements ................................................................................................. xi

1. Introduction ......................................................................................................... 1

1.1 Literature Review ............................................................................................ 2
    1.1.1 Ecological and Economic impacts of *Cronartium ribicola* ....................... 2
    1.1.2 Ecological and Economic impacts of *Cronartium comandae* ............... 5
    1.1.3 Life Cycle of *C. ribicola* and *C. comandae* ...................................... 6
    1.1.4 Recent Hybridization ............................................................................. 9
    1.1.5 Episodic Selection .................................................................................. 10
    1.1.6 Sexual Hybridization .......................................................................... 11
    1.1.7 Population Dynamics and Importance of Diversity ............................. 13
    1.1.8 Implications for Resistance Breeding ................................................... 14
    1.1.9 Novel Pathogenic Characteristics ......................................................... 16

1.2 Objectives ......................................................................................................... 17

2. Methodology ...................................................................................................... 19

2.1 Sample Collection ........................................................................................... 19

2.2 Microscopy ....................................................................................................... 21

2.3 DNA Extraction ............................................................................................... 21

2.4 Polymerase Chain Reaction (PCR) and Electrophoresis ................................ 22

2.5 Real-time PCR and Locked Nucleic Acid Probe Analysis and Design .......... 22

2.6 Genotyping by Sequencing (GBS) and Library Construction ......................... 23

2.7 Data Analysis .................................................................................................. 24
    2.7.1 Fast-GBS pipeline ................................................................................ 24
    2.7.2 Hybridization index ............................................................................. 25
    2.7.3 Analysis of variance (ANOVA) and Kruskal-Wallis H-test .................... 26
    2.7.4 Probability analysis .............................................................................. 26
    2.7.5 sppIDer pipeline .................................................................................. 27
    2.7.6 RepeatMasker ....................................................................................... 28
    2.7.7 Contamination testing ......................................................................... 29

3. Results ................................................................................................................ 30

3.1 Microscopy ....................................................................................................... 31

3.2 PCR, qPCR and LNA probe analysis ................................................................. 31
3.3 Genotyping by Sequencing ............................................................................................................ 34
  3.3.1 Fast-GBS ................................................................................................................................. 34
  3.3.2 Hybridization Index and ANOVA .......................................................................................... 34
  3.3.3 Probability Analysis .................................................................................................................. 35
  3.3.4 sppIDer ..................................................................................................................................... 36
  3.3.5 RepeatMasker ............................................................................................................................ 44
  3.3.6 Contamination Testing .............................................................................................................. 45

4. Discussion ........................................................................................................................................ 48
  4.1 Presence of *C. x flexili* ................................................................................................................ 48
  4.2 Evidence of Introgression .............................................................................................................. 51
  4.3 Climate/site variation ..................................................................................................................... 53
  4.4 Insects Vectoring Pycniospores .................................................................................................... 54

5. Conclusions ...................................................................................................................................... 56

Literature Cited ..................................................................................................................................... 58

Appendices .......................................................................................................................................... 70
  Appendix A: Summary Table of Hybrid Forest Fungal Pathogens ....................................................... 70
  Appendix B: Sampling Site Location Determination ............................................................................. 71
  Appendix C: Excerpt from Zambino General Rust Protocols (1998) ................................................... 72
  Appendix D: Hu’s DNA Extraction Protocol Modifications ................................................................. 75
  Appendix F: Microscopy ....................................................................................................................... 77
  Appendix G: F- Statistics Calculations .................................................................................................. 78
  Appendix H: Statistical Hypotheses and Equations .............................................................................. 80
  Appendix I: sppIDer Pipeline Output Tables ....................................................................................... 82
List of Tables

Table 1. Summary of the known aecial and telial hosts of each Cronartium ribicola, C. comandrae, and C. x. flexili ........................................................................................................ 8

Table 2. Hybridization index values. ........................................................................................................................................ 24

Table 3. Probability equations and hypotheses used to determine the probability that the alleles observed at one locus are from and interspecific hybridization event .................................................. 26

Table 4. Summary of results of identification of C. x. flexili. ............................................................................................................ 29

Table 5. Results from the reference genome quality analysis performed using custom script in Python (N50, total length, total number of scaffolds) as well as the results of variant calling using the Fast-GBS pipeline (Overall SNPs called and Filtered SNPs). ........................................................................ 33

Table 6. Summary of homozygous and heterozygous loci fixed in both C. x flexili samples and one of the parental species, including hybridization index calculations. ........................................................................ 34

Table 7. Summary of sppIDer output, reporting the mapping output of C. comandrae reads .... 36

Table 8. sppIDer output depicting chi-squared results from mapping C. comandrae reads. .... 36

Table 9. Summary of sppIDer output, reporting the mapping output of C. ribicola reads. ........ 39

Table 10. sppIDer output depicting chi-squared results from mapping C. ribicola reads........ 39

Table 11. Summary of sppIDer output, reporting the mapping output of C. x flexili reads. .... 42

Table 12. sppIDer output depicting chi-squared results from mapping C. x flexili reads. ........ 42

Table 13. Summary of sppIDer output, reporting the percent of C. comandrae reads that map to either the masked C. ribicola genome, the C. comandrae genome or are unmapped. ......................... 44

Table 14. sppIDer output depicting chi-squared results from mapping C. comandrae reads to the masked C. ribicola combination reference genome ........................................................................ 44

Table 15. The top 6 hits returned from blasting a subset of unmapped reads onto the NCBI ‘nr’ database .................................................................................................................................................. 45

Table 16. Mean GC content of mapped and unmapped reads resulting from mapping C. x flexili onto the combination reference genome ........................................................................................................ 46

Table 17. Currently recognized and functionally described hybrid forest fungal pathogens. ..... 70
List of Figures

Figure 1. Life cycle diagram of Cronartium ribicola.......................................................... 7
Figure 2. Sampling site location map................................................................................. 20
Figure 3. Sequence alignment chromatogram of samples collected in 2017.......................... 32
Figure 4. Amplification plots ............................................................................................ 33
Figure 5. Probability distributions depicting the probability P(x) that the observed genotype
arose from an interspecific hybridization event.............................................................. 36
Figure 6. Mapping bar plot with percentage of C. comandrae reads mapped by species onto the
combination reference genome including the unmapped data......................................... 38
Figure 7. Mapping bar plot with percentage of C. comandrae reads mapped by species onto the
combination reference genome excluding the unmapped data......................................... 38
Figure 8. Log2 whole genome mean average depth of coverage of C. comandrae sample reads
mapped onto the combination reference genome......................................................... 39
Figure 9. Mapping bar plot with percentage of C. ribicola reads mapped by species onto the
combination reference genome including the unmapped data......................................... 40
Figure 10. Mapping bar plot with percentage of C. ribicola reads mapped by species onto the
combination reference genome excluding the unmapped data......................................... 41
Figure 11. Log2 whole genome mean average depth of coverage of C. ribicola sample reads
mapped onto the combination reference genome......................................................... 41
Figure 12. Mapping bar plot with percentage of C. x flexili reads mapped by species onto the
combination reference genome including the unmapped data......................................... 43
Figure 13. Mapping bar plot with percentage of C. x flexili reads mapped by species onto the
combination reference genome excluding the unmapped data......................................... 43
Figure 14. Log2 whole genome mean average depth of coverage of C. x flexili sample reads
mapped onto the combination reference genome......................................................... 44
Figure 15. Mean guanine-cytosine (GC) content distribution of both mapped and unmapped
reads from mapping of C. x flexili onto the combination reference genome.................. 47
Figure 16. Scanning Electron Images (SEM) showing aeciospores from (A) C. ribicola (B) C.
comandrae (C & D) C. x flexili collected in the Rocky Mountains of Alberta.................. 77
Figure 17. SEM images showing aeciospores from C. x flexili collected in Mckendrick Pass.... 77
Figure 18. Violin plot depicting mapping quality (MQ) score of C. comandrae reads............. 82
Figure 19. Log2 average depth of coverage of C. comandrae sample reads mapped onto the
combination reference genome.................................................................................. 82
Figure 20. Log2 average depth of coverage of C. comandrae sample reads mapped onto the
combination reference genome.................................................................................. 83
Figure 21. Violin plot depicting the mapping quality (MQ) score of C. ribicola reads............ 84
Figure 22. Log2 average depth of coverage of C. ribicola sample reads mapped onto the
combination reference genome.................................................................................. 84
Figure 23. Log2 average depth of coverage of C. ribicola sample reads mapped onto the
combination reference genome.................................................................................. 85
Figure 24. Violin plot depicting the mapping quality (MQ) score of C. x flexili reads............ 86
Figure 25. Log2 average depth of coverage of C. x flexili sample reads mapped onto the
combination reference genome.................................................................................. 86
Figure 26. Log2 average depth of coverage of C. x flexili sample reads mapped onto the
combination reference genome.................................................................................. 87
Acknowledgements

I would like to express my deep gratitude to Dr. Hamelin, Dr. El-Kassaby and Dr. Feau, my research supervisors, for their guidance, creative problem solving and useful critiques of this research work. I would like to express my very great appreciation to Dr. Feau for his invaluable mentorship and support during the entirety of this research work. His willingness to give his time so generously has been very much appreciated. I would also like to extend my thanks to the members of the Hamelin laboratory, especially Dr. Capron and Padmini Herath, without whose mentorship I would not have been able to accomplish this project. I wish to thank various people for their contribution to the field collections of this project, Alex Woods, Richard Reich, and my undergraduate research/field assistants Naomie Herpin-Saunier, Kenny Forbes, Janek Knowicki and Ethan Schnell.

Special thanks should be given to both of my direct supervisors, Dr. Hamelin and Jason Kipp, for their support and flexibility in allowing me to pursue my studies as well as my career with the B.C. Wildfire Service.

Finally, I wish to thank my parents and my oldest friend, Taylor Wale, for their support and encouragement throughout my study.
1. Introduction

Forest pathogens, particularly alien ones, have emerged as a major threat to forest ecosystems in recent decades (Dukes et al., 2009). Climate change, along with globalization, is influencing forest and plant pathogens worldwide through facilitating their spread and establishment to new areas (Gonthier et al., 2007). This, in turn, has led to novel pathosystem interactions, including increased invasive species establishment and hybridization between species (Allendorf et al., 2001). The impact of such hybridization events is yet to be determined, but in some cases increased host range or pathogenicity have been demonstrated (Newcombe et al., 2001; Brasier, et al., 1998).

There are different mechanisms associated with hybridization in fungi, which can occur through asexual or sexual processes. Asexual hybridization occurs vegetatively through the transfer of nuclear DNA fragments, mitochondria, plasmids, and viruses; while sexual hybridization occurs through the transfer of chromosomes via sexual cells (Schardl & Craven, 2003). Hybridization can also occur within a species (intraspecific) or between species (interspecific) (Taylor et al., 2000). For interspecific hybridization to occur, the two parents must be determined as distinct species through pre-established criteria, which can be morphological, biological and phylogenetic (Schardl & Craven, 2003). Importantly, hybridization can act as a mechanism for speciation if the hybrid has high fitness and is self-propagating (Brasier, 1995).

Fungal pathogens are particularly interesting in an evolutionary sense due to their suitability to undertake rapid microevolution (Brasier, 1995). Barrett et al. (2007) note that plant pathogens are able to evolve new virulence at an alarming rate. This has been attributed to the relatively short generation times of many rust and other fungi species (Barrett et al., 2007), which can be mere weeks or months (Helfer, 2013). These short generational times, in combination with
large numbers of propagules, give fungi an evolutionary advantage to selection pressures over their plant host species (Helfer, 2013). Many plant hosts have much longer generation times, several years for most tree species and smaller population effective sizes which results in a reduction of fitness as they have fewer opportunities for the mechanisms of evolution to act (Helfer, 2013). However, this is dependent on the mode of reproduction, if the pathogen reproduces asexually then it might have less evolutionary potential than a recombinant species, regardless of short generation times (e.g. Ophiostoma ulmi in some regions) (Brasier, 1998). Brasier (2000) identify interspecific hybridization as a means to quicken evolutionary processes via combination of genomes.

Although many agricultural pathogens and pathosystems have been studied relatively intensively, forest fungal pathogens, and particularly their hybridization, has hardly been explored (Brasier, 1995). Human-influenced hybridization is pertinent to conservation biology as hybrid pathogens may result in unpredictable implications for forest ecosystems (Lamour, 2013). My thesis project looked at the impacts of a fungal hybridization event that has occurred in western North America between two closely related pathogenic species; the introduced species Cronartium ribicola J. C. Fisch (Basidiomycota, Pucciniales), and the native Cronartium comandrae Peck (Basidiomycota, Pucciniales). Cronartium rusts are some of the most impactful fungal tree pathogens in North America; these include economic impacts on the forest sector and disturbances to ecosystem functions (Loo et al., 2009; Woods et al., 2000).

1.1 Literature Review

1.1.1 Ecological and Economic impacts of Cronartium ribicola

White pine blister rust (WPBR), caused by the fungal pathogen Cronartium ribicola, was introduced to North America from Europe and Asia in the 19th century and is responsible for one
of the most devastating forest disease outbreaks worldwide (Hunt, 2009). This pathogen attacks
the stems and branches of five-needle *Strobus* (white) pines and alternates on *Ribes, Pedicularis*
and *Castilleja* spp. (McDonald et al., 2006). This has both ecological and economic significance
in the Pacific Northwest, where the climate, distribution of white pines and abundance of
alternate hosts has resulted in an ideal situation for *C. ribicola* establishment (Tombback &
Achuff, 2010).

White pines are widespread throughout both Western and Eastern North America, they
have a vast geographic distribution and occur throughout a large elevation gradient (Tombback &
Achuff, 2010). Tomback and Achuff (2010) note that at least one species of white pine occurs in
a majority of coniferous forests in the Pacific Northwest. These pines are primary succession
species, that is, they regenerate rapidly following natural disturbance such as wildfire (Tomback
& Achuff, 2010). In high elevation sites with extreme environmental conditions white pines are
crucial as primary colonizers to maintain appropriate microclimate conditions for secondary
succession species to establish (Tomback et al., 2016). Thus, *Pinus* species in the subgenus
*Strobus* are critical in providing opportunities for increased biodiversity and ecological benefits
for these alpine ecosystems (Tomback et al., 2016).

White pine blister rust, along with fire suppression, mountain pine beetle (*Dendroctonus
ponderosae* Hopkins) and climate change, have resulted in serious threats to white pine
ecosystems in Northwestern North America (Gelderman et al., 2016). White pine blister rust can
be considered the most significant factor in their decline; the disease can cause malformation,
reduced vigour and death of trees and seedlings, in some cases causing 100% mortality rates of
white pines in high hazard zones (Brar et al., 2015). Projected temperature and precipitation
regimes will likely enhance these stressors and lead to a mismatch between the environment that
the trees are currently in, and the environment that is suitable for their survival (Tomback et al., 2016). WPBR is responsible for whitebark pine to be the first tree species to be listed as endangered on both the Species at Risk Act (SARA) and the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) in Canada, as well as a candidate to be listed on the US Endangered Species Act (Gelderman et al., 2016).

The great losses of white pine species have not only affected ecological systems, but have also had a major impact on harvest and export of white pine timber. From 1880 to approximately 1937, logging of white pines climbed steadily, peaking in 1937, and has been continuously declining until today where western white pine has been nearly excluded from the market (Tomback & Achuff, 2010). This is due to the aforementioned causes, thus the replanting of white pine has been very limited and has instead been replaced by other conifer species (Harvey et al., 2008).

Fortunately, genetic resistance to WPBR through a single dominant resistance \((R)\) gene has been found for sugar, western white, southwestern white, and limber pine, labelled \(Cr1\), \(Cr2\), \(Cr3\) and \(Cr4\) genes respectively (Schoettle et al., 2014; Kinloch et al., 2002; 1977; 1999). This is known as vertical resistance, referring to the plant’s exhibition of high resistance to an individual pathotype, which is often controlled by one gene (Burdon, 2001). Rust-resistance breeding programs have been established to allow different species of white pines to be once again planted in regeneration programs (Bingham, 1983; Tomback & Achuff, 2010). However, the level of resistance in progeny has a wide range (Bingham, 1983) and resistance can be overcome by the pathogen through different evolutionary mechanisms which result in the pathogen evolving new virulent strains (McDonald & Linde, 2002).
1.1.2 Ecological and Economic impacts of *Cronartium comandrae*

Comandra blister rust, caused by *C. comandrae*, is a North American native fungal pathogen that attacks two and three-needle pines (eg, lodgepole and Jack pines) and alternates on bastard toadflax (*Geocaulon* spp.) (Jacobi et al., 2002). It has a life-cycle nearly identical to that of *C. ribicola*, only involving different telial and aecial host species.

The most notable impacts from *C. comandrae* have been on lodgepole pine (*Pinus contorta* var. *latifolia* Engelm. ex S. Watson) as it is a dominant tree species in the Pacific Northwest of North America (Petrie et al., 2016). This tree species is widespread, occurring from low elevation to subalpine zones, as well as from the coastal to the Rocky Mountain regions (Coops & Waring, 2011). Lodgepole pine is a primary succession species, and will quickly regenerate following wildfire or other disturbance, has a rapid growth rate and produces quality timber (Coops & Waring, 2011).

Due to these factors, *P. contorta* accounts for a large percentage of total harvest and is a favourite species to be planted for tree regeneration in British Columbia (Reich et al., 2015). Like whitebark pine, lodgepole pine stands have been decimated in more recent years, predominantly by the mountain pine beetle and frequent high intensity wildfires, but also by fungal pathogens (comandra blister rust, stalactiform blister rust *Cronartium coleosporioides* Arthur and western gall rust *Endocronartium harknessii* (J.P Moore) Y. Hiratsuka) (Reich et al., 2015). The timber losses due to comandra blister rust outbreaks have been associated with harvesting methods (Reich et al., 2015). The nature of current day regeneration strategies, that is, plantations of rapidly growing, young, mono-species and even-aged stands, result in environments very conducive to pathogen establishment (Blenis & Li, 2005).
The disruption from a balanced plant-pathogen interaction is important from both an economic and ecological perspective, as comandra blister rust infections that reach the stem are known to be nearly always lethal through stem girdling (Hiratsuka et al., 1988). A study by Reich et al. (2015) showed that mortality caused by predominantly comandra blister rust, but also western gall rust, accounted for a majority of the lodgepole pine volume loss in stands up to 24 years of age in the interior of British Columbia. An earlier study by Woods et al. (2000) found there to be up to 15% rotation-length volume losses from *C. comandrae* infections, moreover this has been predicted to be an underestimate. As there is the possibility that *C. comandrae* outbreaks could worsen in association with the influence of climate change on environmental factors (Pike, Bennett, Redding, Werner, & Spittlehouse, 2010; Reich et al., 2015) the impact of *Cronartium* rusts is a critical factor in the future of forest health in the Pacific Northwest.

### 1.1.3 Life Cycle of *C. ribicola* and *C. comandrae*

Both *C. ribicola* and *C. comandrae* have similar life-cycles but differ in the host species involved. Both pathogens are heteroeccious, that is, they require both a telial and an aecial host to complete their life cycle (Leung & Kot, 2015). The life cycle of these pathogens is rather complex (Figure 1), and requires a number of spore stages on the different host plants (Leung & Kot, 2015).
Sexual reproduction in the Uredinales (rust fungi) occurs in the haploid pycnia (spermogonia) produced on the pycnial host plants, which in the case of *C. ribicola* and *C. comandrae* is also the aecial host tree species (Anikster et al., 1999). Pycniospores are the reproductive structures, consisting of dispersing spermata and receptive hyphae, which act as gametes and are used to form dikaryotic aeciospores around 3-5 years after infection, through fertilization by insects (Maloy, 1997). Spermogonia form at the margin of a canker and produce pycniospores in pycnial nectare, a substance with high sugar concentration, which attracts insects and thus facilitates distribution of pycniospores from one cluster to another (Anikster et al., 1999). This process occurs in late summer or early fall, which allows for dikaryotic hyphae to form aecia the following year. Aeciospores on the same canker can be half-sibs, that is, they can arise through multiple fertilization events via the haploid pycniospores (Gitzendanner et al., 1996).

Aeciospores which are produced on infected pine stems or branches, are released in spring and, as they are airborne, can only infect the telial host species (Mielke, 1943). These spores can travel long distances, in some cases more than 1000 km in large wind events (Frank et al., 2008). Once infected, the telial host will have a latency period of around 14 days, after which it will release urediniospores, which is the repeating stage and can infect other telial hosts but over a
much shorter distance than aeciospores (Leung & Kot, 2015). Later in the summer, the infected telial host produce diploid teliospores, which only function to form telia, followed by creating basidia and basidiospores through meiosis (Frank et al., 2008). The basidiospores can infect the aecial host (pine) through the stomata of their needles, but these spores are limited by their short dispersal capacity (Frank et al., 2008). The successfully infected pine trees then produce haploid cankers, which form pycniospores (or spermatia) to complete the cycle.

After the pathogen enters the aecial host through the stomata, the infection can spread 5 cm per year towards the branch and stem of the tree (Schwandt et al., 2013). As the infection spreads, foliage will die and more cankers may emerge (Hunt et al., 2007). Once the infection reaches the stem and girdles it, the outcome will be either stem girdling, which can result in branch death or top kill if the infection is on the main stem (Leung & Kot, 2015). The infection in the telial host causes a reduction of photosynthesis due to loss of photosynthetic tissue; teliospores and basidiospores will continue to be produced until leaf drop in the fall, whereby the plant can shed the disease. Thus, the infection is annual in the telial host (Leung & Kot, 2015) but perennial in the aecial host.

Table 1. Summary of the known aecial and telial hosts of each Cronartium ribicola, C. comandrae, and C. x. flexili

<table>
<thead>
<tr>
<th></th>
<th>Cronartium ribicola</th>
<th>Cronartium comandra</th>
<th>Cronartium x flexili</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aecial Hosts</strong></td>
<td><em>Pinus</em> subsection <em>Strobus</em> spp. 5-needle pines Primarily whitebark, sugar, western and eastern white, limber, and Swiss stone pine</td>
<td>2 and 3 needle pines Primarily <em>P. contorta</em> and <em>P. ponderosa</em></td>
<td>Limber pine (<em>P. flexili</em>) Whitebark pine (<em>P. albicaulis</em>)</td>
</tr>
<tr>
<td><strong>Telial Hosts</strong></td>
<td><em>Ribes</em> spp. (currants and gooseberries) <em>Castilleja</em> spp. (paintbrush) <em>Pedicularis</em> spp.</td>
<td>Toadflax (<em>Geocaulum lividum, Comandra umbellata</em>)</td>
<td>Telial host(s) unknown</td>
</tr>
</tbody>
</table>
1.1.4 Recent Hybridization

The recent discovery of *Cronartium x flexili*, a hybrid between *C. ribicola* and *C. comandrae* was surprising because they are not known to share any hosts. The discovery was made by Joly et al. (2006) who genotyped aeciospore samples collected from 3 widely separated stands of *Pinus flexilis* using 12 codominant polymerase chain reaction-based markers. The stands sampled were in the Rocky Mountains of Alberta, more specifically at sites in Waterton Lakes National Park, Kananaskis Country and Porcupine Hills (Joly et al., 2006). Genetic tests revealed heterozygosity and novel alleles at all loci in up to 29% of the samples from each sample site over 2 years (2003 and 2004). Subsequent analysis showed that this was due to heterozygosity within each of the 12 loci containing one allele from *C. ribicola* and one from *C. comandrae*. Intermediate spore morphology of the hybrid aeciospores between the parental spore types, as revealed via scanning electron microscopy (SEM) provided further evidence of the hybridization event (Joly et al., 2006). Joly et al. (2006) were able to rule out the possibility of contamination by observing that only *C. ribicola* cytoplasmically-inherited mitochondrial alleles were present in the hybrid samples. They hypothesized that this interspecific hybridization event occurred due to the proximity of *C. comandrae* infected lodgepole pines to *C. ribicola* infected limber pine (*Pinus flexilis* James), allowing for spermatization of *C. ribicola* receptive hyphae by *C. comandrae* pycniospores, forming hybrid aeciospores.

Although the pathosystems of both parental species are well described, the impact of their hybridization is yet unknown. Hybrid pathogens are known to jump hosts in unpredictable ways and can be subjected to episodic selection and speciation (Brasier et al., 2004). This hybrid rust has been found on the *C. ribicola* hosts limber pine and on whitebark pine in a separate hybridization event in 2012 near Mckendrick Pass near Smithers B.C. (Hamelin, R., personal
communication, January 2017). It is unknown if it occurs on any of the known alternate (telial) hosts (Joly et al., 2006). The characteristics of the hybrid rust including occurrence, distribution, survivability, viability and host range are currently unknown (Joly et al., 2006). Improving our understanding of hybridization between a native and non-native tree rust, such as *C. x flexili*, will help to address disease outbreaks in terms of forest conservation as well as invasive species considerations and management.

1.1.5 Episodic Selection

The effects of increased globalization and specifically global trade has resulted in the transfer of exotic species worldwide (Gonthier et al., 2007). This has implications not only for the biosecurity of the world’s forests but also may facilitate opportunities for episodic selection and interspecific hybridization to take place (Pyšek et al., 2011). Episodic selection occurs when sudden environmental changes cause large variability in a species’ population structure (Brasier, 1995). There are a number of circumstances that will lead to episodic selection, including novel species (including competitors) or host introductions, a flux in the quality or availability of resources, and environmental changes (Brasier, 1995). Predominantly, the abrupt movement of a pathogen across the landscape from its current biogeographic environment to a novel biogeographic environment can result in rapid genetic change in order to adapt to a new host and compete with already-established species (Gonthier et al., 2007). With the influence of globalization, the prevalence of episodic selection creating hybrid forest pathogens is likely more common than ever before (Helfer, 2013).

Brasier (2000) notes that fungal pathogens native to the same region are less likely to undertake interspecific hybridization due to genetic barriers. He attributed this to fungal pathogens having greater fitness if they can sustain genetic barriers which maintain gene combinations
optimizing adaptation to its own ecological niche, as well as avoid spreading detrimental genetic elements between species. These genetic barriers can be weaker or non-existent in species that are separated geographically (Brasier, 2000; Stukenbrock, 2016). Thus, when previously allopatrically isolated fungi come into contact with one another it can result in hybridization due to the evolutionary advantages that result when the pathogens are in competition for the same ecological niche (Brasier, 1995). Hybridization could be more likely to occur in situations where species of the same genus, or which share a similar host, come in contact in a novel landscape, as could occur between a native and an introduced pathogen (Brasier, 1995).

In fact, Giordano et al. (2014) suggest that in the case that sibling (related) species are introduced to one another, which have been separated geographically for an extended amount of time and have allopatrically evolved, mating and possible hybridization between the species is distinctly possible. Joly et al. (2006) hypothesized that this occurred between Cronartium comandrae and Cronartium ribicola to form C. x flexili. Cases of introduced forest pathogens going through episodic selection and hybridizing with native species, such as could be the case for C. x flexili, likely occur more frequently than current literature suggests (see Appendix A) (Brasier et al., 1995).

1.1.6 Sexual Hybridization

Sexual hybridization occurs when species of fungi are sexually compatible; male and female structures are able to interact, which allows for dikaryotic cells to be produced with genetic material from each parent (Scharld & Craven, 2003). In the phyla Basidiomycota and Ascomycota, cells continue to grow while maintaining nuclei of each parent (Scharld & Craven, 2003). After some time, which varies by species (indefinite for Basidiomycota), the dikaryon fuse and undertake meiosis, the final product of which is haploid spores (Scharld & Craven, 2003). In
comparison to species which reproduce asexually, considerable genetic diversity is created when sexual recombination occurs (Park & Wellings, 2012), and thus increases the evolutionary potential of a species (Brasier, 2000). However, the results of hybridization are extremely variable in terms of pathogenicity; sexual recombination has both increased virulence in progeny (Park & Wellings, 2012) as well as decreased (Green, 1971).

Interspecific hybrids can be difficult to distinguish from the parental populations when the species that are hybridizing are closely related (Schardl & Craven, 2003). Sexual hybridization can result in introgression in which case backcrossing between hybrids and parents results in new genes being passed on to either parent population (Brasier, 1995). When this occurs the genotypes of certain hybrids could be interpreted as the two distinct parental groups existing as a single interbreeding species, thus the hybrids would be overlooked (Schardl & Craven, 2003). This was particularly problematic when a small number of markers were used to conduct genetic studies. Recent advances in molecular techniques have improved identification proficiency (Allendorf et al., 2001).

Furthermore, the method of hybridization can be hard to determine in fungal pathogens, as is the case for the hybrid Melampsora medusae-populina (Spiers & Hopcroft, 1994). The two parental species, M. medusae and M. larici-populina, which infect various Populus species and alternate hosts, were naturally introduced into Australia and subsequently hybridized. The authors hypothesized that the hybrid may have arisen via anastomosis, or through sexual means by the crossing of pycnia on a common alternate host, although they did not identify a common host at the time of the study (Spiers & Hopcroft, 1994). M. medusae-populina has intermediate spore structure and phenotype of the parental strains. However, it has pathogenicity more similar to M. medusae than M. larici-populina, and has not increased host range or virulence (Spiers & Hopcroft,
1994). Both parent species demonstrate variations in their host range and virulence, randomly infecting hosts which they attack irregularly. This *Melampsora* spp. hybridization event provides an example of rust hybridization between two relatively divergent parents (Vialle et al., 2013).

1.1.7 Population Dynamics and Importance of Diversity

McDonald and Linde (2002) outline key elements that contribute to pathogen population genetic structure and thereby evolutionary potential. They define genetic structure as both the quantity and distribution of genetic diversity within and among populations, while genetic diversity is composed of both gene diversity, and genotype diversity. Gene diversity is the number and frequencies of alleles at a single locus found in a population; meanwhile, genotype diversity is the number and frequencies of genotypes of multiple loci in a population (McDonald & Linde, 2002). Genotype diversity is more important than gene diversity for pathogens that reproduce entirely or mostly asexually due to most of their genetic diversity being distributed amongst the different clonal lines (McDonald & Linde, 2002). In asexual lineages genotype flow can occur whereby whole genotypes can be exchanged between these lines; McDonald and Linde (2002) predict that this genotype flow could pose a higher risk than gene flow as genotype flow represents a group of coadapted alleles which have already undergone selection. It is especially important to describe hybrid populations by the abundance of hybrid genotypes, since describing them by allele frequencies can be misleading (Allendorf et al., 2001). Hybridization can increase genetic diversity through genotype/gene flow, which is linked to the potential of a population to evolve, and the pathogens with the greatest evolutionary potential are the most likely to overcome resistance genes or other control methods such as fungicides (McDonald & Linde, 2002).

Hybridization is important from an evolutionary sense in terms of creating novel species (Brasier, 1995). If fungi hybridize via asexual or sexual means followed by chromosomes
doubling, a novel polyploid species can be produced (Olson & Stenlid, 2002). This new species becomes isolated from the parental species because of incompatibility in chromosome number. However, even when chromosome doubling does not occur, speciation can occur following hybridization. This phenomenon is described by Husson et al. (2015) using the Phytophthora alni species complex. As hybridization and speciation in fungi have not been greatly explored, the amount that hybridization contributes to speciation in fungi is likely underestimated (Olson & Stenlid, 2002).

1.1.8 Implications for Resistance Breeding

Fungi can use hybridization as a way to evolve new species to overcome resistant genes (Olson & Stenlid, 2002). Pathogen population dynamics are complex, some species have well-documented boom and bust cycles, especially in cereal rusts (McIntosh & Brown, 1997). What makes some pathogens particularly suitable to be components in these cycles is the presence of ‘gene-for-gene interactions’ with their hosts (McDonald & Linde, 2002). In a gene-for-gene interaction, single-gene resistance (also known as major gene resistance) in the host population is spread over a large geographic area (boom), followed by adaptation in the pathogen population to overcome the resistance gene (bust) (McDonald & Linde, 2002). The pathogens are able to overcome the resistance gene through genetic means, such as favourable mutations, recombinants, introductions with greater adaptations to the resistant plant host, and possibly hybridization (McDonald & Linde, 2002).

McDonald and Linde (2002) describe gene-for-gene interaction as the plant recognizing the elicitor molecules that the pathogens produce when they attack. Given a plant cell recognizes the elicitor, it leads to programmed cell death and stops the pathogen from infecting further (McDonald & Linde, 2002). However, if there is a mutation in the pathogen that transforms or
stops production of the elicitor molecules, the plant host will not recognize the invading pathogen and the pathogen will become virulent on the host. A larger scale breakdown in resistance of the host will occur if pathogen strains with the favourable mutation causing virulence are able to increase in frequency; a positive feedback will occur and virulent mutants will continue to be able to infect and reproduce as until host defense systems can evolve resistance (McDonald & Linde, 2002). Once selection pushes the virulence mutation above a certain threshold the resistance gene in the plant will no longer be effective and the pathogen has ‘broken’ resistance (McDonald & Linde, 2002). Hybrid pathogens could have the ability to overcome resistance through recombination, rather than through mutations, potentially increasing the possibility to overcome bred or natural resistance in host populations.

For *C. ribicola*, the major gene resistance (MGR) genes *Cr1* and *Cr2* found in sugar and Western white pines, which trigger programmed cell death when exposed to the pathogen, can be overcome by new virulent strains evolved through this gene-for-gene interaction mechanism (Kinloch et al., 1999). These strains have *vcr1* and *vcr2* genotypes respectively (Kinloch et al., 1999). However, these two virulent genotypes are specific to the genes found in each respective host, that is *vcr1* is avirulent on *Cr2* and *vcr2* is avirulent on *Cr1* (Vogler et al., 2005). This has obvious implications for the breeding program established to help restore sugar and white pines, and suggests that the other two resistant genes, *Cr3* and *Cr4* found in Southwestern white pine and limber pine could be overcome by virulent genotypes as well (Vogler et al., 2005).

An example where a hybrid pathogen overcame resistance genes in planted populations is *Melampsora x columbiana*, a hybrid poplar rust pathogen between *Melampsora medusae* and *Melampsora occidentalis* (Newcombe et al., 2001). *M. x columbiana* was identified as a hybrid via its uredinia and telia having intermediate morphology of the parental forms and samples often
containing internal transcribed spacer (ITS) sequences from both parental species (Newcombe et al., 2001). *M. x columbiana* was found to be pathogenic on *Populus* clones which were bred with resistance to *M. medusae*. *M. medusae* alternates between *Populus deltoides* and eastern larch (*Larix laricina*), whereas *M. occidentalis* alternates between *Populus trichocarpa* and Douglas fir (*Pseudotsuga menziesii*). Interestingly Newcombe et al. (2001) found thirteen distinct pathotypes of *M. x columbiana*, thus supporting the hypothesis originally put forth by Brasier (2000) that the hybrid *M. x columbiana* has new host specialization in comparison to its parental species. This example is important from an evolutionary stance as it demonstrates the ability of hybrids to not only overcome resistant genes but also to have vastly different pathotypes than their parental species.

### 1.1.9 Novel Pathogenic Characteristics

Hybrids have the potential to rapidly evolve new virulent combinations allowing them to be more aggressive and expand their host range in comparison to their parental species (Dukes et al., 2009). A considerable number of the hybrid forest fungal pathogens known (Appendix A) have dissimilar host specialization and virulence than their respective parental species. Two aforementioned forest hybrids: *P. alni* and *M. X. columbiana* are examples of this phenomenon (Husson et al., 2015; Newcombe et al., 2001). These ‘host jumps’ cause issues in the discovery and prediction of virulent hybrid outbreaks as variants can easily be overlooked or can establish without attracting notice until they have reached high infestation levels.

There are various examples where hybrid progeny have lower overall fitness than their parental strains. Green (1971) found that the hybrid offspring between the cereal rusts *Puccinia graminis tritici* and *Puccinia graminis secalis* were less virulent than either parent and thus not considered a danger to cereal crops. However, he did find that they were able to infect a greater
Green (1971) hypothesized that the hybrids are more similar to primitive *Puccinia graminis* and that the parents represent the progression in cereal rusts from attacking many hosts with low success to attacking few hosts with high virulence. This evolutionary progression hypothesized by Green (1971) could be at play in naturally occurring hybrids which display low virulence to many different host species, but could become more specialized over time.

Similarly, in the case of Dutch Elm disease, which attacks elm trees worldwide (*Ulmus* spp.), the hybridization of two closely related pathogens (*Ophiostoma ulmi* and *O. novo-ulmi*) created progeny with lower fitness than either parental species (Brasier, 1998). Dutch Elm disease has broken out to epidemic proportions twice in recent history, the first of which occurred in the early 1900s and was caused by *Ophiostoma ulmi*, the second of which is caused by the much more virulent *Ophiostoma novo-ulmi* (Brasier, 1998). The two species evolved separately, but over time came in contact, forming natural hybrids. Samples from two locations (Portugal and Poland) were analyzed confirming the genetic hybridization of the species, however, the hybrid was determined to have low fitness due to its rare occurrence and low female fertility (Brasier, 1998). Though, Et-Touil et al (1999) noted that it is possible that the hybrid species could facilitate the transfer of genes between *O. ulmi* and *O. novo-ulmi*, which has future implications for genetic diversity and adaptation of the population.

### 1.2 Objectives

My research objective is to determine the occurrence, level of hybridization and introgression of the hybrid pine stem rust *Cronartium x flexili*. To address these aims I will test the following hypotheses:
H₀:  *C. x flexili* is not present in natural stands where *C. ribicola* and *C. comandrae* pycnia are sporulating in close proximity

H₁:  *C. x flexili* is present in natural stands where *C. ribicola* and *C. comandrae* pycnia are sporulating in close proximity

H₀:  Populations of *Cronartium ribicola* and *C. comandrae* do not exhibit introgression into the natural populations of *C. ribicola* and *C. comandrae*

H₁:  Populations of *Cronartium* ribicola and *C. comandrae* are introgressed in areas where the two species overlap
2. Methodology

2.1 Sample Collection

Aeciospores were collected from white pines and lodgepole pines from sites across British Columbia and Alberta, which included both provincial crown land and federal parks Canada land (Figure 2). Sampling permits are not required for provincial crown land and were obtained for sampling conducted on federal land. Sampling efforts were concentrated where the presence of the hybrid rust is already known to occur, in Mckendrick Pass British Columbia, and three sites in Alberta: Waterton Lakes National Park, Porcupine Hills Public Land Use Zone, and Kananaskis Country (Joly et al., 2006). Additionally, aeciospores of putatively pure parental species were collected from white pine spp. and lodgepole pines outside of the known hybrid range; this included 5 sites in B.C. (Bowser, Chilliwack, Chief Lake Rust Trial Site, Crystal Lake Rust Trial Site and Yoho National Park) as well as Jasper National Park in Alberta (see Appendix B). The Bowser site serves as a positive control population for C. ribicola, as there is no known occurrence of C. comandrae on Vancouver Island (Zeglen S., Personal Communication, April 2017). At each site 10 trees and 10 aecia per tree were sampled. If this number could not be achieved the maximum number of trees up to 10 were sampled.

Sampling was conducted according to the following protocol: cankers were located on stems or branches with developed aecial blisters which had peridial membranes that had not yet ruptured. A sterilized toothpick was used to carefully rupture a single blister at a time and spores were collected in 1.5ml Eppendorf tubes. Blisters were sampled starting from the bottom of the canker and moving upwards to avoid cross contamination. At each sampling site up to 10 trees were sampled, with 5-10 samples being taken per tree depending on availability. Eppendorf tubes were labelled with tree number, sample number, site location and date. Gloves were
sterilized with 70% Ethanol solution between sampling of aecia to avoid cross contamination. Whenever possible GPS coordinates were taken for the sampled trees. The samples were then placed in plastic storage boxes, and placed in a cooler with an ice pack to remain at approximately 4°C until the samples could be placed into short term storage and transferred into the refrigerator. Upon arrival at UBC spores were placed first into short-term storage at 4°C, then long-term storage at -80°C according to Zambino General Rust Protocols (unpublished 1998; see Appendix C).

Figure 2. Sampling site location map displaying sites which were successfully sampled throughout the 2017 and 2018 sampling seasons. Blue markers indicate that aeciospores were sampled from cankers on Pinus strobus spp., red markers indicate that aeciospores were sampled from cankers on infected P. contorta, while the red/blue marker indicates that both white pines and lodgepole pines were sampled for Cronartium spp. aeciospores (ESRI 2011).
2.2 Microscopy

The spore morphology of all samples collected in the hybrid zones were analyzed microscopically to identify intermediate spore types of the hybrids as well as parental species as described by Joly et al. (2006) and as demonstrated by the hybrid spores collected in 2012 in Mckendrick Pass (see Appendix F). The spores were analysed for characteristics including the shape, length (µm) and the presence of apical acumination and/or conspicuous smooth spot (Joly et al., 2006). Morphology was checked either when assessing germination rates, or with a subset being taking from samples and placed onto a small amount of 1% bacto-agar gel on a glass slide. A compound microscope was used to determine if morphology is varying from the pure parental species.

2.3 DNA Extraction

Aeciospores were treated for DNA extraction using two protocols. The first protocol is the Qiagen DNeasy Plant Mini extraction kits® manufacturers protocol (Qiagen Inc., Toronto, Ontario Canada) with the exception of the lysis step. The modified lysis step is as follows: to complete lysis the fungal biomass was contained in 2.0mL Safe-Lock tubes, sterile glass beads were added to the tubes with approximately 10mg of diatomaceous earth. The safe-lock tubes containing the fungal biomass, glass beads and diatomaceous earth were then incubated at -80°C overnight. After incubation the samples were mechanically disrupted using a Mixer Mill instrument (3 minutes @ 1/30 sec). The rest of the DNA extraction was done following the manufacturers protocol for the Qiagen DNeasy Plant Mini extraction kits® (Qiagen Inc., Toronto, Ontario Canada). The second protocol used for fungal DNA extraction was modified from Hu’s (2016) protocol (See Appendix D for modifications).
2.4 Polymerase Chain Reaction (PCR) and Electrophoresis

PCR amplification was performed in 26μl volumes which contain 2.5μl of each 10× PCR Buffer, (10 μm) Fp (LEPFI), (10 μm) Rp (LEPRI), 0.75μl MgCl₂, 0.5μl dNTP, 0.1μl of Taq polymerase (Invitrogen). The program was set for denaturing at 95°C for 5 minutes and 35 cycles of 94°C for 30 seconds, a gradient from 55°C to 60°C for 30 seconds followed by 72°C for 1 minute. The reaction ended with an extension period of 72°C for 10 minutes. The primer used to amplify DNA fragments in PCR is DCON10 from Joly et al. 2006 (Forward: TATGCCAGTTC GGTAAGG; Reverse: CAAGCCGCTGAAAGGTGA). The PCR products are run on 1.5% agarose gels to verify amplification. Samples which were successfully amplified were sent for sequencing at the Université of Laval in Québec using an Applied Biosystems® 3730xl DNA analyzer. Sequence alignment, editing and chromatograph viewing was done using Geneious® 6.0.6.

2.5 Real-time PCR and Locked Nucleic Acid Probe Analysis and Design

To facilitate large-scale screening, the 2018 samples were analyzed using real-time PCR (qPCR) with Locked Nucleic Acid (LNA) probes on Applied Biosystems ViiA™ 7 Real-Time PCR System. The probes were designed using the sequences: *C. ribicola* Dcon10 1 TGGCC+T+C+G+ACGTGT (FAM); *C. comandrae* Dcon10 1 TGGC+C+T+T+C+ACGTGT (HEX) (Capron, unpublished). The + sign indicates a LNA base, the LNA modification "rigidifies" the oligonucleotide, increasing the melting temperature. The probes were created using the discriminant SNPs amplified by the DCON10 primer (Joly et al., 2006) (*Cronartium* Dcon10 1: Forward AAAGCACTCACCCCTTGGTC; *Cronartium* Dcon10 1: Reverse: GACTGGACACATTTG CTCCCT) (see Appendix E). The samples used to create the LNA probe were collected in 2017 and included: 15 *C. comandrae* sequences, which consisted of 8 *C.
comandrae samples collected near Prince George B.C., 6 C. comandrae samples collected near Smithers B.C., as well as 1 bulk C. comandrae from Ting Pu (2012). Additionally, a total of 34 C. ribicola samples were used in the LNA probe design, 20 from Waterton Lakes A.B. and 14 from Smithers B.C. The 2 hybrid specimens used were from samples collected at Mckendrick Pass B.C. in 2012. The results were analyzed using Applied Biosystems QuantStudio™ Real – Time PCR Software V1.3 (Thermofisher).

2.6 Genotyping by Sequencing (GBS) and Library Construction

Genotyping by sequencing (GBS) was completed using Ion Torrent PstI-MstI technology. Ion torrent sequencing detects the release of hydrogen ions as dNTP’s as they are incorporated into a continuously expanding DNA strand via the use of native dNTPs and electronic sensors (Mascher et al., 2013). This approach uses library production to genotype a small fraction of an entire genome, as described in detail by Mascher et al. (2013). GBS is highly cost effective in comparison to previous genotyping methods, and, as there is no size selection step of digested DNAs, it can be performed using relatively small amounts of DNA (<20ng) and is effective at SNP discovery (Sonah et al., 2013).

DNA was extracted following the protocols outlined in section 2.3. The amount of DNA submitted for GBS varied from 2 ng – 20 ng. This variation in concentration was due to the fact that each sample consisted of aeciospores taken from single aecia. Those samples with concentrations of more than 20ng of DNA were normalized to 20ng / μL and used for library preparation to create a single 116-plex library sequenced on a single chip of Ion Proton. 116 samples were included, as was determined by the number of samples successfully collected and the amount of DNA that could be extracted. Single-end sequencing was performed using an Ion
Torrent platform (Institut de Biologie Integrative et des Systèmes (IBIS) at Université Laval). Each sample was tagged with a unique barcode for identification purposes.

2.7 Data Analysis

2.7.1 Fast-GBS pipeline

Processing of Ion Proton reads was done using the Fast-GBS pipeline as described by Torkamaneh et al. (2017). Two Fast-GBS analyses were run, one used the *C. ribicola* reference genome while the other used the *C. comandrae* reference genome. The reference genomes were retrieved from the Tree Agressors Identification using Genomic Approaches (TAIGA) project (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA190829) and NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCA_000464975.1) respectively. Before conducting the analyses, each reference genome was assessed for N50, total length and total number of scaffolds using a custom Python script (Python Software Foundation. Python Language Reference, version 2.7.0; https://github.com/KiahRae/KAllen-Thesis-Scripts/blob/master/Genome_N50_Scaffolds.py).

The Fast-GBS pipeline includes demultiplexing of fastq files performed by Sabre (Shannon et al., 2016), followed by read trimming and cleaning through Cutadapt (Martin, 2011). BWA is used for read mapping onto a reference genome using MEM (maximal exact matches) algorithm, which uses seed alignments followed by seed extensions with the Smith Waterman algorithm (Li & Durbin, 2010). Platypus (Rimmer et al., 2014) then re-analyzes the mapped reads to account for reads mapped to multiple locations and those that were poorly mapped (Torkamaneh et al., 2017). Variants are called using Platypus (SNPs and Indels), as well as haplotype construction and individual-level filtering (Torkamaneh et al., 2017). The output file generated is a VCF. These VCF files were used to extract SNPs shared between the hybrid
and each parental species. The filtering parameters applied to the SNPs included only those sites with a filter flag PASS (the reads of low quality and with low coverage are flagged ‘badreads’ as determined by the Fast-GBS pipeline), only including bi-allelic sites, excluding sites that consisted of an indel, and removing SNPS with more than 80% missing data. The reads from individual samples including both parental species and the hybrid were mapped onto each of the *C. ribicola* and *C. comandrae* genomes using Fast-GBS.

### 2.7.2 Hybridization index

The two hybrid samples which were successfully GBS sequenced were analyzed for level and direction of introgression using the Fast-GBS filtered variant output. First, the overall degree of genetic differentiation was calculated using the $\chi^2$ goodness of fit F-statistic ($F_{st}$) for each filtered SNP loci to determine those that are highly differentiated (see Appendix G for $F_{st}$ calculations). Those with the highest $F_{st}$ values were screened and further analyzed to determine proportion of fixed loci from each parental species using a custom Python script (https://github.com/KiahRae/KAllen-Thesis-Scripts). That is, those loci which are found in both *C. x flexili* samples but fixed in only one parental species were identified.

To calculate an index value for each of the populations analyzed, the *C. x flexili* loci which are homozygous fixed in *C. ribicola* were assigned a value of 0, while those homozygous fixed in *C. comandrae* were given a value of 1, and heterozygous loci (one allele fixed in *C. ribicola* while the other is fixed in *C. comandrae*) were valued at 0.5 (Table 2). Subsequently these values were totalled and averaged over the total number of highly differentiated loci between the putative pure parental populations, to give a hybridization index number between 0 and 1.
Table 2. Hybridization index values with homozygous reference alleles fixed in C. ribicola given a value of 0, homozygous alternate alleles fixed in comandrae given a value of 1, and heterozygous fixed alleles given a value of 0.5.

<table>
<thead>
<tr>
<th>Spp. with fixed loci</th>
<th>Genotype</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronartium ribicola</td>
<td>0/0 or 1/1</td>
<td>0</td>
</tr>
<tr>
<td>Cronartium comandrae</td>
<td>0/0 or 1/1</td>
<td>1</td>
</tr>
<tr>
<td>C. ribicola &amp; C. comandrae</td>
<td>0/1 or 1/0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

2.7.3 Analysis of variance (ANOVA) and Kruskal-Wallis H-test

Both a one-way analysis of variance (ANOVA) and the Kruskal-Wallis H-test (non-parametric version of ANOVA) were used to test differences in fixed loci between the 3 populations: Mckendrick Pass, Bowser and the Rocky Mountain sites (see Appendix H for equations). The sample measurements were the hybridization values of fixed homozygous and heterozygous loci in each population. The Kruskal-Wallis H-test was used as well as the ANOVA to account for the non-normal data distribution. Both tests were run using a custom Python script (https://github.com/KiahRae/KAllen-Thesis-Scripts).

2.7.4 Probability analysis

A probability analysis was completed to determine the probability that the alleles observed at one locus are from an interspecific cross event (i.e. one allele copy from C. ribicola and one from C. comandrae). Only those loci with $F_{st} > 0.80$ were included in the analysis (see Appendix G for $F_{st}$ calculations). The genotypes of all loci in each sample from the two hybrid zone populations, Mckendrick Pass and Rocky Mountain sites, the positive control sites Bowser and Prince George, as well as the two C. x flexili samples were analyzed. The analysis consisted of calculating the probability $P(x)$ that each genotype arose from an interspecific crossing event by comparing the individual genotypes to the positive control allele counts (Table 3).

Using a custom Python script the means of all loci from each Mckendrick Pass and Rocky Mountain sites, the positive control sites Bowser and Prince George, as well as the two hybrid
samples, were compared using ANOVA and a Kruskal-Wallis H-test followed by Mann–Whitney U test (non-parametric equivalent of the two sample t-test) among all populations to determine which populations differed (see Appendix H for calculations). The probability distributions were plotted (https://github.com/KiahRae/KAllen-Thesis-Scripts).

Table 3. Probability equations and hypotheses used to determine the probability that the alleles observed at one locus are from an interspecific hybridization event (i.e. one copy from C. ribicola and one from C. comandrae) or intraspecific cross (i.e. both alleles from one parental species).

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₀: the observed genotype is from an interspecific cross</td>
<td>[ P_{H₀}(x) = \frac{R_{obs₀}}{T₀} \times \frac{C_{obs₁}}{T₁} + \frac{R_{obs₁}}{T₁} \times \frac{C_{obs₀}}{T₀} ]</td>
</tr>
<tr>
<td>H₁: the observed genotype is from an intraspecific cross</td>
<td>[ P_{H₁}(x) = \frac{R_{obs₀}}{T₀} \times \frac{R_{obs₁}}{T₁} + \frac{C_{obs₁}}{T₁} \times \frac{C_{obs₀}}{T₀} ]</td>
</tr>
</tbody>
</table>

Where

\[ P(x) = \text{probability that a given genotype is from an interspecific (H₀) or intraspecific (H₁) cross} \]
\[ R_{obs₀} = \text{observed frequency of allele 0 at that locus in control C. ribicola reference population} \]
\[ R_{obs₁} = \text{observed frequency of allele 1 at that locus in control C. ribicola reference population} \]
\[ C_{obs₀} = \text{observed frequency of allele 0 at that locus in control C. comandrae reference population} \]
\[ C_{obs₁} = \text{observed frequency of allele 1 at that locus in control C. comandrae reference population} \]
\[ T₀ = \text{total frequency of allele 0 in both Cronartium spp. control populations} \]
\[ T₁ = \text{total frequency of allele 1 in both Cronartium spp. control populations} \]

2.7.5 sppIDer pipeline

The Ion Proton reads were also analyzed using the sppIDer pipeline (Langdon et al., 2018). The sppIDer workflow is designed to identify interspecific hybrids through mapping the short-read sequence data to a reference genome which is referred to as a ‘combination reference genome’ as it is built from the genomes of numerous species of interest (Langdon et al., 2018). In this study the combination reference genome was made of both proposed parental species C. ribicola and C. comandrae. The pipeline assesses the genomic contribution of each parental genome, as well as
ploidies and introgressions, to the short read GBS sequence data (Langdon et al., 2018). Three sppIDer analyses were run, each mapping a different species to the reference genome. The short-read sequence data in each run consisted of pooled short read fastq data from only one species of interest, this was comprised from the VCF output of the Fast-GBS pipeline.

The output of the sppIDer pipeline mapping is analysed for mapping quality and proportion of quality of reads that map to each of the reference genomes, as well as for depth of coverage. Those genomes with disproportionate mapping are identified using a chi-squared test (see Appendix H for equation; Langdon et al., 2018). After the combination reference genome is created (as described in Langdon et al., 2018), BWA is used for read mapping using MEM algorithm (Li & Durbin, 2010). Mapping quality is evaluated for each read that maps to each reference genome, SAMtools view and sort then uses the BWA output to filter out reads with mapping quality of less than 3 (Li et al., 2009). The pipeline then calculates the coverage by base pair or base group using bedtools genomeCoverageBed (Quinlan & Hall, 2010). The last process in the pipeline is to calculate the average depth of coverage by species and in windows, that is, the combination reference genome is sectioned in 1000 equal sized windows and the mean coverage is calculated using an R script (Landon et al., 2018).

2.7.6 RepeatMasker

As a relatively high number of C. comandrae reads mapped to the C. ribicola genome rather than the C. comandrae genome, the C. comandrae reads were examined to determine if a proportion were mapping on low complexity regions of the C. ribicola genome. To test this possibility the C. ribicola reference genome was filtered for scaffolds greater than 1000 bases in length, and the repeats were masked using RepeatMasker software on GenSAS using the ‘fungi’ library (Humann et al., 2019; GenSAS v6.0: Available at https://www.gensas.org/gensas). The
sppIDer pipeline was then re-run using only the *C. comandrae* reads that mapped to either parental genome during the previous *C. comandrae* sppIDer analysis, and a combination reference genome comprised of the masked *C. ribicola* genome and the original *C. comandrae* genome.

### 2.7.7 Contamination testing

Due to the high proportion of unmapped reads that resulted from mapping the *C. x. flexili* samples onto the combination reference genome in the sppIDer pipeline, a subsample of 100 reads from the unmapped data were filtered out and queried against the NCBI nucleotide database with BLASTn to determine potential contaminations (Altshul et al., 1990). The unmapped reads were then assembled and the genome assembly was blasted onto the NCBI database to compare the matches to the short-read blast.

The guanine-cytosine (GC) of both the mapped and unmapped data from the hybrid samples was also evaluated. To evaluate this the unmapped and mapped reads were filtered out of the sppIDer output bam file, 10 000 reads were subsampled from each file, and GC content was calculated for each read and plotted using a custom script in Python script (https://github.com/KiahRae/KAllen-Thesis-Scripts).
3. Results

In total, 831 samples from dikaryotic aecia of *C. ribicola* and *C. comandrae* from 13 sites across British Columbia and Alberta were collected for analysis over two sampling seasons. Of the samples collected in 2017, 109 samples were treated for DNA extraction and subsequently analyzed using end-point PCR. From these 109 samples representing 5 sampling locations, the 47 samples which had good band intensity relative to the control were sequenced using Sanger sequencing. All aeciospore samples were assessed for morphological characteristics.

Of the samples collected in both 2017 and 2018 across 13 sites, 167 samples were treated for DNA extraction following Hu’s modified protocol (Appendix D). A total of 71 samples from 7 sites were analyzed using qPCR and the LNA probe. Subsequently, 114 of these samples, and 3 samples collected in Mckendrick Pass in 2012, were GBS sequenced using Ion Torrent GBS. The genotypic data was used to identify potential hybrids and to assess the potential introgression in hybrid zones. A summary of results of hybrid identification are presented in Table 4.

<table>
<thead>
<tr>
<th>Location</th>
<th>Sample Year</th>
<th>Hybrid Morphology Observed</th>
<th>Hybrid Identified with PCR/sequencing using the DCON10 Primer</th>
<th>Hybrid Identified with qPCR using DCON10 Primer/Probe</th>
<th>Hybrid Identified using GBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waterton</td>
<td>2003</td>
<td>Yes (Joly et al., 2006)</td>
<td>Yes (Joly et al., 2006)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Porcupine Hills</td>
<td>2003</td>
<td>Yes (Joly et al., 2006)</td>
<td>Yes (Joly et al., 2006)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kananaskis Country</td>
<td>2003</td>
<td>Yes (Joly et al., 2006)</td>
<td>Yes (Joly et al., 2006)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mckendrick Pass</td>
<td>2012</td>
<td>Yes (Appendix F)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mckendrick Pass</td>
<td>2017</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Location</td>
<td>Sample Year</td>
<td>Hybrid Morphology Observed</td>
<td>Hybrid Identified with PCR/sequencing using the DCON10 Primer</td>
<td>Hybrid Identified with qPCR using DCON10 Primer/Probe</td>
<td>Hybrid Identified using GBS</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>----------------------------</td>
<td>--------------------------------------------------------------</td>
<td>------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Waterton</td>
<td>2017</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>McKendrick Pass</td>
<td>2018</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Waterton</td>
<td>2018</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Porcupine Hills</td>
<td>2018</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kananaskis Country</td>
<td>2018</td>
<td>No</td>
<td>-</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### 3.1 Microscopy

The results of microscopy analysis of all 831 samples collected in 2017 and 2018 did not indicate any abnormal aeciospore morphology present. That is, the *C. ribicola* aeciospores were of normal length and the presence of the conspicuous smooth spot was observed. The *C. comandrae* aeciospores were of normal length and the apical acumination was present.

### 3.2 PCR, qPCR and LNA probe analysis

Alignment and chromatogram viewing of the 2017 *C. ribicola* and *C. comandra* sequences did not reveal the presence of hybrid SNPs in the amplified region. When the *C. x. flexili* PCR products were amplified using the DCON10 primer developed by Joly et al. (2006) there were 16 heterozygous single-nucleotide polymorphism (SNP) in the genomic DNA. In Figure 3, these SNPs are demonstrated, e.g. where the first *C. comandrae* SNP consists of ‘G’, while the *C. ribicola* sequence consists of ‘A’, and in the case of the *C. x. flexili* SNP, allele carries ‘G’, while the other carries ‘A’. For all of the hybrid SNPs the peaks from both parental species alleles are present but at only half the height of homozygous alleles of the putative pure species.
Figure 3. Sequence alignment chromatogram of samples collected in 2017. Samples compared are C. comandrae (CBR) from Chief Lake Rust Trial, C. ribicola from Slesse (WPBR) and of the Hybrid (C. x flexili) DNA extracted from aeciospores collected in Mckendrick Pass in 2012. The aligned segments were amplified by the DCON10 primer, with C. x flexili clearly demonstrating heterozygous SNPs at different PCR amplicons. These heterozygous SNPs indicate that C. x flexili is derived from diploid genomic DNA with the polymorphic positions showing both peaks from both nucleotides simultaneously on the chromatogram (Geneious® 6.0.6).

A total of 71 samples collected in 2018 were analyzed using qPCR and the LNA probe. Results from analysis of the standard curve produced by Applied Biosystems QuantStudio™ Software V1.3 (Figure 4) did not yield any strong evidence of the presence of C. x flexili. All samples tested amplified as either putatively pure C. ribicola or C. comandrae.

Figure 4 demonstrates the results from qPCR testing on each of C. ribicola, C. comandrae and C x flexili respectively. The blue qPCR curve represents C. ribicola, and reaches the amplification threshold at a Ct of 32, while the red qPCR curve representing C. comandrae does not reach the threshold, indicating that this is a pure C. ribicola sample (Figure 4 Plot A). Figure 4 plot B shows the reverse effect, with the red signal representing C. comandrae and crossing the threshold at a Ct of 36 while C. ribicola does not cross. Figure 4 plot C shows the signal from C. x flexili collected in Mckendrick Pass in 2012. The plot C qPCR curves clearly indicate the
presence of both C. ribicola and C. comandrae amplified fragments as both qPCR curves cross the threshold with Ct values of 24 and 26 respectively.

Figure 4. Amplification plots showing the results of the LNA probe using C. ribicola, C. comandrae, and C. flexili material plotted with the variation of log (ΔRn - Rn minus the baseline) against the PCR cycle number. Rn is the fluorescence of the reporter dye divided by the fluorescence of a passive reference dye. (A) Amplification plot of a sample collected at Waterton Lakes AB demonstrating that it is pure C. ribicola. (B) Amplification plot of a sample collected at Mckendrick Pass BC, demonstrating that it is pure C. comandrae. (C) Amplification plot of C. x flexili DNA from Mckendrick Pass BC. (QuantStudio™).
3.3 Genotyping by Sequencing

3.3.1 Fast-GBS

The results of the quality checks of the reference genomes and variant calling using the Fast-GBS pipeline are summarized in Table 5. Of the 138574 possible C. ribicola SNPs called by the Fast-GBS pipeline, 1793 remained after filtering parameters were applied; for C. comandrae of the 140139 possible SNPs called, 1911 remained after filtering. These filtered SNPs were subsequently analyzed using a hybridization index and did not indicate any presence of newly sampled C. x flexili in the 2017 or 2018 sequenced samples. However, the results of genotyping the C. x flexili DNA extracted from the 2012 Mckendrick Pass samples indicated these samples are F1 hybrids, and the data indicated the absence of introgression, as demonstrated by the fixed SNPs from each parental species (Table 6).

Table 5. Results from the reference genome quality analysis performed using custom script in Python (N50, total length, total number of scaffolds) as well as the results of variant calling using the Fast-GBS pipeline (Overall SNPs called and Filtered SNPs).

<table>
<thead>
<tr>
<th>Ref. Genome Species</th>
<th>N50</th>
<th>Total Length</th>
<th>Total # Scaffolds</th>
<th>Overall possible SNPs</th>
<th>Filtered SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. ribicola</td>
<td>5360</td>
<td>71163587</td>
<td>19902</td>
<td>138574</td>
<td>1793</td>
</tr>
<tr>
<td>C. comandrae</td>
<td>2617</td>
<td>68605838</td>
<td>35717</td>
<td>140139</td>
<td>1911</td>
</tr>
</tbody>
</table>

3.3.2 Hybridization Index and ANOVA

Of the highly differentiated loci which were selected from the filtered SNPs, samples from the positive control site were compared to those samples collected in the hybrid zones. Of the Bowser samples analysed, 12 homozygous loci were fixed in C. comandrae and 67 in C. ribicola, compared to 13 and 68 respectively for the Mckendrick Pass samples and 13 and 68 respectively for the Rocky Mountain Sites (Waterton Lakes, Kananaskis Country and Porcupine Hills) (Table 6). We failed to reject the null hypothesis that there is no difference between the means of the
populations (ANOVA F-value = 0.002; p > 0.05; Kruskal-Wallis H-statistic, corrected for ties = 0.005, p > 0.05.

Table 6. Summary of homozygous and heterozygous (one allele from each parent) loci fixed in both C. x flexili samples. C. x flexili genotypes were compared to the genotypes of the positive control population (Bowser), and the two populations where the hybrid was previously sampled, Mckendrick Pass and the Rocky Mountain Sites (Waterton Lakes National Park, Kananaskis Country, and Porcupine Hills). Hybridization index calculations are included as described in section 2.7.2.

<table>
<thead>
<tr>
<th>Sample Location</th>
<th>Total Fixed C. comandrae loci</th>
<th>Total Fixed C. ribicola loci</th>
<th>Total Fixed Heterozygous loci</th>
<th>Overall Hybridization Index (0-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mckendrick Pass</td>
<td>13</td>
<td>29</td>
<td>245</td>
<td>0.4721</td>
</tr>
<tr>
<td>Bowser</td>
<td>12</td>
<td>28</td>
<td>246</td>
<td>0.4720</td>
</tr>
<tr>
<td>Rocky Mountain Sites</td>
<td>13</td>
<td>29</td>
<td>241</td>
<td>0.4717</td>
</tr>
</tbody>
</table>

3.3.3 Probability Analysis

The distribution of the probability \( P(x) \) that the observed genotypes from each locus arose from an interspecific cross are plotted in Figure 5. The means of all loci from each population were compared using ANOVA and Kruskal-Wallis H-test (see Appendix H for calculations). We reject the null hypothesis that all means are equal and find that there is a significant difference between the means of the populations (F-value = 1416.36, \( p < 0.05 \); and we reject the null hypothesis that all means are equal, there is a significant difference between the means of the populations. The computed Kruskal-Wallis H-statistic = 3949.70, \( p < 0.05 \)). The Mann – Whitney U test (non-parametric pairwise t-test) among all populations was used to compare the populations and indicated a significant difference only between Cronartium x flexili and each of the other populations, and between each of the C. ribicola and C. comandrae populations (results not shown).
Figure 5. Probability distributions depicting the probability $P(x)$ that the observed genotype arose from an interspecific hybridization event between *C. ribicola* and *C. comandrae*. The probabilities were calculated including only those loci with Fst values $> 0.80$ between positive control parental populations (Bowser and Prince George sites).

### 3.3.4 sppIDer

#### 3.3.4.1 *C. comandrae* samples mapped onto combination reference genome

Three *C. comandrae* samples were mapped onto the combination reference genome, with 64% of reads mapping onto *C. comandrae*, 21% of reads mapping onto *C. ribicola* and 15% of reads mapping to neither genome (Table 7). A chi-squared test was performed testing the null hypothesis that the reads will be randomly distributed across all genomes (Appendix H). The results of the chi-squared test which determines which parental species from the combination reference genome significantly contribute to each parental species only identified *C. comandrae*
as contributing to the *C. comandrae* samples for those reads which had a high mapping quality (MQ=60) (Table 8).

Table 7. *sppIDer* output depicting chi-squared results, genomes contributing to the tested *C. comandrae* samples can be identified as those producing a statistically significant positive residual(*) in chi-squared test 1 of the counts of reads that map. Chi-squared test 2 is performed only on the MQ = 60 reads and determines which genomes are enriched for high-quality reads.

<table>
<thead>
<tr>
<th>$\chi^2$ Test</th>
<th><em>C. ribicola</em></th>
<th><em>C. comandrae</em></th>
<th>Unmapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of reads mapped</td>
<td>-1304.7362</td>
<td>-466.14969</td>
<td>1770.8859</td>
</tr>
<tr>
<td>MQ60</td>
<td>-251.7183</td>
<td><em>251.7183</em></td>
<td>--</td>
</tr>
</tbody>
</table>

Table 8. Summary of *sppIDer* output, reporting the percent of *C. comandrae* reads that map to each reference genome, respective total number of mapped reads and the mapping quality (MQ). *sppIDer* uses established MQ score (Li et al. 2008) to bin reads by their map-ability on a 0-60 score.

<table>
<thead>
<tr>
<th>Species</th>
<th>Total mapped reads</th>
<th>% of all reads</th>
<th>% Nonzero mapped reads</th>
<th>All average MQ</th>
<th>NonZero average MQ</th>
<th>All Median MQ</th>
<th>Nonzero Median MQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmapped</td>
<td>54028</td>
<td>15.14%</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>C. ribicola</em></td>
<td>73447</td>
<td>20.59%</td>
<td>26.62%</td>
<td>47.85</td>
<td>52.58</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td><em>C. comandrae</em></td>
<td>229308</td>
<td>64.27%</td>
<td>73.38%</td>
<td>32.88</td>
<td>40.92</td>
<td>35</td>
<td>54</td>
</tr>
</tbody>
</table>

The *sppIDer* pipeline outputs a number of figures including percentage of reads mapped to each reference genome and depth of coverage plots (Figures 6-14). The depth of coverage (Figures 8, 11&14) includes a count of how many reads cover each base pair or region of the genome and can be used to infer the species, the parents of the hybrids, and ploidy changes either between or within a genome.
Figure 6. Mapping bar plot with percentage of C. comandrae reads mapped by species onto the combination reference genome including the unmapped data.

Figure 7. Mapping bar plot with percentage of C. comandrae reads mapped by species onto the combination reference genome excluding the unmapped data.
3.3.4.2 *C. ribicola* samples mapped onto combination reference genome

The genomes of 12 *C. ribicola* samples were mapped onto the combination reference genome, with 96% of reads mapping onto *C. ribicola*, <1% of reads mapping onto *C. comandrae* and 3% of reads mapping to neither genome (Table 9). A chi-squared test was performed testing the null hypothesis that the reads will be randomly distributed across all genomes. The results of the chi-squared test which determines which parental species from the combination reference genome significantly contribute to each parental species only identified *C. ribicola* as statistically contributing to the *C. ribicola* samples for both count of reads mapped and for those reads which had a high mapping quality (MQ=60) (Table 10).

![Figure 8. Log2 whole genome mean average depth of coverage of *C. comandrae* sample reads mapped onto the combination reference genome.](image)
Table 9. Summary of sppIDer output, reporting the percent of C. ribicola reads that map to each reference genome, respective total number of mapped reads and the mapping quality (MQ). sppIDer uses established MQ score (Li et al. 2008) to bin reads by their map-ability on a 0-60 score.

<table>
<thead>
<tr>
<th>Species</th>
<th>Total mapped reads</th>
<th>% of all reads</th>
<th>% Nonzero mapped reads</th>
<th>All average MQ</th>
<th>NonZero average MQ</th>
<th>All Median MQ</th>
<th>Nonzero Median MQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmapped</td>
<td>91175</td>
<td>3.29%</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>C. ribicola</em></td>
<td>2667423</td>
<td>96.19%</td>
<td>99.57%</td>
<td>22.96</td>
<td>36.14</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td><em>C. comandrae</em></td>
<td>14616</td>
<td>0.53%</td>
<td>0.43%</td>
<td>9.53</td>
<td>18.85</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 10. sppIDer output depicting chi-squared results, genomes contributing to the tested C. ribicola samples can be identified as those producing a statistically significant positive residual (*) in chi-squared test 1 of the counts of reads that map chi-squared test 2 is performed only on the MQ = 60 reads and determines which genomes are enriched for high-quality reads.

<table>
<thead>
<tr>
<th>Test</th>
<th>C. ribicola</th>
<th>C. comandrae</th>
<th>Unmapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of reads mapped</td>
<td>1418.7814*</td>
<td>-763.5767</td>
<td>-655.2048</td>
</tr>
<tr>
<td>MQ60</td>
<td>593.3249*</td>
<td>-593.3249</td>
<td>--</td>
</tr>
</tbody>
</table>

Figure 9. Mapping bar plot with percentage of C. ribicola reads mapped by species onto the combination reference genome including the unmapped data.
Figure 10. Mapping bar plot with percentage of C. ribicola reads mapped by species onto the combination reference genome excluding the unmapped data.

Figure 11. Log2 whole genome mean average depth of coverage of C. ribicola sample reads mapped onto the combination reference genome.
3.3.4.3 C. x flexili samples mapped onto combination reference genome

Two C. x flexili samples were mapped onto the combination reference genome, with 39.03% of reads mapping onto C. ribicola, 21.09% of reads mapping onto C. comandrae and 39.87% of reads mapping to neither genome (Table 11). A chi-squared test was performed testing the null hypothesis that the reads will be randomly distributed across all genomes. The results of the chi-squared test which determines which parental species from the combination reference genome significantly contribute to each parental species only identified C. ribicola as statistically contributing to the C. x flexili samples for only those reads which had a high mapping quality (MQ=60) (Table 12).

Table 11. Summary of sppIDer output, reporting the percent of C. x flexili reads that map to each reference genome, respective total number of mapped reads and the mapping quality (MQ). sppIDer uses established MQ score (Li et al. 2008) to bin reads by their map-ability on a 0-60 score.

<table>
<thead>
<tr>
<th>Species</th>
<th>Total mapped reads</th>
<th>% of all reads</th>
<th>% Nonzero mapped reads</th>
<th>All average MQ</th>
<th>NonZero average MQ</th>
<th>All Median MQ</th>
<th>Nonzero Median MQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmapped</td>
<td>195412</td>
<td>39.87%</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. ribicola</td>
<td>191291</td>
<td>39.03%</td>
<td>61.3%</td>
<td>29.28</td>
<td>41.56</td>
<td>24</td>
<td>60</td>
</tr>
<tr>
<td>C. comandrae</td>
<td>103379</td>
<td>21.09%</td>
<td>38.7%</td>
<td>34.12</td>
<td>41.46</td>
<td>39</td>
<td>56</td>
</tr>
</tbody>
</table>

Table 12. sppIDer output depicting chi-squared results, genomes contributing to the tested C. x flexili samples can be identified as those producing a statistically significant positive residual (*) in chi-squared test 1 of the counts of reads that map. Chi-squared test 2 is performed only on the MQ = 60 reads and determines which genomes are enriched for high-quality reads. Test 2 is more appropriate for statistically assessing which parents contribute to the hybrids.

<table>
<thead>
<tr>
<th>$\chi^2$ Test</th>
<th>C. ribicola</th>
<th>C. comandrae</th>
<th>Unmapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of reads mapped</td>
<td>-9.8323</td>
<td>-143.3748</td>
<td>153.2071</td>
</tr>
<tr>
<td>MQ60</td>
<td>66.8459*</td>
<td>-66.8459</td>
<td>--</td>
</tr>
</tbody>
</table>
Figure 12. Mapping bar plot with percentage of C. x flexili reads mapped by species onto the combination reference genome including the unmapped data.

Figure 13. Mapping bar plot with percentage of C. x flexili reads mapped by species onto the combination reference genome excluding the unmapped data.
**3.3.5 RepeatMasker**

The results of the repeatmasker analysis indicate that, in comparison to the sppIDer analysis ran with the unmasked genome, nearly 13% of the *C. comandrae* reads that mapped to the *C. ribicola* reference genome were indeed mapping to low complexity regions of the genome (Tables 13 and 14). The proportion of *C. comandrae* reads mapping to the *C. comandrae* genome increased from 64.27% (Table 8) to 76.35% (Table 13) when using the masked *C. ribicola* genome. Table 14 depicts the chi-square results, where *C. comandrae* is determined to be the only contributing parental genome for both the overall count of reads mapped and the reads that mapped with an MQ of 60.

---

Figure 14. Log2 whole genome mean average depth of coverage of *C. x flexili* sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each base pair of the genome.
Table 13. Summary of sppIDer output, reporting the percent of C. comandrae reads that map to either the masked C. ribicola genome, the C. comandrae genome or are unmapped. The output also includes the respective total number of mapped reads and the mapping quality (MQ). sppIDer uses established MQ score (Li et al. 2008) to bin reads by their map-ability on a 0-60 score.

<table>
<thead>
<tr>
<th>Species</th>
<th>Total mapped reads</th>
<th>% of all reads</th>
<th>% Nonzero mapped reads</th>
<th>All average MQ</th>
<th>NonZero average MQ</th>
<th>All Median MQ</th>
<th>Nonzero Median MQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmapped</td>
<td>3055</td>
<td>0.98%</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. ribicola (masked)</td>
<td>70459</td>
<td>22.67%</td>
<td>25.31%</td>
<td>48.74</td>
<td>53.97</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>C. comandrae</td>
<td>237272</td>
<td>76.35%</td>
<td>74.69%</td>
<td>32.83</td>
<td>40.8</td>
<td>35</td>
<td>54</td>
</tr>
</tbody>
</table>

Table 14. sppIDer output depicting chi-squared results, genomes contributing to the tested C. comandrae samples can be identified as those producing a statistically significant positive residual (*) in chi-squared test 1 of the counts of reads that map. Chi-squared test 2 is performed only on the MQ = 60 reads and determines which genomes are enriched for high-quality reads. Test 2 is more appropriate for statistically assessing which parents contribute to the hybrids.

<table>
<thead>
<tr>
<th>( \chi^2 ) Test</th>
<th>C. ribicola (masked)</th>
<th>C. comandrae</th>
<th>Unmapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of reads mapped</td>
<td>-73.3085</td>
<td>356.5545*</td>
<td>-283.2459</td>
</tr>
<tr>
<td>MQ60</td>
<td>-72.6405</td>
<td>72.6405*</td>
<td>--</td>
</tr>
</tbody>
</table>

3.3.6 Contamination Testing

The output of mapping the hybrid GBS data using the sppIDer pipeline resulted in a relatively large percentage of unmapped data (39.87%). To determine the origin of the unmapped data, the unmapped reads were filtered out and 100 randomly selected reads were blasted (blastx) in the NCBI ‘nr’ database. The blast returned a number of alignments, with the top 6 hits belonging to the bacteria species *Methylocystis sp.*, *Phreatobacter sp.*, *Bradyrhizobium spp.* and *Rhodopseudomonas palustris* (Table 13). Thereafter the unmapped reads were assembled, and the whole genome was blasted (blastn) in the NCBI ‘nt’ database, whereby the top six hits were identical to those of the unmapped reads (Table 13).
Table 15. The top 6 hits returned from blasting a subset of unmapped reads onto the NCBI ‘nr’ database in the left two columns and the top 6 hits returned from blasting the assembled unmapped reads onto the NCBI ‘nt’ database.

<table>
<thead>
<tr>
<th>NCBI ID</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE956757.1</td>
<td>Methylocystis</td>
</tr>
<tr>
<td>CP027668.1</td>
<td>Phreatobacter</td>
</tr>
<tr>
<td>LT859959.1</td>
<td>Bradyrhizobium sp. ORS285</td>
</tr>
<tr>
<td>CP000494.1</td>
<td>Bradyrhizobium sp. BTAi1</td>
</tr>
<tr>
<td>CU234118.1</td>
<td>Bradyrhizobium sp. ORS278</td>
</tr>
<tr>
<td>CP000463.1</td>
<td>Rhodopseudomonas palustris</td>
</tr>
</tbody>
</table>

The Guanine-Cytosine (GC) content of both mapped and unmapped reads was then calculated and compared (Figure 15), with the GC mean of the mapped reads equalling 48% and the mean of unmapped reads equalling 62%. For comparison the GC content of the combination reference genome assembled by sppIDer was calculated and determined to be ≈ 39% while the GC content of the bacterial genome ≈ 65% (comprised of the genomes of the bacteria described in Table 15). A t-test was used to compare the means of the mapped and unmapped reads and indicated a significant difference between the means (p = < 1×10⁻⁶) (Table 14). These results indicate that the C. x flexili genome likely has a bacterial contaminated, thus explaining the
relatively high proportion of unmapped *C. x flexili* reads in the sppIDer analysis (see 3.3.4.3).

![Per Sequence Guanine-Cytosine (GC) Content](image)

Figure 15. Mean guanine-cytosine (GC) content distribution of both mapped and unmapped reads of *C. x flexili* mapped onto the combination reference genome.

<table>
<thead>
<tr>
<th></th>
<th>Mean GC Content (%)</th>
<th>T-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapped Reads</td>
<td>48.35</td>
<td>-131.62</td>
<td>&lt;1×10⁻⁶</td>
</tr>
<tr>
<td>Unmapped Reads</td>
<td>62.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination reference genome</td>
<td>38.92</td>
<td>10.63</td>
<td>2.41e-26</td>
</tr>
<tr>
<td>Bacteria combination genome</td>
<td>65.34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Discussion

*Cronartium x flexili* was first discovered in 2003 in the Southern Canadian Rocky Mountains and subsequently identified in a region geographically distant from the Rocky Mountains, in Mckendrick Pass, near Smithers B.C. The discovery of a separate hybridization event in Mckendrick Pass in 2012 was significant as it revealed that the presence of *C. x flexili* could be widespread. There have been no new reports nor any subsequent analysis of this hybrid pathogen since the two aforementioned discoveries. The purpose of this study was to further the understanding of this hybridization event, specifically to determine if *C. x flexili* continues to be prevalent in those forests where it was formerly sampled and whether or not it is introgressing, and thus allowing gene flow to occur between the two putative parental species, *Cronartium ribicola* and *Cronartium comandrae*.

4.1 Presence of *C. x flexili*

Both of the morphological observations as well as the genetic testing did not reveal the presence of *C. x flexili* in any of the samples collected during the 2017 or 2018 sampling seasons. *C. x flexili* was first identified in 2003 and subsequently identified in 2004 in the same locations, indicating that hybridization events were recurring at least in the first filial generation (F1), or that the hybrid was sustaining in nature perennially. The absence of the hybrid in the samples collected in this study could be owing to inferior relative fitness of *C. x flexili* compared to that of the parental species.

It is hypothesized that F1 *C. x flexili* hybrids are generated through plasmogamy of two haploid mycelia (n) belonging to each of *C. ribicola* and *C. comandrae*, which are not previously known to be interfertile species, but may have evolved compatible mating types over time or preserved the compatible mating type from a common ancestor (Joly et al., 2006). This results in
a dikaryotic offspring with two haploid nuclei, one from each parental species. In a majority of Basidiomycetes nuclei are inherited from both parents during sexual reproduction, while mitochondria are inherited from only one of the parental species (Xu & Wang, 2015). This phenomenon occurs in the genus *Cronartium* and was observed by Joly et al. (2006) as they described that mitochondria in *C. x flexili* were inherited only from *C. ribicola* and not from *C. comandrae*.

The presence of solely *C. ribicola* mitochondria in *C. x flexili* is consistent with the pattern of mitochondrial inheritance in the basidiomycete fungi, whereby mitochondrial genomes are inherited from the parent with the larger gamete (Wilson & Xu, 2012). In basidiomycetes the dispersing gametes (spermatia) are smaller whereas the spermagonia are the larger, more complex and sessile ‘female’ gametes (Xu & Wang, 2015). Thus, mitochondria are inherited predominantly from the female structure, which aligns with the theory as to the formation of *C. x flexili* through sexual hybridization between the spermatia of *C. comandrae* and the spermagonia of *C. ribicola*.

The inheritance of mitochondria from only one parental species can reduce hybrid fitness via what is known as ‘hybrid breakdown’ (Ellison & Burton, 2006). Hybrid breakdown refers to a reduction in fitness in hybrids due to incompatibility between mitochondrial and nuclear alleles (Burton & Barreto, 2012). This is explained by the Bateson-Dobzhansky-Muller incompatibility (BDMI) model, which is a biological model of evolution of genetic incompatibility between two differentiated taxa which interbreed. One cause of this incompatibility is via conflicting interactions between nuclear and mitochondrial genomes (Burton & Barreto, 2012). Burton and Barreto (2012) indicate that the origin of these incompatibilities is from both the dependence of mtDNA replication, transcription and translation upon nuclear gene products and of the cellular
metabolic energy production (aerobic generation of ATP). They suggest that when populations which have divergent mitochondrial lineages hybridize, the nuclear and mitochondrial genomes of these populations have not had a common co-evolutionary history and thus these processes are particularly susceptible to disruption.

In fact, there have been several studies on F1 hybrids between species of the same genus that were previously geographically isolated where hybrid breakdown has been observed (Olson & Stenlid, 2001; Giordano et al., 2018). In the fungal hybrid of *Heterobasidion irregulare* and *H. occidentale*, the hybrids which had only inherited *H. irregulare* mitochondria were more virulent than hybrids with *H. occidentale* mitochondria (Olson & Stenlid, 2001). Olson and Stenlid suggest that the reduction of fitness in the hybrids with *H. occidentale* mitochondria could be the result of the conflicting interactions that result from a mismatch between the mitochondrial and nuclear genomes. Although the F1 hybrids *H. irregulare X H. occidentale* are believed to be extremely rare in nature (Olson & Stenlid, 2001).

Giordano et al. (2018) tested the influence of mitochondrial inheritance on the relative fitness of the aforementioned hybrid offspring of *Heterobasidion irregulare* and *H. annosum* (s.s). In their study they examined the difference in saprotrophic growth, as a measure of fitness, between hybrids which had mitochondria inherited from *H. irregulare* and *H. annosum* comparatively. The results indicated significantly lower saprotrophic growth in the heterokaryotic hybrid offspring with mitochondria inherited from *H. annosum* (s.s.) than those with *H. irregulare* inherited mitochondria. Giordano et al. (2018) concluded that the fitness of these hybrids is reduced due to the mismatch between the mitochondrial genome of *H. annosum* s.s. and the nucleus of *H. irregulare*. They propose that this mismatch is evidence of ‘hybrid breakdown’ as explained by the BDMI model.
As *C. ribicola* mitochondrion had not previously interacted with the allopatrically evolved *C. comandrae* nuclear genome, this could represent the first time that the alleles of *C. comandrae* and the mitochondrial genome of *C. ribicola* occur in the same individual. Thus, as per the BDMI model, natural selection has not already tested the compatibility between the alleles of *C. comandrae* and the mitochondrial genome of *C. ribicola*, and if there is an incompatibility then there is a reduction in hybrid fitness. If this is the case, *C. x flexili* could form in the F1 generation but not persist as hybrid breakdown occurs. Further research on the interactions of the nuclear and mitochondrial components of these two species would need to be performed in order to determine the validity of this prediction.

4.2 Evidence of Introgression

Although it is clear that the two parental species are genetically distinct and have attained reciprocal monophyly at several loci, the results of the fixed loci analysis indicate that the *C. x flexili* samples analysed in this study are interspecific hybrids. This is signified by the fixed SNPs in the hybrid from each of the parental species and by the mappability of the *C. x flexili* genome onto both parental genomes relative to the controls. Yet, the comparison of the relative proportion of fixed loci and hybridization index values of the control site (Bowser B.C.) and the two hybrid zones (Mckendrick Pass and Southern Canadian Rocky Mountains) showed no significant difference between populations, which does not provide evidence that *C. x flexili* is back-crossing into either of the introduced *C. ribicola* nor the native *C. comandrae*. However, if the hybrid samples analyzed in this study were undoubtedly F1 we would expect all of the highly differentiated loci analysed to be heterozygous, demonstrating one fixed allele from *C. ribicola* and one fixed from *C. comandrae*, yet we do not find this to be the case (section 3.3.2 Table 6). Evidence supporting the hypothesis that *C. x flexili* exists only in the F1 is shown in the
amplification of the DCON10 gene, whereby all heterozygous sites in *C. x flexili* for each SNP contain one *C. ribicola* allele and one *C. comandrae* allele (section 3.2 Figure 3). However, this is for only one gene and this gene could be in the center of a recombination block in an F2 hybrid, which would result in every SNP remaining as a heterozygote. Considering these results, it is not entirely clear if *C. x flexili* is capable of facilitating gene flow between the parental species, in fact it remains entirely possible that the hybrid could persist past the F1 stage.

The aforementioned BDMI model proposes that genetic incompatibility between allopatrically evolved hybrid species can arise from nuclear allele incompatibilities, and render a hybrid inviable or sterile (Dobzhansky, 1934; Muller, 1942). Xu and He (2011) tested this theory on the yeast *Saccharomyces cerevisiae x S. paradoxus* F2 hybrids formed by auto-diploidization of F1 gametes, by comparing clonal growth rate and sporulation efficiency as proxies for yeast viability and yeast fertility respectively. They found that in all F2 hybrids analysed sporulation was negatively affected to a much greater extent than clonal growth rate was, indicating that the proposed genetic incompatibility had a greater effect on fertility than it did on viability. Hu and Xe (2011) suggest that as mitosis and meiosis require different genes and that the molecular features in the two processes differ, the gene-incompatibility complex as described by the BDMI model will affect these processes differently.

Zanders et al. (2014) found hybrid infertility of the hybrid yeast *Schizosaccharomyces pombe x S. kambucha* was majorly due to chromosomal rearrangements in all three chromosomes analysed in their study. *S. pombe x S. kambucha* hybrids are able to complete meiosis but the gametes produced have very low viability. The first abnormality Zanders et al. found was a chromosomal inversion in *S. pombe* chromosome 1 relative to *S. kambucha*, which maintained its native state. The inversion limits interhomolog crossover and causes an odd number of crossovers
in this region, generating incomplete genomes and contributing to hybrid infertility (2014). The second chromosome structural abnormality they found was a reciprocal translocation in the S. kambucha lineage, which resulted in certain chromosomal combinations becoming inviable as the translocated sequence being disrupted was found to include essential genes.

The mechanisms of hybrid infertility as described by Xu and He (2011) and Zanders et al. (2014) could be at play in the C. x flexili hybrid, rendering it infertile and inconsequential as a proxy for gene flow between its parental species. Although these genetic mechanisms have been proposed to be quickly evolving (Xu & He, 2011) and as such hybrid offspring could become fertile, either speciating themselves, or introgressing into parental populations.

4.3 Climate/site variation

As discussed in section 1.1.3, C. ribicola and C. comandrae have complex life cycles which are impacted by local climate. Sporulation conditions which are most favourable for C. ribicola include high precipitation and humidity levels (Dukes et al., 2009). The yearly variation in local temperature and the local influence of moisture availability in the form of relative humidity causes large variations in sporulation periods and infection success across time and geography (Hunt, 2005; Hunt et al., 2007). Geils et al. (2010) outlined the largest constraint on the spread of rusts is the various site environmental factors.

Local climate could be a driving factor in localized hybridization events, as it is hypothesized that both C. ribicola and C. comandrae pycnia must be sporulating and receptive at the same time and place in order for hybridization to occur. Aecia form only when sufficient temperatures for sporulation are reached, and success will depend on local moisture levels (Eppstein & Tainter, 1976; Krebill, 1968). This is true for both aforementioned Cronartium
species. Thus, the timing of the life cycle, and the formation of gametes, depends almost entirely on climatic factors.

The distribution of *Cronartium* rust infected white pines and lodgepole pines overlap in a limited distribution and only at certain mid-elevation sites, thus there is a small range both geographically and environmentally in which these two species come in contact with one another. Furthermore, the sporulation timing of the two species can vary dramatically. For example, at our sampling site located near Bowser B.C., on Vancouver Island, white pine blister rust aecia sporulated nearly 12 weeks before the *C. ribicola* aecia sporulated in Mckendrick Pass, near Smithers B.C. Similarly, *C. comandrae* tended to sporulate earlier than *C. ribicola* in similar geographic regions, due to the lower elevation of its host species. This suggests that the conditions required for successful sexual hybridization at the pycniospore stage would require both overlapping host species distribution, and ideal spring/summer climatic conditions at this overlapping range in order to facilitate simultaneous sporulation of *C. comandrae* spermatia and receptive *C. ribicola* spermagonia. Thus, if the hybrid does indeed exist naturally, the climatic conditions may not have been favourable for sexual hybridization in the 2017 and 2018 sampling seasons.

4.4 Insects Vectoring Pycniospores

One of the more complicated parts of the life cycle that has been majorly overlooked and understudied in the literature is the influence of insects which are suggested to facilitate dispersal of sexual spores. Unpublished research has shown that this facilitation is principally driven by species in the order *Diptera*, belonging primarily to two families: *Sciaridae*, the fungus gnats, and *Phoridae*, a family of scuttle flies, with specimens also belonging to the families *Drosophilae* and *Cecidomydae*, but in lesser frequencies (Nicole et al., Unpublished).
Other than the work done by Nicole et al., (Unpublished) there has been little to no research on the insects which are responsible for carrying pycniospores. It is likely that the same insect species would be attracted to pycelial nectar from both *C. ribicola* and *C. comandreae*, and thus if both species are sporulating in proximity the insects would not discriminate between species when collecting spores, and facilitate cross-over and potentially sexual hybridization. The question is whether these insect species are influenced by local climate or other factors, which would influence the prevalence of hybridization events, especially in the F1 generation.

Variations in local climate affect insects as they are ectotherms, that is their life cycle is highly dependent on the temperature of their surrounding environment (Jaworski & Hilszczanński, 2013). The same basic climate parameters (i.e. temperature and moisture) that influence the rust life cycle, also influence insects, including their larval stages, phenology, growth and overwinter survival (Jaworski & Hilszczanński, 2013). In temperate climates, such as in North Western North America, the timing of seasons is one of the most important factors in insect activity and phenology (Jaworski & Hilszczanński, 2013). In fact, one of the symptoms of climate change is alteration of the timing of appearance of some species, and the length of their activity (Walther et al., 2002). Local climate could be a major factor influencing hybridization events not only in the *Cronartium* spp. lifecycle directly but also indirectly through influencing the life cycle of their sexual spore dispersal agents.
5. Conclusions

Understanding the consequences of the hybridization event between *C. ribicola* and *C. comandrae* is important from both an evolutionary and a forest management standpoint. Interspecific hybridization events between forest pathogens are greatly understudied, and thus the species and genetic interactions from this event could provide valuable insights to the area of forest pathology. Considering the catastrophic effects of the introduction of *C. ribicola* to North America, the evolutionary consequences of a hybridization event involving this species needs to be evaluated, as it could lead to an alteration of virulence or a potential host jump.

This study indicates that *C. x flexili* is either no longer prevalent in areas where it was previously found, or else prevalent at such low frequencies that it evaded detection in the sampling effort of this study. Further, the populations which were analyzed did not demonstrate strong evidence of introgression, yet did not entirely rule out the possibility that this hybrid could facilitate gene flow between the introduced *C. ribicola* and the native *C. comandrae*. The results of the research presented here suggests that when a hybridization event occurs between *C. ribicola* and *C. comandrae*, hybrids may result but are likely infertile or have low fertility and do not appear to be occurring perennially. Some evidence suggests this could be due a number of factors including low relative fitness of *C. x flexili* compared to parental species, variations in local climate factors influencing life cycle parameters, or sexual incompatibility with either parental species.

Future surveillance of the hybrid zones could be necessary in order to determine if *C. x flexili* persists but evaded detection in this study. If *C. x flexili* is captured again, inoculations and virulence testing should be undertaken. Alternatively, artificial hybridization could be attempted in a laboratory setting to determine if hybrids can be formed from sexual crossing of gametes.
This work adds to the limited literature on the genetics of hybrid fungal pathogens of forests and improves our understanding of the evolutionary mechanisms occurring when allopatrically evolved forest fungal species hybridize.
Literature Cited


Nicole, M. C., Lavallée, R., & Hamelin R. C. (Unpublished). Identification of insects involved in white pine blister rust spermatization by PCR. Natural Resources Canada, Laurentian Forestry Center, Quebec, Canada.


**Appendices**

**Appendix A: Summary Table of Hybrid Forest Fungal Pathogens**

Table 17. Currently recognized and functionally described hybrid forest fungal pathogens.

<table>
<thead>
<tr>
<th>Hybrid Name</th>
<th>Group</th>
<th>Parent Lineages</th>
<th>Origin</th>
<th>Reference(s)</th>
</tr>
</thead>
</table>
| *Cronartium X flexili*             | Basidiomycete | *C. ribicola*  
                                          *C. comandrae*            | Introduced   | Joly et al., 2006           |
| *Heterobasidion irregulare X H. annosum* | Basidiomycete | *H. irregulare*  
                                          *H. annosum*            | Introduced   | Giordano et al., 2014       |
| *Heterobasidion annosum sensu stricto X H. annosum ISG P* | Basidiomycete | *H. annosum sensu stricto*  
                                          *H. annosum ISG P*        | Native       | Gonthier et al., 2007       |
| *Heterobasidion annosum ‘S-P’*    | Basidiomycete | *H. annosum ‘S’-  
                                          *H. annosum ‘P’*          | Native       | Olson & Stenlid, 2002       |
| *Heterobasidion annosum X H. occidentale* | Basidiomycete | *H. irregulare*  
                                          *H. occidentale*         | Native       | Olson & Stenlid, 2001       |
| *Melampsora X columbiana*         | Basidiomycete | *M. medusae*  
                                          *M. occidentalis*        | Introduced   | Newcombe et al., 2000       |
| *Melampsora lini AB*              | Basidiomycete | *M. lini AA*  
                                          *M. lini BB*             | Native       | Barrett et al., 2007        |
| *Melampsora medusae-populina*     | Basidiomycete | *M. medusae*  
                                          *M. larici-populina*     | Introduced   | Spiers & Hopcroft, 1994     |
| *Ophiostoma ulmi X O. novo-ulmi*  | Ascomycete   | *O. ulmi*  
                                          *O. novo-ulmi*           | Native       | Brasier et al., 1998        |
| *Phytophthora X alni*             | Oomycete     | *P. uniformis*  
                                          *P. x multiformis*       | Introduced   | Husson et al., 2015         |
| *Phytophthora X multiformis*      | Oomycete     | *P. unknown spp.*  
                                          *P. unknown spp.*        | Native*      | Aguayo et al., 2016         |

*The parental lineage and origin of the *Phytophthora x multiformis* hybrid is currently unknown (see Aguayo et al., 2016).*
Appendix B: Sampling Site Location Determination

Sampling sites were determined according to 1) where the hybrid pathogen has previously been identified and 2) where the parental species were the least likely to come in contact with one another. The Hybrid sampling sites were in Smithers BC, and Waterton Lakes AB. Joly et al. (2006) described three sampling sites in the Rocky Mountains of Alberta, Kananaskis County, Porcupine Hills and Waterton Lakes where he identified C. x flexili. Both sampling sites for pure C. ribicola are in western white pine plantations on the coast of British Columbia, specifically Bowser on Vancouver Island and Chilliwack. As these sites occur outside of the range of C. comandrae it is highly unlikely that contamination will have occurred. The sampling sites for pure C. comandrae were selected according to their elevation gradient (low) and proximity to Pinus strobus populations. The lower elevation plantation of lodgepole pine near to Prince George BC, and naturally occurring population in Kimberley BC were chosen as ideal locations.
Appendix C: Excerpt from Zamibino General Rust Protocols (1998)

Short-term 4 C storage of aeciospores and urediniospores. Urediniospores and aeciospores can be stored at 4C 31RH chambers for over seven weeks, but gradually lose viability, and urediniospores develop a need for heat shock for optimal germination. Germination drops off quickly if spores are stored at 0 RH or in a water- saturated environment.

Long-term storage of aeciospores and urediniospores. For small amounts of spores, gel capsules partially filled with spores are placed in small racks, such as inverted lids of 200 pipette tip boxes. at 4C in a 0 or 31 RH desiccation chamber (see Table below) for 7 days, after which the spores and bottom portion of each capsule are transferred to a screw-top 2.0 ml cryovials and frozen at -80 C. Although not tested at our lab, desiccated spores might also be in glass ampules for liquid nitrogen storage.

Desiccation chambers and desiccants: An inexpensive chamber for desiccating spores is made by placing desiccant into the bottom of a 5 cup square, flat Tupperware container, placing the bottom of a 3 cup container on the desiccant, and sealing the 5 cup container with a lid that indicates the target RH. A layer of Calcium chloride granules saturated with water provides 31 RH, suitable for desiccation of all spore stages before either short term storage a 4 C or long term storage at -80 C. The calcium chloride dissolves as it absorbs moisture; heating the solution in a glass container in an oven will drive off the water and the calcium chloride to be reused. Silica gel and Calcium sulfate (e.g., drierite) provide an RH near 0, useful for routine desiccation prior to DNA extract on or long-term storage of urediniospores. Lining 5 cup containers with wet paper towels provides 100 RH for rehydrating spores.

Rehydrating spores. Rapid changes in hydration from dry to wet conditions are damaging to rust spores, especially after low temperature storage. Blister rust urediniospores scraped from infected leaves by toothpick germinate readily without hydration, but even the act of collecting fresh spores into gelatin capsules using cyclone collectors will dry them sufficiently so that they will need rehydration for good germination. Rehydration is required for both aeciospores or urediniospores. Rehydration is accomplished by 4-5 hours at 100RH at 4C. Spores must be in a thin layer and may be on plastic trays or within microcentrifuge tubes or even in open- topped gelatin capsules within microcentrifuge tubes, however gelatin capsules without the support of a surrounding microcentrifuge tube will become too sticky to handle after exposure to high humidity.

Preparing suspensions of spores in dilute agar. Dilute agar prevents spores from settling to the bottom of the tube during inoculation. In contrast to the mineral oil Soltrol (the standard carrier for cereal rust spores during inoculations), agar is non-toxic to Ribes, other host plants, and Rehydrated aeciospores or urediniospores are suspended in either a solution of 0.25 % Tween 20 in extremely pure sterile water (glass distilled water, or water from a 1Tulti-stage deionizing process) or in sterile water lacking surfactant, before being centrifuged down and being resuspended in a solution of 0.07% select agar (Sigma Chemical Co. - made with highly purified water and cooled to room temperature at least several before use to allow for optimum viscosity.

Spore germination tests. Condensation is dried from the lids of plates of 1.5% agar (Sigma Chemical Co.) and the plates opened for several minutes in a biological safety hood until surface water is not present. Two locations are marked on each plate, each ca. 3/4 x 2 in, and drops of 0.07
% select agar placed so as to divide the location in thirds. To create areas of high and low spore
density, a drop of the spore suspension is added to the end of the location, spread up and down
along the 3/4 inch width using a sterile loop, continuing through the two diluting drops of agar.
Alternatively, fully rehydrated but dry spores can be added to one drop with a sterile toothpick or
loop, mixed well, and streaked as above. The plate is left open until the very thin layer of surface
moisture has evaporated into the agar.

Plates are incubated at 17 C or 20 C to test aeciospore or urediniospore germination,
respectively, although 17 C can be used for both spore types. Light neither promotes nor inhibits
germination of Cronartium aeciospores and urediniospores. For urediniospores, percent
germination is determined using a compound microscope at 200 X after ca. 16 hr. incubation.
Germination of aeciospores is often asynchronous, so germination is checked after 16 hr. and again
at 40 hr.

**Applying spore suspensions using aspirators.** Suspensions of aeciospores or urediniospores in
dilute agar are applied to small and large Ribes plants using perfume bottle bulb aspirators
(available for $3 at many local department stores). The bulb is unscrewed and replaced with a thin
plastic tube that is attached to a source of compressed air. 1 ml of a slightly yellow or orange
colored suspension of aeciospores or urediniospores is sufficient for inoculating a small plant
having leaves. I use 2 g of urediniospores to make ca. 300 ml of a spore suspension the color
intensity of dilute Tang that to inoculate 40 stems of large plants and have extra suspension left
over. For large plant inoculations, large amounts of spore suspension can be placed in the perfume
reservoir for large plant inoculations, without stopping the spray of spores between plants. Large
plants are inoculated in a growth chamber by holding the pot in one hand with the plant nearly
horizontal, the aspirator in the other hand, and spraying each leaf from base to top. When small
plants are inoculated, the plant pot is held in a plastic cup affixed to the inside of a large rubbermaid
container that is slightly inclined at the back, to hold the plant nearly horizontally while leaving
both hands free. The reservoir for the aspirator is not used for applying the spore suspension.
Instead, the tube that supplies the aspirator with suspension is briefly submerged in the
microcentrifuge tube containing the suspension, to provide a short pulse of spores to inoculate
each leaf.

After a plant has been inoculated on its lower leaf surfaces with a spore suspension the
upper surfaces of its leaves are immediately misted with distilled water. Small plants are bagged
in individual clear polypropylene bags, with the sealed end up, and the open end secured around
the pot with a rubber band. The open end should extend to the bottom of the pot, which is in a tray
of water. The inoculation chamber is sprayed with a solution of Alconox and surfaces of the
biological safety cabinet and gloves cleaned with alcohol several times between inoculations.
Small single stem plants are placed in greenhouse isolation boxes (Figure 2) after at least 24 hr
incubation at 20 C in a growth chamber. Large plants are all with the same strain, so they are
bagged 5-7 per plastic garbage bag, with bags sealed at the top with a bull clip. Bagged plants are
incubated in a growth chamber at 20C. Plants are fertilized and supplied with drip-line irrigation
after incubation.

All inoculated plants are kept in a low traffic greenhouses or growth chambers. separated
as far as possible from the greenhouse that contains uninfected stock plants. Only local strains are
produced during the natural growing season of wild Ribes. Urediniospores start to develop in 8
days, and can be collected after an additional 7-10 days. Prior to spore production, lateral buds are removed. Once spores have been harvested, the plants are cut back to soil level. The plant material is autoclaved, and the benches and greenhouse isolation cages misted well with water several times throughout a one-day period, as thorough contact with water causes spores to germinate or to die.
Appendix D: Hu’s DNA Extraction Protocol Modifications

The following modifications were made to Hu’s (2016) extraction protocol:

- Step 1: lyse tissue/spores with liquid nitrogen in a precooled tissuelycer using 30Hz for 1min 30
- Step 5: incubate on ice for 4 minutes
- Step 6: max speed = 14 000rpm (12 000 rpm in 4°C centrifuge)
- Step 7: repeat step 7 with 24:1 Chloroform Isoamylalcool to help remove any contaminating phenol
- Step 10: incubate at room temperature for 8 minutes

Primers/Probes Used:

Primer = DCON10 F+R (Forward AAAGCACTCACCCCTTCGGTCT; Reverse: GACTGGACACATTGGCTCCTT)

Probes: C. ribicola Dcon10 1 TGGCC+T+C+G+ACGTGT (FAM); C. comandrae Dcon10 1 TGGCC+C+T+C+ACGTGT (Cy5). The + sign denotes Locked Nucleic Acids.

Negative control: Buffer
DNA: Aeciospores

Spore preparation:
A small subset of aeciospores is transferred into the 0.1ml qPCR tubes. 40ul of Edwards buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS, with 1% PVPP (polyvinyl polypirrolidone), is then added to each PCR tube, pipetted, and spun. The spore solution is then placed in the thermocycler in incubate mode for 10 minutes at 95°C.

Reaction preparation:

DNA 2 µL (1/5 dilution of DNA extract)
DCON10 F 1 µL (10 µM)
DCON10 R 1 µL (10 µM)
Probe 4 µL (1 µM)
Quantitect mastermix 20 µL (2x)
Water 2 µL
TOTAL 20 µL

Probe were used as simplex, samples were run in duplicate with each probe, for a total of 4 reactions per sample.

Cycling parameters for Quantitect:
95°C 15:00
95°C 00:15
60°C 01:30
40 cycles

NOTES: The extraction buffer is not compatible with qPCR so it must be diluted to 1 in 5 order for the reaction to work. When collecting DNA for dilution some solid material is acceptable (e.g. spores). In order to analyze the results on Applied Biosystems QuantStudio™ Real – Time PCR Software V1.3 (Thermofisher) use fast 96 well plate, analyze results using the standard curve.
Appendix F: Microscopy

Figure 16. Scanning Electron Images (SEM) showing aeciospores from (A) C. ribicola (B) C. commandrae (C & D) C. x flexili collected in the Rocky Mountains of Alberta (Joly et al. 2006).

Figure 17. SEM images showing aeciospores from (A & B) C. ribicola collected in Smithers B.C. 2012 (C & D) C. x flexili collected in McKendrick Pass in 2012 by Alex Woods (E & F) C. commandrae collected in Montana 2012.
Appendix G: F-Statistics Calculations

Gene (allele) frequencies

\[ P = \frac{2 \times AA + Aa}{N \times 2} \]

\[ q = 1 - P \]

Expected genotypic counts under Hardy-Weinberg Equilibrium

Expected AA = AA \times p^2

Expected Aa = Aa \times 2 \times p \times q

Expected aa = aa \times q^2

Local observed heterozygosities of each subpopulation

\[ H_{obs} = \frac{Aa}{N} \]

Local expected heterozygosity of each subpopulation

\[ H_{exp} = 1 - \Sigma(p1^2 + q1^2) \]

Local inbreeding coefficient of each subpopulation

\[ F_s = \frac{H_{exp} - H_{obs}}{H_{exp}} \]

where a positive F means fewer heterozygotes than expected indicates inbreeding

\( P \) (p-bar, the frequency of allele A) over the total population

\[ \frac{2 \times AA_1 + Aa_1}{2N_1} + \frac{2 \times AA_2 + Aa_2}{2N_2} + \frac{2 \times AA_3 + Aa_3}{2N_3} \]

\( \bar{q} \) (q-bar, the frequency of allele a) over the total population

\[ \frac{2 \times aa_1 + Aa_1}{2N_1} + \frac{2 \times aa_2 + Aa_2}{2N_2} + \frac{2 \times aa_3 + Aa_3}{2N_3} \]

Global heterozygosity indices (over Individuals, Subpopulations and Total population)

HI based on observed heterozygosities in individuals in subpopulations

\[ HI = \frac{H_{obs1}(N_1) + H_{obs2}(N_2) + H_{obs3}(N_3)}{N_{Total}} \]

HS based on expected heterozygosities in subpopulations

\[ HI = \frac{H_{exp1}(N_1) + H_{exp2}(N_2) + H_{exp3}(N_3)}{N_{Total}} \]

HT based on expected heterozygosities for overall total population (using global allele frequencies)

\[ HT = 1 - \Sigma(\bar{p}2^2 + \bar{q}2^2) \]

GLOBAL F-STATISTICS
\[
F_{IS} = \frac{H_S - H_I}{H_S}
\]
\[
F_{ST} = \frac{H_T - H_S}{H_T}
\]
\[
F_{IT} = \frac{H_T - H_I}{H_T}
\]

\(P\) = proportion of allele ‘A’ in population
\(AA\) = number of 0/0
\(Aa\) = number of 0/1
\(N\) = number of individuals genotyped
\(q\) = proportion of allele ‘a’ in population
Appendix H: Statistical Hypotheses and Equations

Chi-Squared Test of Independence:

H₀: there is no association between the two variables
H₁: there is an association between the two variables

\[ \chi^2 = \sum_{i,j} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \]

where:
f_{ij} = observed frequency counts of events belonging to both ith category of x and jth category of y
e_{ij} = expected count if x and y are independent

Analysis of Variance (ANOVA):

H₀: the population mean of all of the groups are equal
H₀: \( \mu_1 = \mu_2 = \mu_3 \)

H₁: the population mean of all of the groups is unequal
H₁: \( \mu_1 \neq \mu_2 \neq \mu_3 \) or \( \mu_1 = \mu_2 \neq \mu_3 \) or \( \mu_1 \neq \mu_2 = \mu_3 \) or \( \mu_1 = \mu_2 = \mu_3 \neq \mu_2 \)

\[ y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + e_i \]

where:
b₀ = model intercept
b₁x₁i = coefficient (b₁) and the mean of population 1
b₂x₂i = coefficient (b₂) and the mean of population 2
b₃x₃i = coefficient (b₃) and the mean of population 3
e_i = model error

Kruskal-Wallis H-test:

H₀: the population median of all of the groups are equal
H₀: \( \mu_1 = \mu_2 = \mu_3 \)
H₁: the population median of all of the groups is unequal
H₁: \( \mu_1 \neq \mu_2 \neq \mu_3 \) or \( \mu_1 = \mu_2 \neq \mu_3 \) or \( \mu_1 \neq \mu_2 = \mu_3 \) or \( \mu_1 = \mu_3 \neq \mu_2 \)

The H-test statistic:
\[ H = \frac{12}{N(N + 1)} \left( \sum_{j=1}^{C} \frac{(T_j)^2}{N_j} \right) = 3(N + 1) \]

where:
N = sum of sample sizes for all samples
C = number of samples
\( T_i \) = sum of ranks in the \( j^{th} \) sample
\( C \) = size of the \( j^{th} \) sample

Mann – Whitney U test (Nonparametric Equivalent of the Two Sample T-test)

H₀: two randomly drawn samples from each population come from the same population (i.e. that they both have the same median)
H₁: two randomly drawn samples from each population come from different populations (i.e. that they both have different medians)

\[ U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} \]

where:
\( R \) = sum of ranks in the sample
\( n \) = number of items in the sample
Appendix I: sppIDer Pipeline Output Tables

*C. comandrae* samples mapped onto combination reference genome

**Figure 18.** Violin plot depicting mapping quality (MQ) score of *C. comandrae* reads mapped by species.

**Figure 19.** Log2 average depth of coverage of *C. comandrae* sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each base pair of the genome.
Figure 20. Log2 average depth of coverage of C. comandræ sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each region of the genome.
C. ribicola samples mapped onto combination reference genome

**Figure 21.** Violin plot depicting the mapping quality (MQ) score of C. ribicola reads mapped by species.

**Figure 22.** Log2 average depth of coverage of C. ribicola sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each base pair of the genome.
Figure 23. Log2 average depth of coverage of C. ribicola sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each region of the genome.
C. x flexili samples mapped onto combination reference genome

Figure 24. Violin plot depicting the mapping quality (MQ) score of C. x flexili reads mapped by species.

Figure 25. Log2 average depth of coverage of C. x flexili sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each base pair of the genome.
Figure 26. Log2 average depth of coverage of C. x flexili sample reads mapped onto the combination reference genome where depth of coverage represents a count of how many reads cover each region of the genome.