MOLECULAR CYTOGENETICS IN *PICEA*

by

GARTH ROBERT BROWN

B.Sc., The University of British Columbia, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Forest Sciences

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1995

© Garth Robert Brown, 1995
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature) ____________

Department of Forest Sciences
The University of British Columbia
Vancouver, Canada

Date October 11, 1995
ABSTRACT

A fundamental characteristic of a species is its karyotype, a description of each chromosome. In spruce (Picea) and other conifer genera, chromosome identification is hindered by similarities in chromosome size and morphology within a species. Homeologous chromosomes of related species are also generally indistinguishable. The focus of this Ph.D thesis is the development of a new technology in conifer cytogenetics, in situ hybridization, to address these inherent difficulties and to establish procedures for the physical mapping of the spruce genome.

In situ hybridization augments conventional cytogenetics with the methods of molecular biology and allows the visualization of defined DNA sequences along metaphase chromosomes and interphase chromatin. The chromosomal locations of three tandem repeated DNA sequences, including the genes encoding the 18S-5.8S-26S ribosomal RNA, those encoding the 5S ribosomal RNA, and a centromeric satellite DNA (SGR-31), were determined in white spruce (Picea glauca (Moench) Voss) and Sitka spruce (P. sitchensis (Bong.) Carr.). A molecular description of the 5S ribosomal RNA genes and SGR-31 in white spruce, comprising their nuclear organization, nucleotide sequence and genomic copy number, was also performed.

Combining in situ hybridization data from the 18S-5.8S-26S ribosomal RNA and the 5S ribosomal rRNA genes permitted the first unequivocal identification of each somatic chromosome of a conifer species. All three repeated DNAs were subsequently mapped onto the same metaphase spreads, producing the first cytogenetic maps of white and Sitka spruces. The white spruce map consists of twelve loci, seven corresponding to 18S-5.8S-26S
ribosomal RNA loci, one to the 5S ribosomal DNA, and four to SGR-31 sites. Comparison of this map to that of Sitka spruce (five 18S-5.8S-26S ribosomal RNA loci, one 5S ribosomal RNA site, and five SGR-31 loci) revealed that despite the overall uniformity in appearance among homeologous chromosomes, the repeated DNA complement of these closely related spruce genomes is in flux.

The prospects for the use of in situ hybridization in conifer genome analysis are immense. It's greatest contribution, which hinges on improvements in detection sensitivity, will be in the integration of genetically defined linkage groups with the chromosomes on which they reside. Investigating the relationship between genetic linkage, recombination and chromosome structure or the co-linearity of gene sequences and chromosomal synteny among spruce species and other conifer genera, and the eventual cloning of genes based on map position, are now within the possibilities of conifer genetics.
# TABLE OF CONTENTS

Abstract                                           ii  
Table of Contents                                  iv  
List of Abbreviations                              vii 
List of Tables                                     viii 
List of Figures                                    ix  
List of Common and Scientific Names of Gymnosperms x  
List of Common and Scientific Names of Angiosperms xi 
Acknowledgement                                    xii 
Foreword                                           xiii 

## GENERAL INTRODUCTION

1

| Literature Cited | 5 |

## Chapter One  
Literature Review  
8

- White Spruce and Sitka Spruce  
- The Nuclear Genome of Spruce  
- Organization and Evolution of Repeated DNA  
  - Tandem Arrayed Repeated DNA  
  - Dispersed Repeated DNA  
- Low Copy DNA Sequences in Conifer Genomes  
- *In Situ* Hybridization  
- Literature Cited  

## Chapter Two

Development of *In Situ* Hybridization in Spruce:  
the Chromosomal Location of Genes Encoding  
the 18S-5.8S-26S Ribosomal RNA  
30

<table>
<thead>
<tr>
<th>Introduction</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials and Methods</td>
<td>33</td>
</tr>
<tr>
<td>Plant Material and Chromosome Preparation</td>
<td>33</td>
</tr>
<tr>
<td>Probe Labeling</td>
<td>34</td>
</tr>
<tr>
<td><em>In Situ</em> Hybridization</td>
<td>35</td>
</tr>
<tr>
<td>Microscopy</td>
<td>37</td>
</tr>
<tr>
<td>Karyotype Analysis</td>
<td>37</td>
</tr>
<tr>
<td>Results</td>
<td>38</td>
</tr>
<tr>
<td>Chapter Three</td>
<td>Characterization of the 5S Ribosomal RNA Genes in Spruce</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Introduction</td>
<td>51</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>52</td>
</tr>
<tr>
<td>DNA Isolation</td>
<td>52</td>
</tr>
<tr>
<td>PCR Amplification and Cloning of the 5S rRNA Genes</td>
<td>53</td>
</tr>
<tr>
<td>Genomic Digestion and Southern Hybridization</td>
<td>54</td>
</tr>
<tr>
<td>DNA Sequencing</td>
<td>55</td>
</tr>
<tr>
<td>Estimation of 5S rDNA Copy Number in White Spruce</td>
<td>55</td>
</tr>
<tr>
<td>Chromosome Preparation and <em>In Situ</em> Hybridization</td>
<td>55</td>
</tr>
<tr>
<td>Results</td>
<td>56</td>
</tr>
<tr>
<td>Characterization of 5S rDNA in White Spruce</td>
<td>56</td>
</tr>
<tr>
<td>Physical mapping of the 5S rDNA</td>
<td>64</td>
</tr>
<tr>
<td>Discussion</td>
<td>65</td>
</tr>
<tr>
<td>Literature Cited</td>
<td>68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter Four</th>
<th>Characterization of a Centromeric Satellite DNA in White Spruce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>70</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>71</td>
</tr>
<tr>
<td>Plant Material and DNA Isolation</td>
<td>71</td>
</tr>
<tr>
<td>Screening of a Partial White Spruce Library</td>
<td>72</td>
</tr>
<tr>
<td>DNA Sequencing</td>
<td>74</td>
</tr>
<tr>
<td>Estimation of the SGR-31 Copy Number in White Spruce</td>
<td>74</td>
</tr>
<tr>
<td>Chromosome Preparation and <em>In Situ</em> Hybridization</td>
<td>79</td>
</tr>
<tr>
<td>Results</td>
<td>79</td>
</tr>
<tr>
<td>Characterization of SGR-31 in White Spruce</td>
<td>79</td>
</tr>
<tr>
<td>SGR-31 in <em>Picea</em></td>
<td>80</td>
</tr>
<tr>
<td>Physical Mapping of SGR-31</td>
<td>81</td>
</tr>
<tr>
<td>Discussion</td>
<td>82</td>
</tr>
<tr>
<td>SGR-31, a Centromeric Satellite DNA</td>
<td>82</td>
</tr>
<tr>
<td>Evolution of SGR-31 in <em>Picea</em></td>
<td>84</td>
</tr>
<tr>
<td>SGR-31 and Spruce Centromere Function</td>
<td>85</td>
</tr>
<tr>
<td>Literature Cited</td>
<td>87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter Five</th>
<th>The First Cytogenetic Maps of White and Sitka Spruces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>89</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>90</td>
</tr>
<tr>
<td><em>In Situ</em> Hybridization</td>
<td>90</td>
</tr>
<tr>
<td>Image Analysis</td>
<td>91</td>
</tr>
<tr>
<td>Results</td>
<td>91</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Discussion</td>
<td>96</td>
</tr>
<tr>
<td>Cytogenetic Maps of White and Sitka Spruce</td>
<td>96</td>
</tr>
<tr>
<td>Repeated DNAs of the Spruce Genome</td>
<td>97</td>
</tr>
<tr>
<td>Literature Cited</td>
<td>99</td>
</tr>
<tr>
<td>Chapter Six</td>
<td>101</td>
</tr>
<tr>
<td>The Prospects for Non-Isotopic <em>In Situ</em> Hybridization in Conifer Genome Analysis</td>
<td></td>
</tr>
<tr>
<td>The Sensitivity of <em>In Situ</em> Hybridization</td>
<td>101</td>
</tr>
<tr>
<td>Applications</td>
<td>104</td>
</tr>
<tr>
<td>Continued Mapping of Repeated DNAs</td>
<td>104</td>
</tr>
<tr>
<td>Low Copy Sequence and Gene Mapping</td>
<td>105</td>
</tr>
<tr>
<td>Literature Cited</td>
<td>109</td>
</tr>
<tr>
<td>Appendix</td>
<td>113</td>
</tr>
<tr>
<td>Common fluorochromes and their excitation and emission maxima</td>
<td></td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISH</td>
<td><em>in situ</em> hybridization</td>
</tr>
<tr>
<td>18S-26S rDNA</td>
<td>18S-5.8S-26S ribosomal DNA</td>
</tr>
<tr>
<td>DAPI</td>
<td>4'6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>FITC</td>
<td>fluorescein isothiocyanate</td>
</tr>
<tr>
<td>CMA&lt;sub&gt;3&lt;/sub&gt;</td>
<td>chromomycin A&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>A.R.</td>
<td>arm ratio</td>
</tr>
<tr>
<td>C.I.</td>
<td>centromeric index</td>
</tr>
<tr>
<td>R.L.</td>
<td>relative length</td>
</tr>
<tr>
<td>FL&lt;sub&gt;pter&lt;/sub&gt;</td>
<td>fractional length from terminus of the short arm (p)</td>
</tr>
<tr>
<td>IGS</td>
<td>intergenic spacer of the 18S-5.8S-26S ribosomal DNA</td>
</tr>
<tr>
<td>NTS</td>
<td>nontranscribed spacer of the 5S rDNA</td>
</tr>
</tbody>
</table>
I. *Picea* species included in Southern hybridization experiments with SGR-31. 73
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Molecular structure of a typical 18S-5.8S-26S ribosomal RNA gene in higher eukaryotes.</td>
<td>31</td>
</tr>
<tr>
<td>2.2</td>
<td>Localization of 18S-26S rDNA in white spruce.</td>
<td>39</td>
</tr>
<tr>
<td>2.3</td>
<td>18S-26S rDNA polymorphism in white spruce.</td>
<td>40</td>
</tr>
<tr>
<td>2.4</td>
<td>Localization of 18S-26S rDNA in Sitka spruce.</td>
<td>41</td>
</tr>
<tr>
<td>2.5</td>
<td>Ideograms of the mitotic chromosomes of white and Sitka spruces.</td>
<td>44</td>
</tr>
<tr>
<td>3.1</td>
<td>Origin of the primer pairs P1/P2 and P3/P4 used in the PCR amplification of the 5S rDNA.</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Consensus sequence of 5S rRNA genes in white spruce.</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Southern hybridization of 5S rDNA to partially digested white spruce genomic DNA.</td>
<td>59</td>
</tr>
<tr>
<td>3.4</td>
<td>Estimation of 5S rDNA copy number in white spruce.</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>Localization of 5S rDNA in white spruce.</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Localization of 5S rDNA in Sitka spruce.</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Southern hybridization of SGR-31 to partially digested white spruce genomic DNA.</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td>Nucleotide sequence of SGR-31.</td>
<td>76</td>
</tr>
<tr>
<td>4.3</td>
<td>Southern hybridization of SGR-31 to Sau3A digested genomic DNA from 18 species of <em>Picea</em>.</td>
<td>77</td>
</tr>
<tr>
<td>4.4</td>
<td>Localization of SGR-31 in white and Sitka spruces.</td>
<td>78</td>
</tr>
<tr>
<td>5.1</td>
<td>Co-localization of the 5S and 18S-26S rDNA, and SGR-31 in white and Sitka spruces.</td>
<td>92</td>
</tr>
<tr>
<td>5.2</td>
<td>The first cytogenetic maps of white and Sitka spruces.</td>
<td>93</td>
</tr>
</tbody>
</table>
### LIST OF COMMON AND SCIENTIFIC NAMES OF GYMNOSPERMS

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Scientific Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycas</td>
<td><em>Cycas revoluta</em> Thunb.</td>
</tr>
<tr>
<td>Ephedra</td>
<td><em>Ephedra kokanica</em> Regel.</td>
</tr>
<tr>
<td>Gingko</td>
<td><em>Gingko biloba</em> L.</td>
</tr>
<tr>
<td>Gnetum</td>
<td><em>Gnetum ula</em> Brongn.</td>
</tr>
<tr>
<td>Douglas-fir</td>
<td><em>Pseudotsuga menziesii</em> (Mirb.) Franco</td>
</tr>
<tr>
<td>Western redcedar</td>
<td><em>Thuja plicata</em> Donn ex D. Don</td>
</tr>
<tr>
<td>Pine, Loblolly</td>
<td></td>
</tr>
<tr>
<td>Radiata</td>
<td><em>Pinus taeda</em> L.</td>
</tr>
<tr>
<td>Scots</td>
<td><em>Pinus radiata</em> D. Don</td>
</tr>
<tr>
<td>Slash</td>
<td><em>Pinus sylvestris</em> L.</td>
</tr>
<tr>
<td></td>
<td><em>Pinus elliottii</em> Engelm.</td>
</tr>
<tr>
<td>Spruce, Black</td>
<td></td>
</tr>
<tr>
<td>Blue</td>
<td><em>Picea mariana</em> (Mill) B.S.P</td>
</tr>
<tr>
<td>Candalabra</td>
<td><em>Picea pungens</em> Engelm.</td>
</tr>
<tr>
<td>Chihuahua</td>
<td><em>Picea montigena</em> Masters</td>
</tr>
<tr>
<td>Engelmann</td>
<td><em>Picea chihuahuana</em> Martinez</td>
</tr>
<tr>
<td>Koyama</td>
<td><em>Picea engelmannii</em> (Parry)</td>
</tr>
<tr>
<td>Mexican</td>
<td><em>Picea koyamai</em> Shiras.</td>
</tr>
<tr>
<td>Norway</td>
<td><em>Picea abies</em> (L.) Karst</td>
</tr>
<tr>
<td>Oriental</td>
<td><em>Picea orientalis</em> (L.) Link</td>
</tr>
<tr>
<td>Red</td>
<td><em>Picea rubens</em> Sarg.</td>
</tr>
<tr>
<td>Sargent</td>
<td><em>Picea brachytyla</em> (Franch.) Pritz.</td>
</tr>
<tr>
<td>Schrenk’s</td>
<td><em>Picea schrenkiana</em> (Fisch and Meyer)</td>
</tr>
<tr>
<td>Serbian</td>
<td><em>Picea omorika</em> (Pancic) Purkyne.</td>
</tr>
<tr>
<td>Siberian</td>
<td><em>Picea obovata</em> Ledeb.</td>
</tr>
<tr>
<td>Sitka</td>
<td><em>Picea sitchensis</em> (Bong.) Carr.</td>
</tr>
<tr>
<td>Tigertail</td>
<td><em>Picea polita</em> (Sieb. and Zucc.) Carr.</td>
</tr>
<tr>
<td>White</td>
<td><em>Picea glauca</em> (Moench) Voss</td>
</tr>
<tr>
<td>Yeddo</td>
<td><em>Picea jezoensis</em> (Sieb. and Zucc.)</td>
</tr>
<tr>
<td>(unavailable)</td>
<td><em>Picea purpurea</em> Masters</td>
</tr>
</tbody>
</table>
# LIST OF COMMON AND SCIENTIFIC NAMES OF ANGIOSPERMS

<table>
<thead>
<tr>
<th>Common Name</th>
<th>Scientific Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td><em>Arabidopsis thaliana</em> (L.) Heynh.</td>
</tr>
<tr>
<td>barley</td>
<td><em>Hordeum vulgare</em> L.</td>
</tr>
<tr>
<td>broad bean</td>
<td><em>Vicia faba</em> L.</td>
</tr>
<tr>
<td>Chinese cabbage</td>
<td><em>Brassica campestris</em> L.</td>
</tr>
<tr>
<td>garden pea</td>
<td><em>Pisum sativum</em> L.</td>
</tr>
<tr>
<td>lily</td>
<td><em>Lilium speciosum</em> Thumb.</td>
</tr>
<tr>
<td>kale</td>
<td><em>Brassica oleracea</em> L.</td>
</tr>
<tr>
<td>maize</td>
<td><em>Zea mays</em> L.</td>
</tr>
<tr>
<td>mung bean</td>
<td><em>Vigna radiata</em> (L.) Wilczek</td>
</tr>
<tr>
<td>onion</td>
<td><em>Allium cepa</em> L.</td>
</tr>
<tr>
<td>rapeseed</td>
<td><em>Brassica napus</em> L.</td>
</tr>
<tr>
<td>rice</td>
<td><em>Oryza sativa</em> L.</td>
</tr>
<tr>
<td>rye</td>
<td><em>Secale cereale</em> L.</td>
</tr>
<tr>
<td>soybean</td>
<td><em>Glycine max</em> (L.) Merr.</td>
</tr>
<tr>
<td>tomato</td>
<td><em>Lycopersicon esculentum</em> Mill.</td>
</tr>
<tr>
<td>wheat</td>
<td><em>Triticum aestivum</em> L. em Thell.</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

My gratitude is extended to the members of my supervisory committee: Dr. John E. Carlson, for the opportunity and guidance, Dr. F. Brian Holl, for his continual support and encouragement, and Dr. Yousry El-Kassaby for his attitudes and views, research materials and collaborative support on behalf of Pacific Forest Products, Ltd.

I would like to thank Gyula Kiss of the Kalamalka Research Station, Vernon, B.C. for the collection and steady supply of seeds and additional research materials from the spruce arboretum. I am also indebted to Michael Weis for his microscopy and computer skills and his willingness to share this knowledge, and to Vindhya Amarasinghe for her constructive criticism and reviews of all aspects of this research. I am grateful for the efforts of Craig Newton and for the friendship of Andree, Ann, and Jeff. My thanks go to all other members of Dr. Carlson’s lab, past and present, for an enjoyable and stimulating work environment, to Jim and Scott for showing me the straightest path, and to Paul Bauman for planting the seed of my interest in cytogenetics.
FOREWORD

A portion of the research described in Chapter 2 has been previously published by Mr. Garth R. Brown, Dr. Vindhya Amarasinghe, Mr. Gyula Kiss and Dr. John E. Carlson under the title “Preliminary karyotype and chromosomal localization of the ribosomal DNA sites in white spruce using fluorescence in situ hybridization” in the journal Genome, volume 36, pp. 310-316, 1993. The written permission of the copyright holder to include this material in the thesis has been obtained.

The thesis author, under the supervision of Dr. Carlson, conducted all experimental procedures detailed in the publication with the exception of the operation of the confocal laser scanning microscope (performed by Dr. Amarasinghe). The necessary spruce materials were collected and supplied by Mr. Kiss.

First author

Senior Author
GENERAL INTRODUCTION

Molecular biology provides powerful new tools for the study and manipulation of plant genomes. Variations in the DNA sequence between homologous loci can be used as genetic markers in conventional linkage analysis (Botstein et al., 1980) and genetic linkage maps are in progress to supplement research programs oriented towards the improvement of many agricultural crops. Advances in DNA marker technology, in particular the application of the polymerase chain reaction (Williams et al., 1990), have simplified and reduced the time and cost associated with mapping. Genetic maps can now be rapidly constructed for virtually any sexually reproducing species. Among the species attracting the interest of molecular biologists are several economically important members of the Pinaceae family of conifers, including white spruce, slash pine, loblolly pine, and Douglas-fir (Tulsieram et al., 1992; Nelson et al., 1993; Devey et al., 1994; Jermstad et al., 1994). Genetic mapping of conifer genomes may allow the molecular dissection of both simple and complex traits and will provide a much broader insight into the structure, organization, and evolution of the nuclear genome than previously possible using markers for morphological or biochemical traits. Additionally, if associations between genetic markers and quantitative trait characters in specific crosses can be consistently detected, genetic mapping may provide the means of shortening the long generation time particular to conifer breeding programs through marker-assisted selection.

In studying the genome organization of any species it is prudent to augment genetic linkage analysis with the characterization of repeated DNAs and the underlying physical
relationship of both repeated and low copy DNA sequences along the chromosomes. A large portion of the genomes of higher plants, in fact more than 75% of total DNA in those exceeding 2 picograms per nucleus, is composed of repeated DNA sequences (Flavell 1980; Thompson and Murray 1981). Some repeated DNAs, such as the genes encoding the components of ribosomes, have critical cellular roles while others serve no discernible protein-encoding function. In its preponderance, repeated DNA is clearly a major determinant of chromosome size, structure and functioning. It is, however, less amenable then low copy DNA to linkage analysis due to the occurrence of repeating units of a single family found at multiple loci and difficulties in identifying segregating alleles.

A further incentive to characterize the repeated DNA component of any genome in parallel to genetic mapping arises from the unequal distribution of recombination events observed across both animal and plant genomes (Leitch and Heslop-Harrison 1993; Gustafson and Dille 1992; Lukaszewski 1992; Groover et al., 1995). Since genetic distance is defined by the recombination frequency between markers, chromosomal regions which suppress recombination, like heterochromatic areas composed of tandem arrayed repeated DNA families, will lack genetic markers (Flavell et al., 1985). The primary constriction, or centromere, is one such chromosomal domain and can result in tightly linked markers that in reality lie in the distal portion of opposite chromosome arms separated by hundreds of megabases of DNA. Additional recombination "cold spots" such as inversions and translocations have been described (Lucchesi and Suzuki 1968) as have "hot spots" which result in markers appearing unlinked genetically even though they are physically very near to one another (Steinmetz et al., 1987). Therefore, the genetic distance between markers can frequently be a poor indicator of the true physical distance involved.
**In situ** hybridization (ISH), one of several methods available to generate a physical map, combines conventional cytology and methods of molecular biology, enabling the localization of defined DNA sequences along metaphase chromosomes or interphase chromatin. The technique involves the hybridization of labeled DNA or RNA to cytological preparations of denatured metaphase chromosomes and chromatin which have been fixed on a microscope slide. The sites of probe hybridization can be observed microscopically via radioactive or fluorescent emissions, or by the enzymatic production of a colored precipitate.

ISH has been widely applied to the physical mapping of genes and repeated DNA sequences in animal species and, in particular, is a central component of the Human Genome Project. The presence of a cell wall in plants and the available methods of plant chromosome preparation present difficulties not experienced in human cytogenetics. While these may currently limit the routine detection of gene loci in plants, the physical mapping of repeated DNAs has been achieved in a wide variety of angiosperms. These studies, along with observations that repeated DNAs are a dynamic, rapidly evolving part of the genome not subject to the evolutionary constraints imposed on transcribed sequences and those that regulate them, have illustrated their value as cytological markers not only to investigate genome organization but also in the identification of somatic chromosomes and in the study of chromosome evolution, phylogenetic relationships, introgression and the spatial arrangement of chromatin in the interphase nucleus (Lapitan *et al.*, 1989; Lapitan *et al.*, 1987; Zhang and Dvorak 1989; Bauwens *et al.*, 1991).

By necessity, ISH will play a central role in the physical mapping of conifer genomes, serving as the primary tool in chromosome identification and in integrating genetic and physical maps. Most conifers, including spruce species, have symmetrical karyotypes.
composed of predominantly metacentric chromosomes (Khoshoo 1961) which pose serious
difficulties in identifying each chromosome pair based solely on morphological criteria.
Chromosome identification is, however, an important step in genome analysis.

A variety of chromosome banding techniques has been developed to facilitate
chromosome identification in many animals and some angiosperm plant species. These
methods produce either transverse bands spanning the length of all metaphase chromosomes
or darkly staining regions in particular sites of some or all chromosomes. As many as 2000 G-
bands can be produced along midprophase chromosomes of humans, greatly simplifying
karyotyping. However, the method is generally not applicable to plants (Bickmore and
Sumner 1989). Plant cytogenetics has traditionally relied on techniques which produce much
fewer bands of the second type. For example, Giemsa C- banding, which originated as a
modification of ISH (Pardue and Gall 1970), reveals heterochromatic areas associated with
centromeres, telomeres and some interstitial sites on one to all chromosomes of a
complement. Certain fluorescent dyes which have specificity for GC or AT base pairs, such as
Chromomycin A₃ and 4′-6-diamidino-2-phenylindole (DAPI), respectively, reveal
chromosomal regions with sequence compositions distinct from adjacent areas (Schweizer
1980). Fluorescent bands usually correspond to a subset of the heterochromatic regions
revealed by C-banding. Lastly, N-bands denote the sites of actively expressed ribosomal RNA
genes through the detection of a complex of acidic nonhistone argentophilic proteins bound
with the ribosomal genes (Goessens 1984). Unfortunately, none of these chromosome
banding methods have been applied to spruce cytogenetics. Currently, the best approach to
karyotyping spruce and other conifers is the isolation and physical mapping by ISH of one or
several repeated DNA sequences having genomic distribution(s) specific to each chromosome.
In many agricultural crops the integration of genetically mapped DNA markers with the physical map is readily achieved by filter hybridization to genomic DNA from a variety of well characterized chromosome addition or substitution lines (Chinoy et al., 1991). These cytogenetic stocks have not been produced for any conifer species and, at present, map integration in conifers will likely only proceed by determining the chromosomal location of linkage groups using ISH methods.

At the onset of this research, ISH procedures had been demonstrated for only a single conifer species (Cullis et al., 1988). Repeated DNAs had also not been fully characterized or exploited in any gymnosperm. To correct this situation and to access the potential of ISH for analysing the spruce genome the specific aims of this thesis were: 1) the development of a method to localize repeated DNA sequences to the mitotic chromosomes of spruce species by ISH and 2) the cloning and characterization of repeated DNA sequences yielding easily distinguished chromosome-specific hybridization patterns of white spruce (Picea glauca (Moench) Voss) and Sitka spruce (P. sitchensis (Bong.) Carr.).

LITERATURE CITED


CHAPTER 1

Literature Review

Chapter 1 summarizes the pertinent literature concerning the genus *Picea*, repeated DNAs and the development and use of ISH in genome analysis. This chapter is not intended as a comprehensive review, in particular of tandem repeated DNA organization and evolution which are discussed throughout, but rather for sufficient background information for the reading of this thesis.

**WHITE SPRUCE AND SITKA SPRUCE**

The species of primary interest to this research are the closely related spruces white spruce (*Picea glauca* (Moench) Voss) and Sitka spruce (*P. sitchensis* (Bong.) Carr.), two of the approximately 40 species of spruce recognized (Mikkola 1969). White spruce is one of the most widely distributed spruce species, ranging across boreal America from coastal Alaska to Newfoundland and south into Montana, Wisconsin, Michigan and the New England states. In British Columbia it is found throughout most of the interior on the eastern side of the Coastal Mountains at elevations below 1000 meters. Throughout much of the B.C. interior, the natural range of white spruce overlaps with that of Engelmann spruce (*P. engelmanni* (Parry)), which occupies the higher elevations from 1000 metres up to timberline in the central and southern Rocky Mountains (Owens 1982). Effective barriers to interspecific
hybridization do not exist between these two species resulting in complex hybrid "swarms" in areas where the species are sympatric (Roche 1969), a common occurrence among many different species of spruce (Wright 1955). From an operational perspective, the B.C. Ministry of Forests makes no distinction between white spruce, Engelmann spruce and their hybrids. The species complex is denoted simply as "interior spruce" because of the considerable interspecific hybridization and similarities in the species’ cultural regimes in seedling nurseries and in planting sites.

Sitka spruce ranges from Kodiak Island, Alaska to northern California, confined predominantly to a coastal belt below 700 metres in elevation and seldom wider than 80 kilometers. Along several drainage systems, notably the Nass, Bulkley and Skeena Rivers in the Coastal Mountains and the Skagit River in the Cascades, Sitka spruce may be found as far as 150 kilometers inland where its range also frequently overlaps with that of either white spruce to the north or Engelmann spruce to the south (Hosie 1990). Zones of introgressive hybridization occur between Sitka spruce and both white and Engelmann spruces where their ranges are sympatric (Fowler 1987; Hosie 1990).

Interior spruce is extensively harvested in B.C. owing to its predominance and value as pulpwood, lumber, and as a source of specialty wood products (Owens 1982). It is an important component of reforestation programs in the province with an estimated 100 million seedlings being required for artificial regeneration by the year 2000 (Kiss and Yeh 1988). Sitka spruce is also extensively harvested although its value in reforestation in B.C. is presently limited, in particular owing to a high incidence of plantation failure due to susceptibility to the white pine terminal weevil (Pissodes strobi Peck) (Heppner and Wood 1984).
THE NUCLEAR GENOME OF SPRUCE

With the exception of chromosome numbers and information provided by the kinetics of DNA fragment reassociation in solution it is not known how the nuclear genome of spruce, and indeed any conifer, is organized. Early cytological investigations of the genus noted that all spruce species are diploids with the nuclear genome packaged into 12 chromosome pairs, i.e., 2n = 24 (Sax and Sax 1933; Santamour 1960). Only the smallest 4 chromosome pairs can generally be distinguished in metaphase preparations using conventional stains, the largest 8 pairs being metacentric and of similar sizes.

These preliminary surveys revealed a remarkable conservation of karyotype morphology among spruce species as well as other conifer genera (e.g. in Pinus, Pederick 1970). This fact, and the difficulties in identifying each mitotic chromosome of a species, has severely impacted the field of conifer cytogenetics and limited cytological contributions to phylogenetic studies in spruce and other gymnosperms.

Spruce chromosomes are large in comparison to many angiosperm species and this size is reflected in estimates of the DNA content of the nuclear genome. Using Feulgen microspectrophotometry, the diploid nucleus of white spruce contains 17 picograms of DNA (Dhillon 1987) which translates into approximately 8.5 X 10⁹ base pairs per haploid genome. In comparison, the genome of spruce is 100 times the size of the Arabidopsis genome, the smallest genome among flowering plants, 20 times that of rice and 3-4 times that of maize (Arumuganathan and Earle 1991). Among gymnosperms, nuclear DNA amounts vary approximately 12-fold (Ohri and Khoshoo 1986). Excluding Gnetum with clearly the smallest gymnosperm genome (4.5 pg) the range is reduced to approximately 4-fold. This is in
contrast to the 100-fold variation among diploid angiosperms and reflects the similarities between related species and many genera of gymnosperms observed in cytological studies.

The difference in genome sizes between Arabidopsis and species with large genomes is unlikely to arise from a need for a battery of new gene products. In fact, as has been shown by reassociation kinetics, increasing genome size is positively correlated with an increasing amount of repeated DNA (Lapitan 1992). In Arabidopsis, only about 15% of the genome is repeated (Leutwiler et al., 1984) whereas in species with greater than 2 pg of DNA per nucleus the repeated DNA fraction typically exceeds 75% of the genome, much of which has no apparent cellular function (Flavell 1980). Studies on the reassociation kinetics of conifer genomes confirm this generality: approximately 68% of the white spruce genome reanneals with repeated DNA kinetics (Rake et al., 1980), a value which likely underestimates the true repeated DNA content of the spruce genome since the technique fails to include repeated sequences which are too short or have diverged to such an extent as to prevent stable DNA duplexes forming in solution.

ORGANIZATION AND EVOLUTION OF REPEATED DNA

DNA reassociation kinetics describes repeated DNA sequences solely in terms of their genomic copy number. Only since the development of techniques in molecular biology has the complete characterization of repeated sequences become possible. While no attempts had been made prior to the beginning of this research to study repeated DNA families in conifers or other gymnosperms, a considerable volume of information concerning their organization and evolution in animals and angiosperm plants has been amassed. These studies have
revealed that repeated DNA sequences are found in two distinct classes, either arranged as tandem arrays or interspersed with unrelated repeated or unique DNA sequences (Flavell 1986).

**Tandem repeated DNA sequences**

Tandem arrays are composed of closely related repeating units arranged in a "head-to-tail" manner and are typically the most highly represented DNA sequence families in complex genomes. Many distinct families can coexist within a genome, differing in sequence and complexity, and in genetic activity. Among these are the least complex repeats of the microsatellites (repeating units of 6 bp or less) and the minisatellites (with repeating units of 11-60 bp), the satellite DNAs composed of tandemly repeated units of 150-500 bp lengths and originally named for their distinct banding position away from the main band DNA on cesium chloride density gradients, the genes encoding the 18S, 5.8S and 26S ribosomal RNAs and the 5S ribosomal RNA, and the specialized, highly conserved tandem repeats with the consensus sequence of \((T/A)_nG_{1.8}\) at the telomeres of each eukaryotic chromosome (Zakian 1989; Ganal et al., 1991). The organizational features of relevant repeated DNA families are discussed in detail in the introductions of Chapters 2, 3 and 4.

The origin of tandem repeated DNA sequences in a genome undoubtedly involves some form of amplification event at specific chromosomal sites, such as centromeres and telomeres, or wherever it is tolerated. Although the precise mechanisms are not understood, several possibilities have been proposed, including unequal crossing over between two sequences (Smith 1976), excision followed by a rolling circle type of DNA replication and reintegration (Hourcade et al., 1973), slippage replication (Tautz and Renz 1984) and
aberrant in situ replication (Schimke 1982).

Tandem arrays of a given sequence family are frequently found in similar positions on more than one, if not all chromosomes of a complement. Thus, arrays can be divided or duplicated and transposed to both homologous and non-homologous chromosomes. Again, although the details of intragenomic movement of tandem repeated sequences are sketchy, Flavell (1985) suggests it may result from double crossovers between interacting chromosome segments or the excision of a circular segment of arrays and its reintegration elsewhere. In this regard, the association of chromosomes in intermitotic nuclei ("Rabl polarization"; Rabl 1885) described in some (but not all) plant and animal cell nuclei (Comings 1980) is intriguing. Rabl polarization creates a unique ordering of chromosomes within a cell such that non-homologous chromosomes of most similar length are adjacent to one another. Centromeres are then closely associated as are telomeres on chromosome arms of similar length. Such a physical environment could facilitate the recombination necessary for transposing members of a tandem array onto homologous and non-homologous chromosomes leading to the fixation of an array in a species.

Since most tandem arrays, with the exception of ribosomal RNA genes and the telomeric repeat structure, appear not to have a strictly sequence-dependant function, they accumulate base substitutions, deletions and insertions at higher rates than protein-coding genes. However, both the repeats within a single tandem array and all members of a sequence family in a given genome do not evolve independently. Instead, DNA sequence analysis reveals very high homology among repeats within a species (Lapitan 1992). They therefore appear to evolve "in concert" (Arnheim et al., 1980; Dover 1982) through mechanisms such as unequal crossing over and/or gene conversion events (Flavell 1985). It is precisely because
of comparatively rapid divergence between species and concerted evolution within a species that tandem repeated DNA sequences have attracted considerable attention as phylogenetic tools.

The debate over the role(s) of tandem repeated DNA (or heterochromatin) continues. How does one ascribe function(s) to DNA sequence families which make up a highly variable component of the genome in terms of copy number, length, sequence and organization? Proponents of the "junk" DNA hypothesis maintain that tandem repeated DNA, and repeated DNA in general, have no functional significance, simply accumulating and evolving due to selective neutrality (Orgel and Crick 1980; Doolittle and Sapienza 1980). As Macgregor and Sessions (1986) note: "in no case is there any obvious causal relation between amounts of satellite DNA and morphological change within a group of related animals. It is possible for satellite sequences to be lost from both tissue culture cells and in vivo without obvious somatic effects". Others hold the view that tandem repeated DNA is selected for its involvement in chromosome pairing, recombination or organization of chromosomes in the nucleus (Flavell 1982). At least in Drosophila, meiotic pairing seems to require heterochromatic homology (Irick 1994). Martinez-Zapater et al. (1986) propose the involvement of satellite DNA in the organization of DNA into nucleosomes based on the correspondence of satellite DNA lengths in many species and the distance between nucleosomes. Some satellites are transcribed (although a cellular function cannot be assigned or the transcriptional event is due to read-through from adjacent structural genes: Macgregor and Sessions 1986). The most recent addition to the debate is the enticing postulate of Vogt (1992) in which tandem repeated DNAs permit the establishment and stabilization of specific chromatin folding structures in distinct chromosome regions such as centromeres and
telomeres.

\textit{Dispersed repeated DNA sequences}

Species with very large genomes have not nearly enough tandem repeated DNA sequence families to account for the amount of DNA in excess of protein coding requirements (Smyth 1991). Much of these large genomes, for example more than 50\% of the genome of cereals (Flavell \textit{et al.}, 1981), is composed of repeated DNA sequences found interspersed with other repeated DNAs or unique sequences. Dispersed repeats of 5-2000 bp may be found in large genomes on average every 200-4000 bp of unique sequence DNA (Flavell 1980). By contrast, in small genomes, the length of unique sequence DNA between dispersed repeats may exceed 130,000 bp (Ganal \textit{et al.}, 1988). Dispersed repeats are more difficult to characterize due to their association with diverse neighboring sequences and lower copy number than most tandem arrayed repeated DNAs. However, several have now been studied in a number of angiosperms including rye, tomato, rice, maize and lily (Rogowsky \textit{et al.}, 1990; Ganal \textit{et al.}, 1988; Mochizuki \textit{et al.}, 1992; Schwarz-Sommer \textit{et al.}, 1987; Smyth \textit{et al.}, 1989) revealing considerably more size and structural heterogeneity among different copies in a genome than tandem arrayed sequences.

The genomic organization of dispersed repeats is consistent with the discovery that many have the properties of mobile genetic elements. Several dispersed elements in plant genomes have been identified as transposable elements, mobile via a DNA intermediate and encoding a transposase gene and sequences necessary for recognition and transposition. An example of this type is the \textit{Activator} transposon of maize (McClintock 1948). However, since their copy numbers generally do not exceed several hundred per genome at the most, this class
of dispersed repeat appears to contribute little to genome size (Smyth 1991). Other mobile elements have structural similarities to retrotransposons, transposing via an RNA intermediate. Among these, two distinct classes have been identified based on whether or not long terminal repeats are produced during amplification (Smyth 1991). In contrast to transposable elements, some retrotransposon types of dispersed repeats have been found in high copy number in several genomes, including the IFG element present in approximately 10,000 copies in the sugar pine genome (C.S. Kinlaw, personal communication) and the del2 element at 240,000 copies in lily (Leeton and Smyth 1993). Although not all dispersed repeats share features in common with transposons or retrotransposons, most contain inverted or direct repeats (Flavell 1982) which suggest that they may be the remnants of once active mobile elements. The potential for mobile elements to have contributed significantly to the large genome size of plants is further illustrated by electron microscopic observations that as much as 10% of wheat DNA can form hairpin structures under high stringency conditions (Flavell 1984).

Clearly the frequent amplification and transposition of mobile elements could be a selective disadvantage which might explain why few active elements in plants have been discovered. At least in the case of retrotransposons, subsequent activity may be precluded by the error prone nature of reverse transcriptase, such that inserted elements may already be diverged in sequence and unable to move further. Although gene conversion processes can act to maintain the sequence homogeneity of a dispersed repeat family, Scherer and Davis (1980) noted that this mechanism operates at lower frequency than with tandem arrayed repeats. Thus over time the individual members of a dispersed repeated family may diverge and eventual become part of the single copy sequence component of a genome (Smyth 1991).
LOW COPY DNA SEQUENCES IN CONIFER GENOMES

No discussion on genome organization is complete without mention of the distribution of protein-coding and other single- or low-copy sequences. As noted previously, 25-30% of the genome of white spruce and other conifers reassociates with single- or low-copy kinetics. One might question why in large genomes like those of conifers the repeated DNA fraction does not exceed 95%, relegating the low- and single-copy fraction to one up to several percent. Although the percentage of repeated DNA is positively correlated with increasing DNA content in plant species, single copy sequences are also found to increase in large genomes (Hutchinson et al., 1980). Therefore, conifer genome size is not simply a function of an elevated repeated DNA content.

Prior to RFLP mapping, isozyme linkage analysis was used to investigate the organization of a limited portion of the low copy component of conifer genomes. Conservation of isozyme linkage groups among pine species and other Pinaceae, including white spruce, was used to infer that no major chromosome rearrangements had occurred in these regions during conifer evolution (Conkle 1981).

RFLP mapping vastly improves upon the number of available markers for genetic analysis and consequently yields a much more detailed view of the organization of low-copy sequences. Only in loblolly pine, however, are RFLP maps sufficiently advanced to allow meaningful insights. Of 65 cDNAs mapped by Devey et al., (1994), 24% detected more than one segregating locus that could be mapped. According to the authors, many other cDNAs detected multiple loci but the complexity of the hybridization pattern prevented their mapping. This is in contrast to the genome of Arabidopsis in which 98% of RFLPs map to
single loci (Chang et al., 1988).

While polyploidy in the course of evolution can be responsible for multiple loci of a given gene, as in maize (Helentjaris et al., 1988), this phenomenon has not been a regular feature of conifer evolution (Khoshoo 1959). Thus, numerous multigene families have arisen by other means in the loblolly pine genome, conceivably by transposition, and in part account for the large fraction of single- and low-copy DNA. The possibility that much of this component is actually composed of ancient families of dispersed repeats which have diverged enough to prevent their inclusion in a repeated DNA family has yet to be investigated.

**IN SITU HYBRIDIZATION**

The basic principles of ISH to cytological preparations were independantly established more than 25 years ago by Gall and Pardue (1969) and John et al. (1969). These pioneering efforts used tritium-labeled RNA probes to visualize ribosomal RNA in *Xenopus* oocytes. In the following years, isotopic ISH was adapted to agricultural plants and used to determine the chromosomal location of both repeated DNA sequences (Bedbrook et al., 1980; Hutchinson et al., 1981) and single copy DNA sequences (Shen et al., 1987). Isotopic ISH is very sensitive particularly in human cytogenetics where unique sequences as small as 500 base pairs have been detected on metaphase chromosomes (Jhanwar et al., 1983). The technique, however, suffers from several drawbacks relating to the use of radioactive labels, including long exposure times, limited resolution due to scattering of radioactive emissions and their capture in an emulsion overlay, a high degree of background signal necessitating statistical analysis of hybridization data, and safety and disposal concerns.
The major advances and widespread application of ISH techniques have arisen with the development of non-radioactive labeling methods. The most common method is indirect labeling in which hapten-like reporter molecules (typically biotin) are incorporated into hybridization probes and subsequently detected by the appropriate enzymatic or fluorescent affinity reagents. While labeling methods have been developed in which nucleotides labeled with fluorescent molecules are incorporated directly into hybridization probes, they currently provide lower sensitivity than indirect methods in most cases.

Langer et al. (1981) described the synthesis and use of deoxyuracil-5'-triphosphate (dUTP) to which biotin, a member of the vitamin B complex, had been attached to the pyrimidine ring. Analogs of dUTP, dCTP and dATP are now commercially available and each biotin-labeled nucleotide can be readily incorporated into hybridization probes by nick translation (Rigby et al., 1977), random primer labeling (Feinberg and Vogelstein 1983) or the polymerase chain reaction (Saiki et al., 1985; Mullis et al., 1986). Additional reporter molecules include digoxigenin and dinitrophenol which may be incorporated into probes in similar fashion, and aminoacetylfluorene, mercury and sulfonate which are attached to DNA through chemical reactions (Trask 1991).

The detection of biotin-labeled probes after hybridization to chromosomes is achieved through enzymatic or fluorescent conjugates of either anti-biotin antibodies or avidin, a glycoprotein extracted from egg white. The affinity of avidin (and related molecules such as streptavidin and ExtrAvidin) for biotin is higher than immunological methods making it the method of choice for biotin detection. Horseradish per oxidase which catalyses the polymerization of 1,2-diaminobenzidine in the presence of hydrogen peroxide, producing a brown precipitate at the hybridization site, has been used as an avidin-enzyme conjugate.
(Rayburn and Gill 1985). Amplification of the hybridization signal by prolonging the reaction can improve the sensitivity of this method for the detection of less abundant DNA sequences (Jiang and Gill 1994).

Fluorescent tags are increasing in popularity since first reported (Langer-Safer et al., 1982). A variety of fluorochromes, such as fluorescein isothiocyanate, rhodamine, Texas Red, amino methyl coumarin acetic acid, Cy3, and Cy5, each with distinct spectral properties, have been conjugated to either antibodies or avidin, enabling the simultaneous hybridization and independent localization of different DNA probes (Leitch et al., 1991). (The excitation and emission characteristics of commonly used fluorochromes are presented in Appendix I). Signal amplification can also be achieved using anti-avidin or anti-biotin antibodies (Pinkel et al., 1986). The most impressive use of fluorescent ISH to date has undoubtedly been in human cytogenetics where as many as 7 cosmid probes have been physically mapped in a single experiment by combinatorial labeling (Ried et al., 1992).

Another important advantage of fluorescence ISH in relation to other reporter systems is the development of sophisticated digital imaging instruments including the confocal laser scanning microscope and the cooled charge coupled device camera. Not only can different fluorochromes be detected separately using specific filters, but digital images are generated which can be analysed by computer software providing an added degree of precision to physical mapping.

The simplicity and benefits of non-radioactive ISH have led to its broad use in the analysis of plant genomes. Rayburn and Gill (1985) were the first to report the use of biotin-labeled probes in plants, determining the location of a tandem repeated DNA family on the chromosomes of wheat. In the past decade, additional tandem repeated sequences including
the multigene families of the ribosomal RNAs, the 5S rRNA and specific cloned sequences have been mapped in many angiosperms, such as wheat, rye, barley, rice, soybean, tomato, rapeseed, onion and Arabidopsis (Appels et al., 1980; Bedbrook et al., 1980; Leitch and Heslop-Harrison 1992; Song and Gustafson 1993; Skorupska et al., 1989; Ganal et al., 1988; Xia et al., 1993; Ricroch et al., 1992; Bauwens et al., 1991). Recently ISH in several species of pine has also been reported (Cullis et al., 1988; Karvonen et al., 1993; Doudrick et al., 1995). Among the useful contributions to the understanding of plant genome structure provided by these studies have been the detection of additional ribosomal RNA sites not detectable by Southern hybridization (Mukai et al., 1991) and the molecular characterization of telomeric heterochromatin in cereals (Bedbrook et al., 1980), heterochromatic knobs in Arabidopsis (Maluszynska and Heslop-Harrison 1991) and Giemsa C-bands in rye (Mukai et al., 1992). As a karyotyping tool, the physical mapping of tandem repeats has enabled the identification of somatic chromosomes in tomato (Lapitan et al., 1989), and in combination with fluorescent dye banding patterns, the chromosomes of slash pine (Doudrick et al., 1995). The chromosomal distribution of dispersed repeated sequences has also been demonstrated in several angiosperm genomes (Ganal et al., 1988; Moore et al., 1991).

Genomic ISH, an extension of the basic procedure, is an important development in the analysis of polyploid species and hybrids (Le et al., 1989; Schwarzacher et al., 1989). By combining labeled total DNA from one parental species with unlabeled total DNA from the other parental genome(s) in the hybridization mixture, only species-specific DNA sequences of the labeled genome, which in all likelihood are repeated sequences, remain single stranded and capable of producing hybridization signals. Providing that the genomes contributing to the polyploid or hybrid species are sufficiently diverged with respect to their repeated DNA
compositions, the chromosomes or chromosomal segments from the different genomes are clearly differentiated after ISH. This approach, therefore, does not require the molecular cloning of species-specific repeated DNA sequences. Genomic ISH has been used to determine genome origin, relatedness and evolution in allopolyploids such as hexaploid wheat (Mukai et al., 1993). From a breeding perspective, the method has been used to monitor the introgression of alien chromatin from rye into wheat breeding lines (Schwarzacher et al., 1992). Wide crossing strategies provide valuable sources of genetic variability in agricultural breeding programs. Genomic ISH can efficiently and accurately reveal the location of translocation breakpoints and the amount of alien chromatin remaining in the crop species in subsequent backcross generations.

Non-radioactive ISH of low copy or unique DNA sequences to plant chromosomes has proven difficult, although not impossible. A number of factors inherent in plant cytogenetics likely account for the limited success, including the presence of the plant cell wall, the low mitotic index of root tips and the methods of plant chromosome preparation. These factors and the methods available to circumvent the difficulties they present in the physical mapping of low copy DNA sequences are discussed in detail in Chapter 6.

LITERATURE CITED


Moore, G., Cheung, W., Schwarzacher, T. and Flavell, R.B. 1991. BIS1, a major component of the cereal genome and a tool for studying genomic organization. Genomics, 10: 469-476.


CHAPTER 2

Development of *In Situ* Hybridization in Spruce: the Chromosomal Location of Genes Encoding the 18S-5.8S-26S Ribosomal RNA

INTRODUCTION

The genes encoding the 18S-5.8S-26S ribosomal RNA (18S-26S rDNA) were among the first isolated and have been studied at the molecular and cytological levels in a wide variety of organisms. From 500 to 40,000 copies of the 18S-26S rDNA, in some cases comprising up to 10% of the total DNA, are found in tandem arrays at one to several chromosomal sites in plant genomes (Rogers and Bendich 1987). Each repeating unit is typically from 7-12 kb in length in angiosperm plants and consists of the coding region for the RNA products and associated spacers (Fig. 2.1).

Within the nontranscribed or intergenic spacer (IGS), sequence repetition is also observed in all organisms. In some lower eukaryotes, such as yeast and slime mold, the 5S rDNA is located here (Appels and Honeycutt 1986). In higher eukaryotes, the 5S rDNA is not linked to the 18S-26S rDNA, and instead, a small tandem subrepeat is found. For example, there are 11 copies of a 131-133 bp element in wheat (Appels and Dvorak 1982). In *Drosophila* and *Xenopus*, the subrepeats share homology with the site of RNA polymerase I transcription initiation and, while this feature is not observed in all species analysed, it has led to the postulation that the subrepeats serve as sites for the loading of RNA polymerase I.
Fig. 2.1. Molecular structure of a typical 18S-5.8S-26S ribosomal RNA gene in higher eukaryotes. The precursor rRNA consists of an external transcribed spacer (ETS) preceding the 18S rRNA gene at the 5' end of the transcript, the 5.8S rRNA separated from both the 18S and 26S rRNA by intervening spacers (ITS-1 and ITS-2), and the 26S rRNA at the 3' end of the transcript. The vertical lines within the intergenic spacer denote sequence repetition observed in all organisms (Appels and Honeycutt 1986). In white spruce, the restriction enzyme BgII has a single recognition site within this subrepeat. Several subrepeats have been cloned and sequenced, revealing a 131 bp repeating element (Brown, Newton and Carlson, unpublished).
Transcription of the rDNA can be observed cytologically both in interphase and during cell division. During interphase, the nucleolus forms around the 18S-26S rDNA arrays within which the chromatin is dispersed and transcribed. For this reason, the 18S-26S rDNA loci are frequently referred to as the nucleolus organizer regions (NOR). The rRNA molecules associate with proteins and accumulate within the nucleolus before being exported to the cytoplasm. As the cell cycle continues, rRNA synthesis decreases in prophase and the nucleolus disintegrates. Although the chromatin which was actively transcribed in the nucleolus condenses along with the remainder of the metaphase chromosome, it remains associated with nucleolar proteins and stains poorly in comparison to adjacent chromatin, giving rise to a visible gap or secondary constriction (Goessens and Lepoint 1974). It is now generally accepted that secondary constrictions are the sites of rRNA genes that were actively transcribed in the preceding interphase. Because of the high GC content of the 18S-26S rDNA, secondary constrictions are also easily revealed by fluorescent dyes which preferentially bind GC or AT base pairs, such as Chromomycin A3 and DAPI, respectively. In the former case they appear as brightly staining regions and, in the latter, as negatively staining regions.

Since the 18S-26S rDNA has both a large physical size and a visible site on metaphase chromosomes, it is an ideal DNA sequence to use in developing ISH methodology. Additionally, although the IGS exhibits both sequence and length variation in many plant species (Rogers and Bendich 1987), the coding sequences are highly conserved across a wide evolutionary spectrum obviating the need to screen genomic libraries or to isolate hybridization probes by other means. In this chapter, a heterologous probe from soybean (Zimmer et al., 1988) was used to determine the chromosomal location of 18S-26S rRNA
genes in white spruce and Sitka spruce. The findings described herein have been published by the author in the journal *Genome* (Brown *et al.*, 1993).

**MATERIALS AND METHODS**

*Plant material and chromosome preparation*

Open pollinated seeds from one white spruce plus tree (Parent #5548 of the British Columbia Ministry of Forests Tree Improvement Program) originally located northeast of Prince George, B.C. were provided by Gyula Kiss of the Kalamalka Research Station. Sitka spruce seeds, collected from a Sitka spruce seed production orchard were provided by Dr. Yousry El-Kassaby of Pacific Forest Products, Ltd. Both species are diploids with $2n = 24$ and 1 or more B chromosomes occasionally present. For uniform germination, Sitka spruce seeds were imbibed in distilled water overnight and stratified for 3 weeks at 4 °C. This treatment was not necessary with the white spruce seedlot used. Seeds of both species were germinated on water saturated filter papers in a growth chamber for 6 days in the dark. To accumulate root tip cells in metaphase, the 1-2 cm long germinants were treated with 0.2% (w/v) aqueous colchicine for 6 hours. Following overnight fixation in 3:1 ethanol:acetic acid, root tips were stored in 70% ethanol at 4 °C until used.

Microscope slides were cleaned thoroughly by washing in 2 M chromic acid for three hours followed by several rinses in distilled water and storage in ethanol. Prior to squashing, root tips were rinsed several times in 0.01M sodium citrate/citric acid buffer (enzyme dilution buffer). A 10X stock solution of this buffer was prepared by mixing 40 ml of 0.1 M citric acid and 60 ml of 0.1 M trisodium citrate. The cell walls were digested to facilitate chromosome
spreading and the penetration of hybridization probes by incubating root tips in a solution of 2% cellulase (Calbiochem: 11,300 U/g), 1% macerase (Calbiochem: 3100 U/g) and 2% liquid pectinase (Sigma: 9 U/g) in 1X enzyme dilution buffer for 1 hour at 37 °C. Meristematic regions were then excised, teased apart in 45% acetic acid, and squashed under a coverslip. Slides were scanned for quality metaphase spreads on a Zeiss Axiophot microscope under phase contrast. The coverslip was removed from high quality slides with a razor blade after freezing on dry ice. Slides were then immersed briefly in 100% ethanol and air dried. These could be stored at room temperature for several weeks without noticeable effect on hybridization.

Chromosomes were stained prior to ISH in 0.2 μg/ml 4',6-diamidino-2-phenylindole (DAPI) in phosphate buffered saline (PBS, 0.13 M NaCl, 7 mM Na₂HPO₄, pH 7.4). DAPI staining was visualized using a Zeiss UV filter block and well spread metaphases were photographed on TMAX 400 film at ASA 1600. A micrometer scale bar was also photographed at the same magnification (100 X). Chromosomes were then destained in 3:1 methanol:acetic acid for 30 minutes at room temperature, rinsed three times in methanol and air dried.

Probe labeling

pGmR1, supplied by E. Zimmer (Zimmer et al., 1988), contains a 7.9 kb EcoR1 fragment in pBR325 encoding the 18S-5.8S-26S rDNA and the intergenic spacer sequences of soybean (Glycine max (L.) Merr.). Plasmid DNA was purified from 250 ml bacterial cultures by alkaline lysis and cesium chloride ultracentrifugation (Sambrook et al., 1989). The entire plasmid molecule was labeled with biotin-14-dATP using a commercially available nick
translation kit according to the supplier's recommendations (BRL Life Technologies). The 50 μl labeling reaction consisted of 1 μg pGmR1, 20 μM each of dCTP, dGTP, dTTP and biotin-14-dATP, 50 mM Tris-HCl (pH 7.8), 5 mM MgCl₂, 10 mM 2-mercaptoethanol, 2 U of E. coli DNA Polymerase I and 200 pg of DNase I. Labeling was performed at 16 °C for 1.5-2 hours and the reaction was then terminated by adding EDTA to 30 mM and SDS to 1.25%. Unincorporated nucleotides were removed on Sephadex G-50 (Pharmacia) spun columns or by ethanol precipitation. Lastly, an aliquot (50 ng) of the purified probe was run on a 2% agarose minigel in 1X TAE (0.4 M Tris-acetate, 0.001 M EDTA, pH 8) to verify that the average fragment size was less than 500 bp.

**In situ hybridization**

Slides were incubated with 20 μg/ml DNase-free RNAase A (Boehringer-Mannheim: 500 μg/ml) in 2X SSC (0.6 M NaCl, 0.06 M sodium citrate, pH 7) at 37 °C for 1 hour. Following three washes in 2X SSC at room temperature for 5 minutes each, slides were immersed in 0.01 M HCl for 2 minutes and then incubated in a solution of pepsin (Sigma: 3700 U/mg of protein used at 5 μg/ml in 0.01 M HCl) for 8 minutes at 37 °C. The reaction was stopped by rinsing slides in distilled water for 2 minutes. Slides were then washed 3 times in 2X SSC for 5 minutes at room temperature and dehydrated in a 70, 90, and 100% ethanol series.

Chromosomal DNA was denatured in one of two ways. Initial experiments performed this step by incubating slides in 70% formamide/2X SSC, pH 7.0, for 2 minutes at 72 °C. Formamide used in the denaturing solution and hybridization mixture was de-ionized with BIO-RAD AG 501-X8 ion exchange resin for 4 hours and filtered. Slides were then
immediately transferred through an ice-cold ethanol series and air dried. The hybridization mixture, consisting of 50-150 ng of labeled pGmR1, 7.5 μg of sheared salmon sperm DNA, 50% formamide, 10% dextran sulfate, 2X SSC, and 0.1% SDS, was then denatured at 80 °C for 10 minutes, chilled on ice, and applied to the slide. Hybridization proceeded at 37 °C for 12-16 hours in a humidity chamber. Although this method was effective, the development of equipment to automate chromosome denaturation (Heslop-Harrison et al., 1991) simplified the ISH protocol and improved consistency between experiments. Chromosomal DNA denaturation was subsequently performed using the Omnitirne Hybaid Temperature Cycler. In this situation, the probe hybridization mixture was denatured as described above, chilled on ice, and applied to air dried slides following pepsin treatment. Once a coverslip was put in place, slides were incubated at 80 °C for 10 minutes in the temperature cycler to denature the chromosomal DNA and then slowly cooled to 37 °C according to Heslop-Harrison et al. (1991). Slides remained in the temperature cycler at 37 °C for 12-16 hours.

After hybridization, slides were washed twice in 50% formamide/2X SSC for 5 minutes at 42 °C, 2X SSC for 5 minutes at 42 °C, and 2X SSC for 5 minutes at room temperature.

For the detection of biotin-labeled probes, slides were washed in 0.2% (v/v) Tween-20 in 4X SSC (detection buffer) for 5 minutes at room temperature. They were then incubated in 5% BSA in detection buffer at room temperature for 5 minutes to block non-specific avidin binding. Slides were treated with 100 μl of ExtrAvidin conjugated to fluorescein isothiocyanate (FITC) (Sigma: diluted 1:100 in 5% BSA in detection buffer) at 37 °C for 1 hour and then washed 3 times in detection buffer at 37 °C for 8 minutes. Chromosomal DNA was counterstained for 3-5 seconds in 0.1 μg/ml propidium iodide, rinsed in 2X SSC and
mounted in 50% glycerol in PBS.

Microscopy

In situ hybridized chromosomes were imaged using a BIO-RAD MRC 600 confocal laser scanning system equipped with an argon/krypton laser. FITC and propidium iodide were excited simultaneously using 488 nm and 538 nm laser lines. The resulting green and red emissions were separated to the two photomultipliers using K1 and K2 filter blocks. Fluorescein images were obtained in the photon counting mode (with signal accumulation to peak) and the propidium iodide images were taken in the direct mode with Kalman averaging. Images of hybridization sites and counterstained chromosomes were merged using Adobe Photoshop 3.0 software on a Macintosh Quadra 840 personal computer.

Karyotype analysis

Photographs of DAPI stained metaphases (and micrometer scales printed at the same enlargement) were digitized using an AGFA Studioscan II flatbed scanner. Chromosomes were randomly numbered 1-24 and measured using the public domain NIH Image program (written by W. Rasband at the U.S. National Institutes of Health and available from the Internet by anonymous ftp from zippy.nimh.nih.gov or on floppy disk from NTIS, 5285 Port Royal Rd., Springfield, VA 22161, part number PB93-504868). After calibrating the software using the known distance on the micrometer scale, total chromosome length was obtained by computer measurement of a line drawn down the middle of each chromosome. The positions of the centromere and secondary constriction were marked and, using the Plot Profile command, values for the length of the short arm (p), long arm (q) and the fractional length of
the secondary constriction from the terminus of the short arm (FLpter) were derived. The arm ratio (q/p), and centromeric index (p/p+q x 100) were then calculated. Homologous pairs were identified based on the measurements and hybridization patterns of 5 well spread metaphases and validated by the hybridization patterns from an additional 20 cells. Chromosome pairs were ordered according to convention from longest to shortest based on the relative length of each chromosome pair expressed as a percentage of the diploid cell complement.

RESULTS

Of critical importance to the successful application of ISH to plant chromosomes is the ability to generate many well spread metaphases free of the cell wall and cytoplasmic debris. Conventional methods of obtaining squash preparations from plant root tips entail the use of HCl hydrolysis to soften the middle lamella and improve chromosome spreading. However, cell wall material remains and, in addition, acid hydrolysis depurinates nucleic acids potentially resulting in the loss of chromosomal DNA target sequences. Cell walls and cytoplasmic components will also inhibit the penetration of probe molecules and interact with elements of the detection system causing high levels of background signal.

Three factors designed to ameliorate these difficulties were optimized for the preparation of acceptable spruce root tip metaphases. Because spruce chromosomes are comparatively long, an extended treatment with a colchicine solution considerably stronger than used with most angiosperm germinants was required to bring about an adequate degree of chromosome condensation and to allow satisfactory spreading. The cell wall was
Fig. 2.2. Localization of the 18S-26S rDNA in white spruce. a) Fluorescence ISH of biotin-labeled pGmR1 (green) to the mitotic chromosomes of white spruce (2n = 24). Chromosomal DNA was counterstained with propidium iodide (red). b) The same metaphase stained with DAPI. *Asterisk*, an unpaired B chromosome. Bar represents 10 μm.
Fig. 2.3. 18S-26S rDNA polymorphism in white spruce. a) Fluorescence ISH localization of pGmR1 and b) DAPI staining reveal that only 13 rDNA loci are present in this metaphase. Arrows, chromosome pair showing rDNA polymorphism. Asterisk, an unpaired B chromosome. Bar represents 10 μm.
Fig. 2.4. Localization of the 18S-26S rDNA loci in Sitka spruce. 
a) Fluorescence ISH of biotin-labeled pGmR1 (green) to the mitotic 
chromosomes of white spruce (2n = 24). Chromosomal DNA was 
counterstained with propidium iodide (red). b) The same 
metaphase stained with DAPI. Bar represents 10 μm.
effectively removed from most cells by incubating the root tips in a mixture of cellulase, macerase and pectinase prior to squashing. Lastly, most of the cytoplasm which still enclosed many metaphases in the squash preparations was degraded by a pepsin incubation prior to chromosome denaturation and ISH.

Hybridization sites of the biotin-labeled 18S-26S rDNA probe were detected as green fluorescence from the FITC conjugated ExtrAvidin. Non-hybridizing chromosomal sequences fluoresced red due to the propidium iodide counterstain. Images were acquired using a confocal laser scanning microscope and, while the optical sectioning capability of this system is of little added benefit on thin squash preparations, the ability to record the FITC signal independently from that of the propidium iodide counterstain signal provides higher sensitivity in detecting hybridization signals compared to epifluorescent microscopy. Over the course of developing the ISH technique, more than 50 intact metaphases of both white spruce and Sitka spruce were analysed. The number of pGmR1 hybridization sites observed in white spruce varied from 9 to 14 which was likely due to the quality of metaphase spreads given that loci on overlapping chromosomes can be masked from probe or detection reagents. Quality improved with experience and the 10 best spreads in which one or no chromosomes overlapped were selected for further analysis. In 6 metaphases (one shown in Fig. 2.2), 14 pGmR1 hybridization sites were counted. The same chromosome spread stained with DAPI, in which secondary constrictions appear as negative staining regions, shows that each hybridization site is associated with a secondary constriction, indicating that all loci were transcriptionally active in the previous interphase. The probe did not hybridize to the B chromosome. Therefore, the 18S-26S rDNA in white spruce are located on 7 of the 12 chromosome pairs. In the remaining 4 cells, only 13 hybridization sites and secondary
constrictions were observed (Fig. 2.3) suggesting that one of the white spruce chromosome pairs is polymorphic for this character. Of additional note in these metaphases was a very elongated secondary constriction found on one of the large metacentric chromosomes.

ISH and DAPI staining results in Sitka spruce are shown in Fig. 2.4. In all metaphases observed, 10 pGmR1 sites were evident, each with a corresponding secondary constriction. Although not present in the metaphase pictured, pGmR1 did not hybridize to the B chromosome(s) of Sitka spruce.

The mitotic chromosomes of white and Sitka spruce are represented by the ideograms in Fig. 2.5 arranged from longest to shortest according to the relative length of each pair averaged over 5 well spread metaphases. Pairing of homologous chromosomes was based primarily on the length, arm ratio, and presence and location of the secondary constriction or pGmR1 hybridization signal (given as the FLpter value). In most cases a simple visual assessment was sufficient to identify and pair 20 of the 24 chromosomes in each species, the exceptions being chromosomes 2 and 5 in Fig. 2.5. With careful measurements of high quality metaphases these can be distinguished by total length differences. Somewhat fortuitously, chromosome 5 also contains the site of the 5S rRNA genes in both species as described in Chapter 3.

Although in situ hybridization is not an entirely quantitative procedure, in white spruce the FITC signal on chromosome 6 was usually the most intense while that on chromosome 10 was the faintest, sometimes appearing as a discrete dot on each chromatid. Therefore it is likely that the 18S-26S rDNA arrays on chromosome 6 and 10 correspond to the longest and shortest, respectively, in the white spruce genome. Additionally, karyotyping spreads showing the possible rDNA polymorphism indicated that the potentially polymorphic locus in white
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.R.</td>
<td>1.08 (0.05)</td>
<td>1.12 (0.03)</td>
<td>1.03 (0.02)</td>
<td>1.21 (0.05)</td>
<td>1.05 (0.04)</td>
<td>1.09 (0.05)</td>
<td>1.04 (0.03)</td>
<td>1.29 (0.06)</td>
<td>1.72 (0.07)</td>
<td>1.33 (0.10)</td>
<td>1.17 (0.08)</td>
<td>2.02 (0.17)</td>
</tr>
<tr>
<td>C.I.</td>
<td>0.49 (0.01)</td>
<td>0.46 (0.00)</td>
<td>0.50 (0.01)</td>
<td>0.45 (0.01)</td>
<td>0.49 (0.01)</td>
<td>0.48 (0.01)</td>
<td>0.49 (0.01)</td>
<td>0.44 (0.01)</td>
<td>0.37 (0.01)</td>
<td>0.43 (0.02)</td>
<td>0.46 (0.02)</td>
<td>0.33 (0.02)</td>
</tr>
<tr>
<td>R.L.</td>
<td>10.01 (0.11)</td>
<td>9.28 (0.10)</td>
<td>9.25 (0.23)</td>
<td>9.24 (0.36)</td>
<td>8.88 (0.17)</td>
<td>8.50 (0.32)</td>
<td>8.37 (0.18)</td>
<td>8.33 (0.22)</td>
<td>7.78 (0.23)</td>
<td>7.46 (0.18)</td>
<td>6.81 (0.18)</td>
<td>6.10 (0.09)</td>
</tr>
<tr>
<td>FLpter</td>
<td>0.74 (0.01)</td>
<td>0.32 (0.01)</td>
<td>0.19 (0.01)</td>
<td>0.23 (0.01)</td>
<td>0.13 (0.02)</td>
<td>0.72 (0.01)</td>
<td>0.19 (0.13)</td>
<td>0.25 (0.01)</td>
<td>0.66 (0.01)</td>
<td>0.20 (0.01)</td>
<td>0.22 (0.01)</td>
<td>0.13 (0.01)</td>
</tr>
</tbody>
</table>

Fig. 2.5. Ideograms of the mitotic chromosomes of white spruce (top) and Sitka spruce (bottom), both 2n = 24. The upper arm is the short arm (p) by convention. A.R. = arm ratio, C.I. = centromeric index, R.L. = relative length, FLpter = fractional length of the 18S-26S rDNA locus from the terminus of the p arm, as described in the Materials and Methods.
spruce was found on chromosome 2. Only slight variation in hybridization signal intensity was observed among the five Sitka spruce loci.

As expected among closely related species with a high degree of sexual compatibility, many of the chromosomes share morphological features in white and Sitka spruces, including ranking, arm ratios, and the location of the 18S-26S rDNA. Notable discrepancies between the species include the absence of the 18S-26S rDNA arrays on chromosome 8 and 10 in Sitka spruce, different positions of the 18S-26S rDNA of chromosomes 2 and 3, and slight morphological differences between the smaller three chromosome pairs.

DISCUSSION

A fluorescence ISH procedure for the detection of highly repeated DNA sequences on metaphase chromosomes of spruce was developed and used to map the location of the genes encoding the 18S-26S rDNA in white and Sitka spruce. Cross-hybridization of the highly conserved coding regions of the ribosomal DNA repeat from soybean revealed that the 18S-26S rDNA in white spruce are distributed at one site on seven chromosome pairs. In Sitka spruce these loci are found on five pairs. All loci correspond to the positions of secondary constrictions at metaphase implying that each is actively transcribed during interphase in the spruce root tip meristem.

ISH of rDNA probes has recently been reported in a number of pine species. Ten major sites and several minor sites were observed in radiata pine by Cullis et al. (1988), and 16 sites were observed in both Scots pine and slash pine (Karvonen et al., 1993a; Doudrick et al., 1995). Considered with the data presented here on white and Sitka spruce, the genes
encoding the 18S-26S rRNA are more extensively distributed in these conifer genomes in comparison to angiosperms.

The tandem organization of a repeated DNA family confers a susceptibility to molecular mechanisms capable of altering the number of repeats in a given array. Unequal crossing over, whereby improper meiotic pairing and recombination results in a deletion in one chromatid or chromosome and a corresponding duplication in the other, is one of several suggested mechanisms and has been reported to occur in 18S-26S rDNA arrays in yeast (Szostak and Wu 1980). Since a hierarchy of repeat organization exists within the 18S-26S rRNA array, consisting of an IGS subrepeat embedded within the repeating unit itself, both the size of the 18S-26S rDNA repeat itself and the number of repeats at a given locus are subject to change.

Variation in the length of the 18S-26S rDNA repeat can arise by unequal crossing over between subrepeats within the IGS. Among angiosperms, heterogeneity has been found at all levels, for example, between species related to maize (Zimmer et al., 1988), between individuals of wheat (Appels and Dvorak 1982), and within individuals of a number of species, the extreme example being broad bean in which as many as 20 different length variants within a single plant have been observed (Rogers et al., 1986). Although a molecular characterization of the 18S-26S rRNA repeating unit of white and Sitka spruce was not undertaken in this study, Bobola et al. (1992) investigated the gene family in two other North American spruces, black spruce (Picea mariana) and red spruce (P. rubens). In both species, the coding regions and intergenic spacer, comprising as much as 4% of the genome, ranged from 32 kb to greater than 40 kb in length, considerably larger than that found in angiosperm plants. Restriction mapping revealed that as many as five different repeat sizes are found
within any individual. Up to four different 18S-26S rDNA repeat lengths within an individual tree of Scots pine have also been reported (Karvonen et al., 1993b).

Although the individual members of a tandemly repeated DNA family can show considerable sequence or length variation in a single genome, repeats within any one array are frequently more homogeneous than would be expected if each was evolving independently (Arnheim et al., 1980). The concerted evolution of 18S-26S rDNA repeating units within the two nucleolus organizers of pea was suggested by Polans et al. (1988) who noted that two 18S-26S rDNA length variants segregated in a Mendelian manner, indicating that the composition of rDNA repeats at each NOR is distinct. Although linkage analysis of rDNA loci in conifers is hampered by the multitude of sites and the difficulty in identifying segregating alleles, the number of length variants in black spruce corresponds with the number of secondary constrictions observed by Nkongolo and Klimaszewska (1994).

Unequal crossing over between repeating units outside the IGS subrepeat will give rise to variation in the total number of 18S-26S rRNA genes among individuals as well as the size of array at a particular chromosomal site. Extensive variability in rDNA copy number has been reported in many populations of angiosperms (Rogers and Bendich, 1987). In conifers, Strauss and Tsai (1988) reported a 5-fold variance among 54 individuals sampled from the range of Douglas-fir and a 28-fold variance within individuals of radiata pine was found by Cullis and Teasdale (1985). Such variation is presumably tolerated because only a small fraction of the 18S-26S rDNA, perhaps as little as 10%, is transcribed (Rogers and Bendich, 1987).

The observed variance in rDNA copy number among individuals predicts that in outcrossing species, such as spruce, heterozygosity in copy number at a particular locus
should arise. While more difficult to document, Miller et al. (1980) clearly demonstrated its occurrence in wheat using radioactive ISH in which grain counts provided a direct measure of rDNA copy number. The detection of only 13 of the 14 sites in well spread metaphases of white spruce could likewise be ascribed to great differences in rDNA copy number on the homologs of chromosome 2 as opposed to the complete absence of an array. Since the alternate homozygous state (i.e., 12 pGmR1 hybridization sites due to the absence of a rDNA locus on both homologs) was not observed among the well spread metaphases analysed, the maternal parent from which the open pollinated seeds were collected may be heterozygous for a rare 18S-26S rDNA deletion. That the secondary constriction found on one homolog of chromosome 2 in these metaphases was very elongated suggests that unequal recombination and/or the decondensation and transcription of greater numbers of rDNA genes has occurred.

Alternatively, if Engelmann spruce differs from white spruce in its rDNA distribution, specifically lacking the locus on chromosome 2, than a white X Engelmann hybrid would carry only 13 hybridization sites. While the possibility of contaminating Engelmann spruce pollen cannot be ruled out, preliminary DAPI staining and ISH with pGmR1 in Engelmann spruce suggest that the two species have identical 18S-26S rDNA chromosome locations (data not shown). Therefore, interspecific hybridization is unlikely to account for the 13 18S-26S rDNA loci observed in some metaphases.

Mapping additional repeated DNA sequences in spruce suggests that the ISH procedure described here is at least capable of detecting 100 kb of DNA on a metaphase chromosome. Therefore, given the large physical size of the 18S-26S rDNA repeat in spruce, one to several copies of the gene are expected to reside at the undetected site on chromosome 2. This situation exemplifies the current difficulty in interpreting ISH results in
plants for which the lower limit of detection is seldom known, that is, that in the absence of reproducible methods to localize single copy DNA sequences one cannot be certain that all genomic locations of a given sequence have been identified.

LITERATURE CITED


Characterization of the 5S Ribosomal RNA Genes in Spruce

INTRODUCTION

Ribosomal 5S RNA (5S rRNA) is a component of all ribosomes except in some mitochondria. Like the 18S-26S rDNA, the structure and organization of genes encoding the 5S rRNA have been extensively studied. Nucleotide sequences of at least 30 5S rRNA and 42 5S rDNA from higher plants are known (Barciszewska et al., 1994). In a typical eukaryotic genome, hundreds to thousands of genes are maintained as long tandem arrays of repeating units composed of a highly conserved 120 nucleotide coding sequence and a nontranscribed spacer (NTS). In plants, spacer lengths ranging from 95 bp to 730 bp in mung bean and radiata pine, respectively, have been reported (Hembleden and Werts 1988; Moran et al., 1992). Sequence and length variability within the NTS is commonly observed both within and among species. In genomes containing more than one spacer length, the variant classes are generally restricted to discrete chromosomal sites, as in rye (Reddy and Appels 1989), in concurrence with the concerted evolution of tandem repeated DNA families.

ISH has been used to identify the chromosomal location of 5S rRNA genes in a number of plant species, including wheat, rye, barley, maize, pea, tomato, and rice (Appels et al., 1980; Leitch and Heslop-Harrison 1993; Mascia et al., 1981; Ellis et al., 1988; Lapitan et al., 1991; Song and Gustafson 1993). In general, one to three chromosomes carry the 5S
rDNA. In all higher eukaryotes examined to date, these genes are not associated with the
genes encoding the 18S-26S rRNA, although in wheat and rye the two arrays are situated near
one another (Appels et al., 1980).

Most analyses of plant 5S rRNA genes have involved angiosperm species to date.
Consequently, little is known of its organization and chromosome location in the genome of
gymnosperm plants. Cullis et al. (1988) and Moran et al. (1992) characterized the molecular
structure of the 5S rDNA in radiata pine while Doudrick et al. (1995) determined its
chromosomal locations in slash pine. The present study was initiated not only to develop
additional cytological markers for spruce chromosomes but to extend the knowledge of the 5S
rRNA gene family in gymnosperms. As a rapid alternative to isolating 5S rDNA on
actinomycin-D/CsCl density gradients (Lawrence and Appels 1986), the polymerase chain
reaction was employed to amplify the 5S rDNA from the genome of white spruce using
primers derived from known evolutionarily conserved sequences. The 5S rDNA was
sequenced and its organization and copy number in white spruce determined. Additionally,
fluorescence ISH was used to physically map the 5S rDNA in white and Sitka spruce.

MATERIALS AND METHODS

DNA isolation

Genomic DNA was isolated from white spruce needle samples collected from the
Kalamalka Research Station of the British Columbia Ministry of Forests, Vernon, B.C. Five
grams of needles were ground to a fine powder in liquid nitrogen using a mortar and pestle
and stirred into 75 mls of ice cold extraction buffer. The extraction buffer consisted of 50
mM Tris-HCl, pH 8, 5 mM EDTA, 0.35 M sorbitol, 0.1% BSA, 10% PEG 4000, 0.1% spermine and 0.1% spermidine. 2-mercaptoethanol was added to 0.1% just prior to use. The homogenate was filtered through four layers of cheesecloth and one layer of miracloth. Nonfiltered plant material was re-extracted with 25 mls ice cold extraction buffer and the two filtrates combined and centrifuged at 9000 xg for 15 minutes at 4 °C. The resulting pellet was resuspended gently in 5 mls of wash buffer (50 mM Tris-HCl, 25 mM EDTA, 0.35 M sorbitol and 0.1% 2-mercaptoethanol), transferred to a 15 ml Falcon tube, and 1/4 volume of 5% sarkosyl added. The solution was gently mixed and left at room temperature for 20 minutes. For each ml of solution, one gram of CsCl was then added. This solution was then transferred to an Oakridge tube containing 300 µl of 10 mg/ml ethidium bromide and ultracentrifuged at 45,000 xg for 16 hours at 20 °C. The band of genomic DNA was removed and placed in a quick-seal tube which was then filled with 1 g/ml CsCl and ultracentrifuged for a second time. Genomic DNA was again removed and the ethidium bromide extracted with water saturated n-butanol. Finally, the DNA was ethanol precipitated, washed twice with 70% ethanol, vacuum dried and resuspended in 50-150 µl of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8).

**PCR amplification and cloning of the 5S rRNA genes**

PCR amplification (Saiki *et al.*, 1985; Mullis *et al.*, 1986) was carried out in a Perkin-Elmer Cetus DNA Thermal Cycler. The reaction mixture consisted of 5 ng white spruce template DNA, 10 mM Tris-Cl, pH 8.3, 50 mM KCl, 0.2 mM dNTPs, 1 µM primer pair, 2 mM MgCl₂ and 1.25 U Amplitaq DNA Polymerase in a 50 µl volume. The primer pairs used were synthesized based on the sequences of seven angiosperms reported in Goldsborough *et al.*, (1982) and included (P1: 5'-GGGTGCGATCATACCAGCGT-3' and P2: 5'-
GGTGCAACACTAGGACTTC-3') and (P3 5'-GAGTTCTGATGGGATCCGGTG-3' and 
P4 5'-CGCTTGGGGCTAGAGCAGTAC-3'). Reactions were initially denatured at 94 °C for 3 
minutes and subsequently subjected to 20 cycles of 94 °C for 1 minute and 55 °C for 10 
seconds. Amplifications were completed by a 72 °C final extension for 10 minutes. Reaction 
products were resolved on 2% agarose gels in 1X TAE and PCR products to be cloned were 
excised and gel purified. Following reamplification under the identical cycling conditions, the 
reaction mixture was cloned into the EcoRV site of pBluescript KS+ using the dideoxy-tailing 
method described by Holton and Graham (1991). Approximately 10 ng of the ligation 
mixture was used to transform competent DH5 alpha E. coli cells (BRL).

Genomic DNA digestion and Southern hybridization

Genomic DNA (1 μg) was partially digested with 10 units of BamHI or ScaI at 37 °C 
for varying amounts of time (from 0-120 minutes). Reactions were stopped by adding EDTA 
to 25 mM and the restriction fragments resolved in 1.6% agarose gels in 1X TAE. Following 
a 5 minute depurination in 0.25 M HCl, DNA was blotted to Hybond N+ (Amersham) 
membranes by capillary action using 0.4 M NaOH as the transfer buffer. The insert from 
pWS11, a PCR product from amplification with primers P3 and P4, was excised from 
pBluescript by EcoRI/XhoI digestion, gel purified and labeled with 32P-dCTP by random 
priming using a commercially available kit (Boehringer-Mannheim). Following a 1 hour 
prehybridization at 65 °C in 6X SSC, 20 μg/ml sheared, denatured salmon sperm DNA, 5X 
Denhardt's and 0.5% SDS, the labeled probe was denatured at 100 °C for 5 minutes, chilled 
on ice, and incubated with the membrane at 65 °C overnight. (1X Denhardt's is 0.02% Ficoll, 
0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin). Membranes were subsequently
washed in 2X SSC for 15 minutes, 0.2X SSC at 65 °C for 30 minutes and twice in 0.1X SSC at 65 °C for 30 minutes before autoradiography.

**DNA sequencing**

Sequencing of both strands of each PCR clone was performed by the dideoxy chain termination method (Sanger *et al.*, 1977) with the T7 Sequencing Kit (Pharmacia) using the T3 and T7 promoter primers.

**Estimation of 5S rDNA copy number in white spruce**

The copy number of the 5S rDNA in white spruce was estimated by reconstruction experiments. Various dilutions of pWS11, corresponding to the expected weight of 100, 500, 1000, 2000, 5000, 7500, 10,000 and 20,000 copies, were prepared in sterile distilled water. NaOH and EDTA were added to to final concentrations of 0.4 M and 0.01 M, respectively. The DNA was denatured at 80 °C for 10 minutes, chilled on ice, and immobilized on a nylon membrane using a slot blotting apparatus (BIO-RAD). Equivalent molar amounts of pBluescript were slotted as controls for cross-hybridization to vector sequences. Aliquots of white spruce genomic DNA (0.5 and 1.0 μg) were also applied. Two replicates of each filter set were included. After probing with the insert of pWS11, membranes were washed at high stringency (as above) and the resulting autoradiograms analyzed by densitometry.

**Chromosome preparations and in situ hybridization**

Plant materials and procedures performed were as described in Chapter 2 (pp. 33-37). Fifteen well spread metaphases from both white spruce and Sitka spruce were analysed.
RESULTS

Characterization of 5S rDNA in white spruce

The 5S rDNA was amplified from white spruce genomic DNA by the polymerase chain reaction. Primers P1 and P2 were designed to amplify the 120 bp coding region only and primers P3 and P4 to amplify the entire repeating unit (with the exception of a 12 bp region between the 5' ends of P3 and P4, Fig. 3.1). PCR amplification with P1 and P2 gave the expected 120 bp product. Primers P3 and P4 yielded a prominent amplification product of approximately 210 bp and lesser products of approximately 430 and 650 bp (data not shown). Reamplification of either the 430 or 650 bp product produced the 210 bp product, consistent with these minor PCR products being dimers and trimers, respectively, of a 5S rDNA monomer.

To obtain the sequence of the repeating unit, both the 120 and 210 bp PCR products were cloned and four representatives of each analyzed. Alignment of the sequences from each primer pair revealed that the amplification products represented a 221 bp 5S rRNA gene consisting of the 120 bp coding region, as determined by sequence comparison with other plant 5S rRNA genes, and a 101 bp NTS. Sequence variation among clones was observed corresponding on average to less than one nucleotide substitution for coding region sequences and seven nucleotide substitutions for NTS sequences.

A consensus sequence derived from the eight PCR products is presented in Fig. 3.2. The coding region shows complete sequence identity with the 5S rRNA of Scots pine (Mashkova et al., 1990), and the 5S rDNA from radiata pine (Moran et al., 1992) and Douglas-fir (Amarasinghe and Carlson, unpublished). Upstream of the transcription start site,
Fig. 3.1. Origin of primer pairs P1/P2 and P3/P4 used in the PCR amplification of 5S rDNA. The sequences of primer pairs P1/P2 and P3/P4 are given in the Materials and Methods. Line, nontranscribed spacer. Open box, coding region. Shaded region, 12 bp of coding region sequence not amplified using primers P3/P4.
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGGTGGCATC</td>
<td>ATACCACCGT</td>
<td>TAATGCACCG</td>
<td>GATCCCATCA</td>
</tr>
<tr>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>GAACCTCCGCA</td>
<td>GTTAAGCGCG</td>
<td>CTTGGGCTAG</td>
<td>AGTAGTACTG</td>
</tr>
<tr>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
</tr>
<tr>
<td>GGATGGGTGA</td>
<td>CCTCCCGGGA</td>
<td>AGTCCTAGT</td>
<td>TTGCACCCCTT</td>
</tr>
<tr>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
</tr>
<tr>
<td>CCCCCCTTTT</td>
<td>GCATGGCTCC</td>
<td>GCGATGGATC</td>
<td>GGGCGCTTTT</td>
</tr>
<tr>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td>TAAGCCTCTCC</td>
<td>CCGCCGGCTC</td>
<td>GGTGATGTCC</td>
<td>ATTCAAGAGG</td>
</tr>
<tr>
<td>210</td>
<td>220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGGAGGGGGGC</td>
<td>CTGATCCTTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3.2. Consensus sequence of the 5S rRNA genes in white spruce. Nucleotides in bold represent the coding region. The nontranscribed spacer begins at position 121. Regulatory elements referred to in the text (dotted lines) are found in the 3' region of the nontranscribed spacer. The BamHI (single line) and ScaI (double line) restriction sites are noted.
Fig. 3.3. Southern hybridization of pWS11, a P3/P4 primer pair amplification product. Genomic DNA of white spruce was digested for varying amounts of time with *BamH1* or *ScaI*. Arrows, hybridizing fragments not corresponding to multimers of the 221 bp repeating unit. Molecular weights shown are in kilobases.
Fig. 3.4. Estimation of the copy number of 5S rDNA. Dilutions of white spruce genomic DNA, pWS11, and pBluescript KS+ corresponding to the expected number of copies shown on the left were denatured, slotted onto nylon membranes and hybridized at high stringency with the $^{32}$P labeled insert of pWS11. Hybridization to the plasmid vector pBluescript KS+ was not detected. Copy number calculations were based on densitometric readings and the assumption that the 1C DNA content of white spruce is $8.5 \times 10^9$ bp.
Fig. 3.5. Localization of the 5S rDNA (green) in white spruce. a) Fluorescence ISH of biotin-labeled pWS11 (green). Chromosomal DNA was counterstained with propidium iodide (red). b) The same metaphase stained with DAPI. *Arrowhead*, secondary constriction adjacent to 5S rDNA site. *Asterisk*, B chromosome. Bar represents 10 μm.
Fig. 3.6. Localization of the 5S rDNA (green) in Sitka spruce. a) Fluorescence ISH of biotin-labeled pWS11 (green). Chromosomal DNA was counterstained with propidium iodide (red). b) The same metaphase stained with DAPI. Arrowhead, secondary constriction adjacent to 5S rDNA site. Asterisk, B chromosome. Bar represents 10 μm.
from nucleotide -1 to -33, the 5S rDNA sequences of radiata pine and Douglas-fir also show 80% identity to that of white spruce. Within this region are elements suggested to regulate transcription (Venkateswarlu et al., 1991), including a cytosine at -1, a GC-rich region between -11 and -17, and an AT-rich region between -24 and -30. Lastly, a stretch of four thymidines immediately downstream from the 3' end of the coding region likely represents the signal for transcription termination.

The genes encoding the 5S rRNA molecule are typically arranged in a genome as long tandem arrays. To verify this organization in white spruce, Southern blots of genomic DNA partially digested with either BamHI or ScaI, which have only one restriction site within the repeating unit, were probed with pWS11, a clone with a white spruce insert from the P3 and P4 primer pair amplification (Fig. 3.3). The rationale for this experiment is that partial digestion will randomly restrict only a limited number of all available sites. Therefore, for a tandemly repeated DNA, a group of restriction fragments based on multimers of the repeating unit will be generated which, when resolved in an agarose gel and analysed by Southern hybridization, appears as a "ladder" of hybridizing fragments. When either partial BamHI or ScaI digests were probed with pWS11, a ladder of bands based on an approximately 220 bp monomer was apparent. With increasing digestion time, fragments of approximately 600, 1000 and 1200 bp which do not correspond exactly to multiples of the repeating unit were apparent. These latter fragments, more evident in the completed ScaI digests (which is not methylation sensitive) shown in lanes 5-10 in Fig. 3.3, likely represent minor length variants of the 5S rDNA genes in the white spruce genome (with the 1200 bp fragment possibly being a dimer of the 600 bp fragment). Although various PCR reaction and cycling conditions were tested, amplification products corresponding to these length variants were not seen.
The copy number of the 5S rDNA genes in white spruce was calculated from densitometry readings of slot blots probed with pWS11 (Fig. 3.4). A value of approximately 1170 +/- 82 copies per haploid genome was derived. Given that chloroplast and mitochondrial DNA sequences can account for a considerable portion of total DNA, this value is likely an underestimate of the copy number in white spruce.

*Physical mapping of the 5S rDNA*

Using fluorescence ISH, the 5S rRNA genes were assigned to a single metacentric chromosome pair in white spruce very near to a secondary constriction (Fig. 3.5). There was very little difference in the length of the long and short arm of this chromosome and, therefore, fractional lengths were calculated from the end of the arm to which the probe hybridized. From 10 chromosomes, the FLpter for the 5S rDNA and the NOR were estimated as 0.28 +/- 0.02 and 0.23 +/- 0.01, respectively. The average relative length of the hybridizing chromosome pair indicates that it should be designated chromosome 5.

Different size classes of 5S rDNAs are often found sequestered at discrete chromosomal sites (Dvorak *et al.*, 1989; Reddy and Appels 1989). However, even with amplification of the ISH signal using overlays of biotin-labeled anti-avidin antibodies and FITC conjugated ExtrAvidin, no additional hybridization sites were detected (data not shown). These results suggest that all 5S rDNA size classes in white spruce are found at this one chromosome site, either in adjacent tandem arrays or with some degree of interspersion.

A single chromosomal site for the 5S rDNA was observed in all metaphases of Sitka spruce as well (Fig. 3.6), on what appears to be the homeologous chromosome. Fractional lengths calculated from the end of the chromosome arm to which the 5S rDNA and pGmR1
hybridized were determined to be 0.28 +/- 0.02 and 0.22 +/- 0.01, respectively.

DISCUSSION

Characterization of the 5S rDNA amplification products by DNA sequencing, Southern hybridization and fluorescence ISH have shown that the white spruce 5S rDNA is typical of that found in higher plants. The primary form of 5S rDNA observed was a 221 bp repeating unit composed of the 120 bp coding sequence and a 101 bp nontranscribed spacer. Among the clones analysed, sequence variation observed within the NTS was 7-fold greater than in the coding region and likely reflects natural variation rather than misincorporation by Taq DNA polymerase. A comparison of the 5S rDNA sequences among several genera in the Pinaceae family showed a similar degree of conservation as that observed among families of angiosperm plants, both in the coding region and in sequences immediately upstream.

The 120 bp coding sequence was arranged into the generalized secondary structure model proposed for plant 5S rRNAs (Barciszewska et al., 1994). An irregularity in the double helical region of Stem II was noted where a thymidine at nucleotide 20 must base pair with a guanine at nucleotide 58. In a recent review of 5S rRNA sequences at the RNA and DNA levels, Barciszewska et al. (1994) noted that sequence data from 5S rDNA are often not colinear with the sequence of the mature 5S rRNA molecule. Possible explanations include the sequencing of non-transcribed pseudogenes or the potential existence of a mechanism to edit the 5S rRNA. Although in this study a number of PCR products were sequenced to minimize errors inherent in PCR, it may still be that sequence comparisons among organisms may be suspect unless performed at the RNA level. With this in mind, it is interesting to note
that the irregularity in Stem II base pairing found in white spruce is not found in those angiosperms analyzed to date, but has been reported in 6 of 7 gymnosperms studied, namely Gingko, Cycas, Ephedra, and Scots pine at the RNA level (see Barciszewskia et al., 1994 for references) and in radiata pine and Douglas-fir at the DNA level (Culliscia et al., 1988; Moran et al., 1992; Amarasinghe and Carlson, unpublished).

The tandem organization of the 5S rDNA genes in white spruce was confirmed by Southern hybridization. The presence of additional size classes of hybridizing fragments suggests that spacer length variation exists but that the PCR conditions used were ineffective in amplifying all size classes. While PCR amplification may be a suitable means of generating homologous probes for ISH, established methods of cloning rDNAs may be more appropriate for the complete characterization of these gene families.

One major 5S rDNA locus was identified on the metaphase chromosomes of both white spruce and Sitka spruce, adjacent to a secondary constriction. Hybridizing chromosomes also showed morphological similarities between the two species, suggesting their homeologous nature. Assuming that all 5S rDNA repeats are found at this site, the array is approximately 260 kb in length. While minor sites were not detected even with amplification of the hybridization signal, it is difficult to rule out their existence given the Southern hybridization evidence for length variants in white spruce and the preferred clustering of such variants at discrete loci characteristic of most plants analysed. As in rye (Reddy and Appels 1989), minor 5S rDNA sites may be revealed as the sensitivity of ISH improves.

In most plants and animals, the 5S rDNA repeats outnumber the 18S-26S rDNA genes (Appels et al., 1980). Although the copy number of the 18S-26S rDNA in white spruce has
not been reported, more than 12,000 genes are present per haploid genome in Sitka spruce (Ingle 1975). This is more than 10-fold higher than the 1170 copies of the 5S rDNA estimated here in white spruce. Additionally, the chromosomal distribution of the 5S rDNA arrays has not paralleled the multilocus nature of the 18S-26S rDNA in these spruces.

The primary goal in developing the ISH procedure at this stage was to establish a means of identifying spruce chromosomes quickly and without reliance on chromosome measurements. In the exceptional metaphase spread in which squashing had not appreciably distorted any of the chromosomes, 10 of the 12 chromosome pairs could be distinguished by morphology and the location of the 18S-26S rDNA (revealed by either DAPI staining or ISH using pGmR1). Length measurements were needed to discern chromosomes 2 and 5 in Chapter 2. The 5S rDNA locus in both white spruce and Sitka spruce mapped to chromosome 5 which allowed all the chromosomes to be reliably identified.

In both species, the 5S rDNA bearing chromosomes are conserved in size, morphology and the location of two repeated DNA families to date. The single 5S rDNA locus could serve as the link to integrate the physical chromosome with genetically mapped molecular markers through segregation analysis of polymorphic restriction fragments within the array. Pulsed field gel electrophoresis could also be used in the genetic mapping of very large fragments or the entire 5S rDNA array assuming that parents heterozygous for the total array length could be identified. Such a study would provide the first look at the synteny of a specific chromosome between conifer species and reveal insights into chromosome evolution within the genus.
LITERATURE CITED


Characterization of a Centromeric Satellite DNA in White Spruce

INTRODUCTION

A significant portion of genomic DNA from many plant and animal species can be separated from the main band of DNA using buoyant density ultracentrifugation. This "satellite" DNA is composed of tandem arrays of one or more highly repeated DNA family. Other tandem repeated sequences, described as cryptic, have densities similar to the bulk of genomic DNA and consequently band at the same position. Many distinct satellites and cryptic satellites (herein referred to as satellites) of differing complexities can exist in a plant's genome. These include the small tandem repeated microsatellites and minisatellites, and those with larger repeating units of typically 150-500 bp (Lapitan 1992). Satellite DNAs are typically the most highly repeated DNA sequences in plant genomes. For example, in Arabidopsis, there are 6000 copies of a 185 bp repeat comprising 1.5% of the genome (Simoens et al., 1988) and, in rye, 10% of the genome is made up of four satellite sequences located near the telomeres of all chromosomes (Bedbrook et al., 1980).

While several tandem repeated DNA families are conserved across a broad evolutionary spectrum and have a known or hypothetical function, the majority of tandem repeats are a dynamic, rapidly changing component of genomes. The distribution of a particular sequence within a taxonomic family may range from its presence in many or all
species to its confinement to a single genus or species. Many tandem repeated DNAs have been classified as "junk" DNA as a reflection of their rapid divergence, high degree of methylation and lack of protein coding capacity. Consistent with this apparent genetic inertness, satellite DNA is generally confined to constitutively heterochromatic regions of chromosomes. These areas, denoted by Giemsa C-bands, include subtelomeric regions and centromeres as well as some interstitial sites of some or all chromosomes of a complement. The correlation of C-bands with the location of characterized satellite DNAs has been observed by ISH in a number of plants (and animals), including rye (Mukai et al., 1992) and onion (Irifune et al., 1995).

Changes in satellite DNA may represent the primary form of variation in genome size among related species (Flavell et al., 1977). As such, they have been used to investigate the evolution of species and chromosomes and to support established taxonomic relations (Lapitan et al., 1987; Harrison and Heslop-Harrison 1995). Since they are a major component of the nuclear genome, studies on genome organization are lacking without knowledge of the types, numbers and distribution of satellite DNAs. To date, however, satellite DNA in conifers has not been investigated or exploited. This chapter details the isolation and characterization of a white spruce satellite DNA, SGR-31 (Spruce Genomic Repeat), and investigates its chromosomal location and presence in the genus.

MATERIALS AND METHODS

Plant material and DNA isolation

Plant material used in Southern hybridizations is listed in Table I. Spruce samples
were collected from a spruce arboretum at the Kalmalka Research Station and provided by Gyula Kiss. Douglas-fir and western redcedar samples were collected from trees on the University of British Columbia campus. Radiata pine DNA was provided by a post-doctoral fellow in Dr. Carlson’s Lab, Dr. Yong-Pyo Hong. Genomic DNA was isolated from needle samples as described on pp. 52-53 of Chapter 3. Plant materials used for cytogenetic analysis are described on pp. 33 of Chapter 2.

**Screening of a partial white spruce library**

Cloning and preliminary screening were performed by Dr. Craig H. Newton of B.C. Research, Inc. as part of an unrelated research project. Genomic DNA was fractionated by CsCl/Hoechst 33258 equilibrium ultracentrifugation (Douglas 1988). The gradient fraction from which ribosomal DNA had been mostly subtracted was identified, digested to completion with Sau3A according to the manufacturer's recommendations and ligated into the BamHI site of dephosphorylated pUC8. Approximately 10 ng of the ligation mixture was used to transform competent DH5 alpha *E. coli* cells (BRL). Two thousand white colonies were screened for those containing repeated DNA inserts by colony hybridization (Sambrook *et al.*, 1989). White spruce genomic DNA was labeled with $^{32}$P-dCTP by random priming using a commercially available kit (Boehringer-Mannheim) and hybridized to colony lifts under the conditions described on pp. 53 of Chapter 3. Twenty five colonies representing a range of signal intensities were picked and supplied to the author for further analysis.

Selected clones were screened by Southern hybridization. Genomic DNA (2 $\mu$g) was digested with 4-10 units of *DraI*, *HaeIII*, *HindIII*, *HinfI*, *MvaI*, *RsaI*, *Sau3A* or *TaqI* for 16 hours at 37 °C and resolved in 1% agarose gels using 1X TAE as the running buffer.
Table I  Picea species included in Southern hybridization experiments with SGR-31.

<table>
<thead>
<tr>
<th>Species</th>
<th>Species Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Picea glauca</em></td>
<td>Canada, N. United States</td>
</tr>
<tr>
<td><em>Picea sitchensis</em></td>
<td>W. Canada, W. United States</td>
</tr>
<tr>
<td><em>Picea pungens</em></td>
<td>W. United States</td>
</tr>
<tr>
<td><em>Picea mariana</em></td>
<td>Canada, N. United States</td>
</tr>
<tr>
<td><em>Picea rubens</em></td>
<td>E. Canada, N.E. United States</td>
</tr>
<tr>
<td><em>Picea chihuahuana</em></td>
<td>Mexico</td>
</tr>
<tr>
<td><em>Picea mexicana</em></td>
<td>Mexico</td>
</tr>
<tr>
<td><em>Picea jezoensis</em></td>
<td>E. Asia</td>
</tr>
<tr>
<td><em>Picea polita</em></td>
<td>Japan</td>
</tr>
<tr>
<td><em>Picea koyamai</em></td>
<td>E. China, Japan</td>
</tr>
<tr>
<td><em>Picea asperata</em></td>
<td>central China</td>
</tr>
<tr>
<td><em>Picea purpurea</em></td>
<td>central China</td>
</tr>
<tr>
<td><em>Picea montigena</em></td>
<td>central China</td>
</tr>
<tr>
<td><em>Picea obovata</em></td>
<td>Scandinavia, N. Europe, Asia</td>
</tr>
<tr>
<td><em>Picea schrenkiana</em></td>
<td>E. China</td>
</tr>
<tr>
<td><em>Picea omorika</em></td>
<td>Serbia</td>
</tr>
<tr>
<td><em>Picea orientalis</em></td>
<td>S. Russia, N. Turkey</td>
</tr>
<tr>
<td><em>Picea abies</em></td>
<td>W. Europe, Scandinavia, W. Russia</td>
</tr>
</tbody>
</table>
Southern blotting was performed as described on pp. 54. The insert from each clone was excised from the plasmid by double digestion with EcoRI and HindIII, gel purified in 0.8% low melting agarose in 1X TAE, and labeled with $^{32}$P-dCTP by random priming. The hybridization and post-hybridization wash schedule were carried out as described on pp. 54-55 of Chapter 3.

For the partial digestion of genomic DNA, 1 µg aliquots were digested with 4 units of Sau3A at 37 °C for varying amounts of time. Reactions were stopped by adding EDTA to 25 mM. Genomic DNA (1 µg) from the spruce and other Pinaceae species listed in Table I was digested overnight with 4 units of Sau3A at 37 °C. These digests were resolved on 1.6% agarose in 1X TAE. Southern blotting and hybridizations were performed as described above.

**DNA sequencing**

Prior to sequencing, the EcoRI/HindII fragment, representing the white spruce insert and the multiple cloning site of pUC8, was subcloned into pT7T3 (Pharmacia) to make use of this vector's T3 and T7 bacteriophage promotor sites. Sequencing of both strands of SGR-31 was performed by the dideoxy chain termination method (Sanger et al., 1977) with the T7 Sequencing Kit (Pharmacia).

**Estimation of SGR-31 copy number in white spruce**

The copy number of SGR-31 in white spruce was estimated by reconstruction experiments. Various amounts of SGR-31, corresponding to the expected weight of 100, 500, 1000, 2000, 5000, 7500, 10,000 and 20,000 copies were applied to a nylon membrane using a slot blotting apparatus (BIO-RAD). Equivalent molar amounts of pT7T3 were slotted
Fig. 4.1 Southern hybridization of $^3$P-labeled SGR-31. One microgram aliquots of white spruce genomic DNA were digested for 1 hour with varying amounts of Sau3A (indicated at the top of each lane). Molecular weights are in base pairs.
Fig. 4.2. Nucleotide sequence of SGR-31. The Sau3A sites defining the repeating unit are underlined.
Fig. 4.3. Southern hybridization of SGR-31 to Sau3A digested genomic DNA from 18 species of *Picea*. Douglas-fir (*Pseudotsuga menziesii*), western red cedar (*Thuja plicata*) and radiata pine (*Pinus radiata*) were included to represent three other conifer genera. Molecular weights given are kilobases.
Fig. 4.4. Localization of SGR-31 in a) white spruce and b) Sitka spruce by fluorescence ISH. SGR-31 (green) hybridized to the centromeres of four chromosome pairs in a) and five in b). Chromosomal DNA was counterstained with propidium iodide (red). Several interspecific differences SGR-31’s chromosomal distribution were apparent including one observed on chromosome 12 (arrowheads). Bar represents 10 μm.
separately and served to control for cross-hybridization to vector sequences. Aliquots of white spruce genomic DNA (0.5 and 1.0 μg) were also applied. After hybridizing with the 32P-labeled insert from SGR-31, membranes were washed at high stringency and the resulting autoradiograph analyzed by densitometry. Two replicates for each clone were included.

**Chromosome preparation and in situ hybridization**

The procedures performed were as described on pp. 33-37 of Chapter 2.

**RESULTS**

**Characterization of SGR-31 in white spruce**

The monomer unit of satellite DNAs in plants (and animals) with small genomes can generally be observed and cloned directly from ethidium bromide stained gels of genomic DNA digested with the appropriate restriction enzyme. In contrast, the much higher number of restriction fragments generated by digestion of large genomes, like those of conifers, prevents the direct observation of satellite DNA monomers and necessitates the cloning of restriction digests and screening of recombinants.

Clones containing repeated DNA sequences were isolated from a partial white spruce library by comparing relative signal intensities when probed with total genomic DNA. To investigate the genome organization of selected clones, each was hybridized to Southern blots of white spruce DNA digested with the restriction enzymes *DraI*, *HaeIII*, *HindIII*, *HinfI*, *MvaI*, *RsaI*, *Sau3A* and *TaqI* (data not shown). Of the 25 clones screened, 3 clones revealed a prominent ladder of hybridizing fragments in *Sau3A* digests indicative of a tandem arrayed
repeated sequence. Dot blot hybridizations showed that these clones represented the same sequence family (data not shown) and, therefore, one clone designated SGR-31 was selected for further analysis.

To confirm its genomic organization, SGR-31 was hybridized to Southern blots of white spruce DNA partially digested with Sau3A (Fig. 4.1). These results clearly indicated the presence of an approximately 140 bp repeated DNA element tandemly arrayed in the genome. The size of the monomer corresponded to the estimated insert size of SGR-31 indicating that the entire repeating unit had been cloned. Faint hybridization to approximately 0.8 and 2 kb fragments is also evident suggesting a different organization of some SGR-31 related sequences in the genome.

The nucleotide sequence of SGR-31 was determined to be 138 bp in length with an A+T content of 60% (Fig. 4.2). No significant internal subrepeats were observed although small (6 bp or less) direct and inverted repeats were found. A search of the EMBL and GENBANK databases revealed no sequences with greater than 57% homology to SGR-31. Approximately 10,000 copies of this repeated DNA family are found per haploid genome of white spruce as calculated from densitometric analysis of slot blots (data not shown). This corresponds to 0.02% of the genome based on a haploid genome size of 8.5 x 10⁹ bp (Dhillon 1987) and is likely to be an underestimate given that considerable amounts of the total DNA are chloroplast or mitochondrial sequences.

SGR-31 in Picea

Seventeen additional species of spruce were assessed for the presence and genomic organization of SGR-31 by Southern hybridization to Sau3A digested genomic DNA.
Representative species from three other genera of Pinaceae were also included on the filters. Fragments homologous to SGR-31 were observed in all spruces as tandem repeats with the exception of Chihuahua spruce (P. chihuahuana) (Fig. 4.3). The size of the monomer unit was identical to that of white spruce in all species to the limit of gel resolution. In Chihuahua spruce, faint hybridization was seen to only the higher molecular weight fragments of 0.8 and 2 kb common to all spruce species examined. Differences in SGR-31 copy number among the spruces is suggested by relative hybridization intensities between lanes although this observation should be treated cautiously since the genome size of most of the species, and hence the number of genome copies per lane, has not been determined. SGR-31 appears to be specific to spruce species since no hybridization to the DNA of radiata pine, Douglas-fir or western redcedar was observed.

**Physical mapping of SGR-31**

The SGR-31 tandem repeated sequence was found within the primary constriction of four of the twelve chromosome pairs in white spruce by fluorescence ISH (Fig. 4.4a). Hybridization sites appeared to lie within the primary constriction and not in paracentromeric regions of the metaphase chromosomes examined although the condensed state of the chromosomes limited a higher resolution analysis. SGR-31 is not evenly distributed among these chromosome pairs as is evident by variation in signal intensity consistently observed. The highest copy number is found on chromosome 12, easily identified by its size and submedian centromere position, where the hybridization signal was usually seen as a broad band encompassing the entire region of the primary constriction. The other three chromosome pairs, identified as chromosomes 4, 6 and 8, showed much smaller, discrete
signals on each chromatid.

In Sitka spruce, SGR-31 hybridization sites were fairly uniformly distributed over the primary constriction of five of the twelve chromosome pairs including chromosomes 1, 4, 6, 7, and 8. (Fig. 4.4b). Distinct differences in the distribution of SGR-31 were apparent between homeologous chromosomes of the two species. The SGR-31 sites on chromosome 1 and 7 of Sitka spruce are absent in white spruce and the prominent SGR-31 site on chromosome 12 in white spruce is not found in Sitka spruce.

DISCUSSION

SGR-31, a centromeric satellite DNA

SGR-31, a 138 bp satellite DNA, was isolated from a white spruce genomic library. Fluorescence ISH analysis localized SGR-31 sequences to the centromeres of four white spruce chromosome pairs and five Sitka spruce pairs. Hybridization sites clearly fell within the boundaries of the primary constriction. Despite the abundance of repeated sequences characterized from a wide variety of plant genomes, only in Arabidopsis, barley, wheat, and several Brassica species have centromeric tandem arrays been isolated (Murata et al., 1994; Dennis et al., 1979; Xia et al., 1993; Harrison and Heslop-Harrison 1995). No significant homology between these centromeric repeats and SGR-31 was found.

SGR-31 is typical of most characterized satellite DNAs in that its simple sequence offers little insight into its origin and role, if any, in the spruce genome. Benslimane et al. (1986) maintain that some satellite DNAs arise from a tRNA gene ancestor although no homology of SGR-31 with any of the tRNA genes was observed. It has also been suggested
that plant satellite DNAs originate from a small 30 bp unit, first duplicated or triplicated, and then amplified as an entire block (Ingham et al., 1993). No internal subrepeats of greater than 6 bp were observed to support this argument. A potential role for SGR-31 in the spruce genome is no less clear. The folding of centromeric DNA into constitutive heterochromatin has been suggested by Vogt (1992) to be mediated by a protein binding capacity of tandem repeated structures involving a particular sequence domain and/or a stable curvature of the DNA. The sequence unit (GGAAT)\textsubscript{n} has been implicated as a component of this interaction based on its extreme evolutionary conservation (from yeast to sea urchins, maize, chickens and humans), centromeric location, unusual hydrogen bonding properties, high affinity for specific nuclear proteins and similarities to the functional centromeres isolated from yeast (Grady et al., 1992). However, no such sequence motifs are found in the primary sequence of SGR-31.

The most intriguing feature of SGR-31 is its confinement to only a subset of spruce chromosomes. Other centromeric satellite sequences isolated from plants, with the exception of pBcKB4 and pBoKB1 from Chinese cabbage and kale, respectively (Harrison and Heslop-Harrison 1995), were physically mapped to all somatic chromosomes of the particular species. While SGR-31 sequences may be completely absent from non-hybridizing centromeres, it is conceivable that SGR-31 copy numbers on these chromosomes present target sequences below the sensitivity of the ISH technique used. Alternatively, non-hybridizing chromosomes may contain SGR-31 related sequences that have diverged sufficiently to preclude hybridization under high stringency conditions. Low stringency hybridization of pBcKB4 and pBoKB1 detected additional centromeric locations with divergent sequences. In the chromosomes of humans and other primates, most centromeres have one or more diverged
alpha satellite subsets specific to that particular chromosome type (Willard 1990). Low stringency ISH might reveal that diverged SGR-31 related sequences are organized in a similar chromosome-specific manner as pBcKB4 and pBoKB1 and the alpha satellite.

_Evolution of SGR-31 in Picea_

The evolution of SGR-31 was investigated by Southern hybridization to 18 species of spruce and by comparative ISH to the metaphase chromosomes of Sitka spruce. The taxonomic relationship of spruce species is not fully elucidated because of a lack of fossil records for many species, relatively few changes to the basic spruce form having occurred during differentiation into the present day array of species, and geographic, as opposed to genetic, isolation mechanisms having been predominantly involved in speciation (Wright 1955). It was hoped that SGR-31 might contribute to current taxonomic views, however, Southern hybridizations provided no additional information. Both the size and organization of SGR-31 sequences are well conserved within all but one of the species assayed. Since only 4 of the 138 bp of SGR-31 are sampled for sequence variation by digestion with Sau3A a more thorough analysis involving sampling numerous restriction enzymes or DNA sequencing of SGR-31 in other species may contribute more to a taxonomic study.

SGR-31 should be considered an ancestral tandem repeated DNA family amplified early in the evolution of the genus since it is present and organized as such in all but one species assayed. These included _P. koyamai_, considered to be the most primitive and generalized spruce, and Yeddo spruce (_P. jezoensis_), the probable link between the older Asiatic species and those found in North America (Wright 1955). In Chihuahua spruce, an isolated Mexican species with limited range, SGR-31 is not organized in tandem arrays. SGR-
31 related sequences do exist in the Chihuahua spruce genome as the 0.8 and 2 kb Sau3A fragments present in all spruces but the molecular nature of these has not been investigated.

In contrast to Southern hybridization results, comparative ISH mapping of SGR-31 revealed differences between its chromosomal distribution in white and Sitka spruce. At some point since the divergence of these species intragenomic movement of SGR-31 has resulted in the species-specific sites on chromosomes 1, 7 and 12. Experimental observations in plants and animals that tandem arrays of a given sequence family are usually found on many, if not all, chromosomes of a species (after presumably originating from an amplification event at a single site) suggest that satellite DNAs can be transposed between homologous and non-homologous chromosomes. Flavell (1986) suggests several mechanisms of transposition including a double crossover between homologous chromosomes or the excision and integration of DNA fragments into non-homologous chromosomes. Alternatively, some arrays may be prone to transposition due to the sequence itself.

By whatever means it has occurred, the deletion and transposition of SGR-31 has had little effect on chromosome pairing during meiosis since fertile hybrids are invariably produced by interspecific hybridization. It would be interesting to determine the chromosomal distribution of SGR-31 in other spruces, and in particular Yeddo spruce, perhaps leading to a better understanding of taxonomic relations within the genus and the dynamic nature of the spruce genome.

**SGR-31 and spruce centromere function**

The kinetochore is the trilaminar structure at the outer surface of the centromere to which microtubules attach in mediating the movement of chromosomes during mitosis and
meiosis. Its assembly must involve the interaction of specific centromeric DNA sequences with kinetochore components (Willard 1990). It has been postulated from studies on mammalian centromere organization that the process might rely on the regular spacing of a particular DNA curvature or sequence motif embedded in a larger repeated DNA "superstructure" (Vogt 1990; 1992). Two putative motifs have been characterized including the \((\text{GGAAT})_n\) repeat (Grady et al., 1992) described earlier and the CENP-B box, a 17 bp motif found in a subset of alpha satellite sequences of humans and the mouse minor satellite. CENP-B, a major protein component of human centromeres, binds the CENP-B box both \textit{in vitro} and \textit{in vivo} (Masumoto et al., 1989; Haaf et al., 1992) and has been immunolocalized to the central domain of human metaphase chromosomes underlying the kinetochore (Cooke et al., 1990). Vogt's concept is appealing since only the short motifs and their regular spacing are selected for as the critical requirement for kinetochore assembly. Within an array of tandemly repeated satellite DNA many copies of the sequence or curvature will be present. If the interaction with kinetochore proteins requires the correct spacing of only some of these motifs, then amplification, deletion and rearrangement of the surrounding repeated sequences (the spacing domain; Vogt 1992) can result in the chromosome-specific satellite DNA compositions characteristic of mammalian centromere organization.

Little is known about the specific sequence organization of plant centromeres other than the few repeated DNAs listed previously. While it appears that SGR-31 is not a required component of the spruce centromere, Vogt's postulate suggests that some so-called junk DNAs might have a cellular role despite chromosome specificity and variation in copy number between chromosomes. Whether SGR-31 plays a structural role in centromere function or is simply tolerated in the spruce genome is presently unknown.
LITERATURE CITED


87


CHAPTER 5

The First Cytogenetic Maps of
White and Sitka Spruces

INTRODUCTION

Providing a location on the emerging ISH map of spruce for a new DNA clone necessitates that its hybridization site can be assigned to a recognized chromosome and related to other markers in its proximity. In mammalian cytogenetics, new markers can be quickly assigned not only to a chromosome but to a well defined region by G-banding after ISH (Bhatt et al., 1988; Lemieux et al., 1992). Although G-banding in plants is uncommon, Jiang and Gill (1993) successfully applied modified N- or C-banding protocols following ISH to identify wheat chromosomes hybridizing repeated DNAs. Similarly, ISH of the 18S-26S rDNA and the 5S rDNA in combination with fluorescent banding patterns produced by DAPI and Chromomycin A₃ distinguished most of the chromosomes of slash pine (Doudrick et al., 1995).

In spruce, C- or N- banding procedures have not been defined and fluorescent banding with DAPI and Chromomycin A₃ are less informative than in slash pine (Brown and Carlson, unpublished data; Hizume et al., 1991). However, the wide distribution of 18S-26S rRNA genes is sufficient to provide a chromosomal designation for the majority of probes. In most cases, all that it is needed is a DAPI stained image of the metaphase spread to determine to
which chromosome a probe has hybridized. In others, particularly involving chromosomes 2 and 5, additional information such as the chromosome-specific 5S rDNA locus is needed. In this case, multiple probes must be physically mapped on the same metaphase spread.

DNA clones can be co-localized on the same metaphase spread by either simultaneously hybridizing distinctly labeled probes (Leitch et al., 1991) or by multiple probings of the same slide (Heslop-Harrison et al., 1992) in a manner analogous to the repeated hybridizations afforded by nitrocellulose or nylon membranes in Southern analysis. However, wide differences in target sequence copy numbers can make analysis of simultaneously hybridized probes difficult and the extra detection reagents required can lead to unacceptably high levels of background signal. In this chapter, the method of Heslop-Harrison et al. (1992) was employed using slides previously hybridized with SGR-31 (Chapter 4). By stripping these of old reagents and reprobing them, first with the 5S rDNA and then the 18S-26S rDNA clones, all three repeated DNAs were co-localized on the same metaphase spreads. This approach should be sufficient to designate specific chromosomal locations for additional repeated and low copy DNA sequences in spruce.

MATERIALS AND METHODS

In situ hybridization

After acquiring SGR-31 hybridization images (Chapter 4), slides could be immediately stripped of old detection reagents, hybridized probe sequences and mounting medium or stored in the dark at 4 °C for at least six months. Stripping was accomplished by washing slides 3 times in 0.1% Tween-20 (v/v) in 4X SSC for 1 hour each and twice in 2X SSC for 5
minutes each. Slides were then dehydrated through an ethanol series and air dried. Biotin-labeled pWS11 (white spruce 5S rDNA clone) was then denatured at 80 °C, chilled on ice, and applied to the slide. Chromosomal DNA denaturation, ISH and detection were performed as described on pp. 33-36 of Chapter 2. This procedure was repeated a third time for the ISH of pGmR1 (18S-26S rDNA).

**Image analysis**

To generate the images in Fig. 5.1 of propidium iodide stained chromosomes hybridized with the three repeated DNA sequences, the pWS11 image was first merged with the propidium iodide image using Adobe Photoshop 3.0 software and converted to an 8-bit grey scale image. This image was then merged with the 8-bit images of SGR-31 and pGmR1 hybridization, producing a 24-bit RGB color image in which the images of pWS11/propidium iodide, SGR-31 and pGmR1 were arbitrarily assigned to the red, green and blue channels, respectively. The two hybridization sites of pWS11 were then each selected and pseudocolored to more clearly distinguish them from the counterstain.

**RESULTS**

The three repeated DNA sequences described in this research were co-localized on the same metaphase spreads in white spruce and Sitka spruce (Fig. 5.1). Chromosome morphology was not affected by the repeated stripping and hybridization. Of the twenty spreads observed after each hybridization only one suffered any chromosome loss. Although some amount of DNA loss is inevitable, these results suggests the potential for further rounds
Fig. 5.1. Fluorescence ISH colocalization of the 5S and 18S-26S rDNA, and SGR-31 on the mitotic chromosomes of a) white spruce and b) Sitka spruce (2n = 24). Fluorescent images of hybridization sites for the 18S-26S rDNA (purple), the 5S rDNA (pale blue) and SGR-31 (green) were merged with the propidium iodide counterstained image (red). The distinguishing features of each chromosome are listed in the text. Bar represents 10 μm.
Fig. 5.2 The first cytogenetic maps of white spruce and Sitka spruce. Map positions of the 18S-26S rDNA (small gaps) and the 5S rDNA (open circles) are given from the terminus of the short arm. SGR-31 sites at the centromeres are denoted by filled circles.
of hybridization with other repeated DNAs.

Chromosome pairs of white and Sitka spruce were readily identified by morphology or ISH pattern and are shown diagramatically in the ideograms in Fig. 5.2. Distinguishing features of each white spruce chromosome are described below:

**Chromosome 1** - Longest chromosome. No hybridization sites.

**Chromosome 2** - Metacentric. Distinguished from #3 by position of 18S-26S rDNA in the middle of long arm. The extended secondary constriction in one homolog frequently observed in this seedlot may account for different arm designations and ratios between white and Sitka spruces.

**Chromosome 3** - Metacentric. Easily distinguished from all other chromosomes by the 18S-26S rDNA locus proximal to the centromere on the long arm.

**Chromosome 4** - 18S-26S rDNA site on short arm and SGR-31 site at centromere.

**Chromosome 5** - Metacentric. 18S-26S rDNA and 5S rDNA sites on the same arm. This arm has been designated the short arm since arm length measurements were variable.

**Chromosome 6** - Easily distinguished from all chromosomes by the subtelomeric 18S-26S rDNA locus on the short arm. SGR-31 locus also at the centromere.

**Chromosome 7** - Metacentric. Smaller than #1. No hybridization sites of SGR-31, 5S rDNA or 18S-26S rDNA.

**Chromosome 8** - 18S-26S rDNA locus on the long arm and SGR-31 site at centromere.

**Chromosome 9** - Easily identified from all other chromosomes by centromere position and arm ratio. No hybridization sites of SGR-31, 5S rDNA or 18S-26S rDNA.

**Chromosome 10** - Smallest chromosome with 18S-26S rDNA site.

**Chromosome 11** - Smallest metacentric chromosome. No hybridization sites.

The unique features of each Sitka spruce chromosome are listed below, with differences between their probable white spruce homeologues noted. Chromosomes 2-6, 9 and 11 cannot be distinguished between white spruce and Sitka spruce based on morphology or ISH patterns of the repeated DNAs used.

Chromosome 1 - Longest chromosome. Centromeric site of SGR-31 in Sitka spruce which is absent on #1 in white spruce.

Chromosome 2 - Similar 18S-26S rDNA locus in middle of chromosome arm as on #2 of white spruce but chromosome measurements place it on the short arm in Sitka spruce. 18S-26S rDNA site polymorphism not observed.

Chromosome 3 - Similar proximal 18S-26S rDNA locus as on #3 of white spruce although chromosome measurements place the NOR in Sitka spruce on the long arm.

Chromosome 4 - Similar to #4 of white spruce.

Chromosome 5 - Similar to #5 of white spruce.

Chromosome 6 - Similar to #6 of white spruce.

Chromosome 7 - Similar morphologically to #7 of white spruce but with a centromeric SGR-31 site in Sitka spruce.

Chromosome 8 - Medium sized metacentric with a centromeric SGR-31 hybridization site. Although the arm ratio, centromeric index and relative length correspond to #8 of white spruce, no 18S-26S rDNA locus is found in Sitka spruce.

Chromosome 9 - Similar to #9 of white spruce. Easily identified by its size and arm ratio.
Chromosome 10 - Submetacentric. Third smallest in the complement. No hybridization sites. Lacks the 18S-26S rDNA loci on the long arm of #10 in white spruce.

Chromosome 11 - Second smallest chromosome. Distinguished from #10 and #12 in Sitka spruce by both length and arm ratio.

Chromosome 12 - Smallest chromosome of the complement. Lacks the prominent site of SGR-31 hybridization on #12 of white spruce.

DISCUSSION

Cytogenetic maps of white and Sitka spruce

The three repeated DNA families used in this research have allowed the placing of the first 12 loci onto the cytogenetic map of white spruce, seven corresponding to sites of the 18S-26S rDNA, one to the 5S rDNA, and four to SGR-31 loci. Eleven loci (five 18S-26S rDNA, one 5S rDNA and five SGR-31 loci) have likewise been mapped in Sitka spruce. While morphology and the locations of most repeated loci mapped here are common among the two species several clear differences are apparent. As mentioned previously, chromosomes 8 and 10 in Sitka spruce lack the 18S-26S rDNA loci found in white spruce. Chromosomes 1, 7 and 12 have different SGR-31 distributions in each species. Differences were also observed in the positioning of the 18S-26S rDNA on chromosomes 3 and 4.

Despite earlier cytologic observations that karyotype morphology among species of spruce is highly conserved, the differences in molecular cytogenetic maps between white and Sitka spruce indicate that the molecular forces responsible for repeated DNA turnover (e.g., deletion, amplification and transposition) are clearly active. The tools of genetic mapping,
DNA sequencing and ISH are now available to conifer geneticists to approach the question of whether or not karyotype conservation reflects gene order colinearity within and potentially between genera. Meiotic chromosome pairing studies in white X Sitka or other interspecific hybrids would also assist in studying chromosome synteny and genome rearrangements coincident with speciation in the genus.

Additional repeated or low copy sequences mapped by ISH can now be easily assigned to a spruce chromosome and in most cases the chromosome arm by reprobing the slide with repeated DNAs of known location. Currently the 5S and 18S-26S rDNA are satisfactory references and could be supplemented with new DNA sequences mapped in the future. Reprobing the same metaphase spread also ensures that the linear relation of sequences in proximity can be determined with certainty, in contrast to comparing hybridization sites between spreads. This is a very practical contribution to spruce cytogenetics since high quality root tip preparations and metaphase spreads, generally in limited supply, can be reused several times with minimal loss of chromosomal material.

*Repeated DNAs of the spruce genome*

While this thesis focussed on the ISH localization of tandem repeated DNA sequences, two lines of evidence support the notion that families of dispersed repeated DNAs may be the larger component of the spruce repeated DNA fraction and, therefore, be responsible for the large genome size. First, while developing ISH probes at the beginning of this research, a black spruce genomic library in a lambda bacteriophage vector (constructed and provided by Linda De Verno of Petawawa National Forestry Institute) was screened with white spruce genomic DNA to identify recombinants with highly repeated DNA inserts. Of approximately
10,000 clones screened, 30 corresponding to those with the most intense hybridization signal were selected. The inserts, on average 17 kb in length, were digested with SalI producing 52 restriction fragments. Twenty six of these were then assigned to six distinct repeated DNA families by dot blot hybridization. Four of the six families were determined to be dispersed nuclear repeats based on the "smear" of fragments observed following Southern hybridization (data not shown). The remaining two families were tandem repeated sequences, one encoding the 18S-26S rDNA as inferred by homology to pGmR1, and the other corresponding to tandem repeated sequences found in the large intergenic spacer of the 18S-26S rDNA repeating unit (Brown, Newton and Carlson, unpublished). While this survey represents only a limited examination of the spruce genome, it does suggest that the most highly repeated sequences, apart from the 18S-26S rDNA, are dispersed as opposed to tandemly arranged.

Secondly, chromosome staining with the fluorescent dyes DAPI and CMA3 revealed few AT- or GC-rich genomic regions. Fluorescent bands reflect the reiteration of satellite DNAs with a base composition distinct from the surrounding chromatin. DAPI stains the chromosomes of white, Engelmann and Sitka spruces and Picea brachytyla homogenously, with the exception of centromeres and secondary constrictions which appear as negatively staining regions (Brown and Carlson, unpublished; Hizume et al. 1991). The centromeric tandem arrays of SGR-31 do not induce a positive DAPI band probably because the sequence composition differs little from the genome average of 63% (Miksche and Hotta 1973). In the same species, CMA3 stains all secondary constrictions intensely, consistent with the high GC content of the 18S-26S rDNA. In white, Engelmann and Sitka spruces, additional CMA3 positive bands are found at several centromeric sites. Most of these are associated with the ISH site of the GC-rich tandem repeat found in the IGS of the 18S-26S rDNA and cloned
from the black spruce library described above (data not shown). In summary, it appears that neither GC- nor AT- rich tandem repeats, apart from the 18S-26S rDNA, have been amplified to any appreciable extent in the spruce genome. It would be instructive to optimize Giemsa C-banding protocols, which do not rely on differential base composition to reveal heterochromatic regions, for spruce chromosomes. This would then enable the full extent of tandem repeated DNA sequences found in telomeric, centromeric and interstitial heterochromatin to be determined, and provide a clearer view of their contribution to the spruce genome.

LITERATURE CITED


The Prospects for Non-Isotopic In Situ Hybridization in Conifer Genome Analysis

THE SENSIVITY OF IN SITU HYBRIDIZATION

Technical advances directed towards improving the detection sensitivity of nonisotopic ISH are central to its future applications in the analysis of conifer and other plant genomes. Sensitivity defines the lower size limit of chromosomal target sequences that can be visualized. While in human cytogenetics, sensitivities as low as 0.5 kb allow the routine detection of cDNAs, low copy DNA sequence mapping in plants has generally required chromosomal targets of 10 kb or more (Ambros et al., 1986; Simpson et al., 1988; Schaff et al., 1990; Leitch and Heslop-Harrison 1993). Only in rice does ISH sensitivity appear to rival that of human ISH (Gustafson and Dille 1992; Song and Gustafson 1995). Gustafson and Dille (1992) successfully mapped 23 genomic DNA clones ranging in size from 0.7-3.4 kb on the mitotic chromosomes of rice, albeit hybridization sites were detected in only 6% of metaphases analysed and were usually observed on only a single chromatid. Dong and Quick (1995) also physically mapped a 2.6 kb low copy sequence in wheat and rye but failed to report the percentage of cells showing hybridization signals.

The discrepancy in sensitivity between ISH to human and plant chromosomes arises primarily from the presence of the plant cell wall, the source of dividing material and the
methods of metaphase chromosome preparation. The preparation of human chromosomes relies exclusively on well established tissue culture methods which yield synchronized cell populations with a high mitotic index. Chromosome spreads are typically obtained by dropping a hypotonic cell suspension onto microscope slides resulting in many well spread metaphases with little or no cytoplasmic debris. In contrast, plant chromosomes prepared for ISH are usually produced by the squash technique following enzymatic digestion of the cell wall, as used in this research. This approach is effective in producing greater numbers of well spread metaphases in comparison to conventional squashes which omit the degradation of the cell wall. However, cell wall and cytoplasmic debris invariably remain and appear to hinder the hybridization and detection of low copy sequences and increase the non-specific binding of labeled probes and detection reagents. These factors, coupled with the generally low mitotic index of root tip meristems and the frequent detection of hybridization signals on only a single chromatid in a very low percentage of metaphases analysed, may raise questions concerning the authenticity of hybridization sites of small, low copy DNA probes in plants.

Most successes in low copy DNA sequence mapping in plants have been attributed to the use of protoplasts derived from suspension cultures or root tips, and to dropping protoplast suspensions onto microscope slides to spread the chromosomes. Tissue culture methods for conifers, including white spruce and several pine species (Attree et al., 1989; Gupta and Durzan 1987) are well established. Of particular relevance to genome analysis in spruce is the recent demonstration that embryogenic cultures of black spruce can be synchronized by treatment with hydroxyurea and arrested effectively in metaphase using colchicine. Mitotic indices in these cultures approached 40% (Nkongolo and Klimaszewska 1994). Adapting these culture treatments to other conifers and developing protoplast
methods for chromosome preparation (Dille et al., 1990) will significantly improve the
sensitivity of ISH.

Raap et al. (1995) described a new detection principle for fluorescence ISH in human
cytogenetics. To detect biotin-labeled probes, a layer of avidin conjugated to horseradish
peroxidase is first applied, as in many non-isotopic ISH procedures. The method is based on
the subsequent use of the peroxidase substrate tyramine conjugated to fluorochromes or
biotin. The action of peroxidase on the tyramine derivatives produces many highly reactive
intermediates that effectively immobilize fluorochrome or biotin molecules at or near the
peroxidase molecule. Fluorochromes can then be visualized by fluorescence microscopy,
while biotin molecules require an additional avidin conjugate layer. This detection method,
described as "ultra-sensitive" by the authors should be easily adapted to plants and may also
help alleviate low ISH sensitivities.

Even using the squash technique, good hybridization signals from low copy DNA
sequences could be generated using DNA probes with large inserts, such as yeast or bacterial
artificial chromosomes. If the typical conifer genome is composed of 75% repeated DNA
then inserts of 200 kb should contain approximately 50 kb of low copy sequence. Even
cosmids or P1 bacteriophage clones (with more easily managed insert sizes of 50 and 100 kb,
respectively) may contain enough low copy sequence to produce readily detected signals. It is
necessary to suppress the hybridization of repeated sequences within these clones although
this is routinely achieved in the ISH of yeast artificial chromosomes in human cytogenetics by
competition with Cot-1 DNA (Landegent et al., 1987). Genomic DNA in the appropriate
molar ratio might also serve as a competitor. Problems with the hybridization of repeated
sequences in large clones may also be circumvented by using clones from unrelated species
(although this approach might require the verification of linkage relationships between the species).

APPLICATIONS

Continued mapping of repeated DNAs

The value of physically mapping tandem arrayed repeated DNAs in spruce has been clearly demonstrated in this research as a means of identifying the morphologically similar chromosomes of conifers and in comparing the distribution and evolution of repeated DNAs between homeologous chromosomes. Even in the absence of technical advances to increase the sensitivity of ISH, this method and the ribosomal RNA probes described can be likewise applied to other conifers as a basis for comparing their genome organization and evolution with other plant forms.

The chromosomal mapping of dispersed repeated DNAs will also be a useful tool for those foresters and geneticists in British Columbia interested in the presence and extent of introgression among natural populations of Sitka spruce and interior spruce (white, Engelmann and their hybrids). While the tandem repeated DNA families studied here would be sufficient to distinguish parental species and F_1 white X Sitka spruce hybrids, meiotic recombination in the hybrids and the limited genome coverage of the repeated DNA families available restricts their ability to assess species composition in later generations of introgressive hybridization. The efforts of Sutton et al. (1991a; 1991b), using species-specific restriction fragment length polymorphisms (RFLP's) of the organelle genomes and more recently RFLP's of the intergenic spacer of the nuclear-encoded 18S-26S rDNA (Sutton et al.,
1994), are presently more reliable and simpler approaches to investigating introgression than ISH.

Some form of ISH with dispersed repeated DNA probes would be a valuable cytological complement to RFLP analysis of introgression in spruce. Genomic ISH, whereby labeled white spruce genomic DNA and an excess of unlabeled Sitka spruce DNA (or vice versa) are hybridized to chromosome preparations of individuals from putative hybrid zones or seedlots, is one possibility. However, since the ease of detecting differences between species by genomic ISH decreases with increasing overall sequence homology between the species (Jiang and Gill 1994), the apparently high degree of genome relatedness of white and Sitka spruces may preclude this approach. Alternatively, species-specific dispersed repeats present throughout only one of the two introgressing genomes could be used as hybridization probes. Either of these ISH approaches provides broader coverage of the genome than RFLP's, giving a more direct, easily interpreted analysis of the contribution of each parental species.

Low copy DNA sequence and gene mapping

A future goal of genome mapping in conifers, and one in which ISH plays a central role, is the assignment of genetic linkage groups and individual genes or other low copy DNA loci to specific chromosomal sites. This accomplishment will serve both academic and practical purposes. Correlating the genetic and physical maps of a species will allow the status of the genetic map to be assessed. It will also serve to unite different linkage groups on the same chromosome and verify that genetic markers at the end of linkage groups actually lie very near to the telomere. Integrated maps will augment future comparative genetic mapping
studies in conifers based on RFLP linkages by providing a chromosomal context in which gene order co-linearity is assessed. The results could provide insight into the extent and consequence of chromosomal rearrangements undergone during evolution and the potential application of one species map to another species.

It is becoming increasingly clear that the physical distance between genes is often markedly different than the observed genetic distance (Dvorak et al., 1984; Song and Gustafson 1995). Recombination in the distal region of cereal chromosomes is significantly more frequent than in regions proximal to the centromere (Lukaszewski 1992). Relating the genetic map to underlying chromosome structure will enable recombination rates across the conifer genome to be evaluated. The effect of reduced recombination in some genomic regions on the usable genetic variation available to tree breeders and on the value of genetic diversity estimates derived from biochemical and molecular markers may need to be addressed in the future.

Gene transfer methods for white spruce and other conifers have received considerable attention as avenues to integrate new traits into existing populations quickly. This approach to tree improvement would also benefit from successes in gene mapping by ISH as the location of the integrated transgene and the number of copies in the genome may affect gene expression.

From a practical perspective, low copy DNA sequence mapping by ISH will facilitate constructing higher density maps of conifer genomes. The current focus of genome analysis in tree improvement is the construction of genetic linkage maps and the search for associations between molecular markers and desirable economic traits. To use this information in map-based cloning or marker-assisted selection the subsequent step in defining the genomic region
involved is to isolate more tightly linked and flanking markers. This requires constructing a higher resolution map.

In genetic mapping, the smaller the map distance between loci, the larger the segregating population that is needed to ensure, at an acceptable level of probability, that loci are properly ordered. For conifer species, the construction of higher resolution genetic maps may be difficult since large numbers of segregating individuals from controlled crosses may not be available and obtaining progeny from specific genotypes or crosses can take many years. Physical mapping by ISH is a particularly valuable complement to linkage analysis in such cases since it requires seed from only one individual. The locus need not be polymorphic and segregation analysis is not required. Additionally, Lichter et al. (1990) demonstrated that DNA clones 1.1 Mb apart can be resolved across the width of each chromatid. The resolution of ISH to human metaphase chromosomes is generally 1-3 Mb (Joos et al., 1994), equivalent to roughly 1-3 centiMorgans in humans. Although it has been suggested that plant metaphase chromosomes are more condensed than human chromosomes (Greilhuber 1977; Jiang and Gill 1994), resolution of metaphase chromosome mapping in plants would still likely be less than 10 Mb.

ISH provides a quick and efficient tool in constructing higher resolution maps. To randomly add markers to a physical map, several probes can be distinctly labeled, hybridized and independently localized to chromosomes in a single experiment. Clones which map to the same position on low resolution genetic maps may also be quickly ordered. Once the chromosomal location of a marker linked to a desirable trait is known, then clones mapping in proximity to that marker can be readily identified. To find the most tightly linked flanking markers would still require segregation analysis. However, if large insert clones were used as
ISH probes in generating the higher resolution map, the proximal markers could be screened for highly informative microsatellite or minisatellite polymorphisms to maximize the possibility of detecting recombination between the markers and the trait locus.

Generating a high resolution map of an entire conifer genome is an enormous task. The random addition of molecular markers to a genetic map is both time and resource consuming and could be simplified if approaches to target directly a specific chromosome or chromosomal region of interest were developed. The construction of chromosome-specific libraries using flow sorting methods permits this (Arumuganathan et al., 1991; Wang et al., 1992), although it may not be possible given the overall uniformity in conifer chromosome size. Alternatively, microdissection methods (Jung et al., 1992) or the sorting of individual chromosomes by fluorescence ISH in conjunction with flow sorting or avidin-conjugated magnetic particles are possible (Dudin et al., 1988; Gray et al., 1986). The chromatin isolated could then be cloned into cosmids or other large insert vectors and individual clones rapidly ordered along the chromosome of interest.

To what extent high resolution maps of conifer genomes will be required is uncertain. For dissecting quantitative traits or marker-assisted selection, the resolution offered by ISH mapping on metaphase chromosomes is likely sufficient. Cloning genes from a conifer genome based on their map position would require the generation of a much higher resolution physical map and the development of techniques to manipulate and sequence large tracts of DNA. Conifer genomes are large and primarily repetitive which hampers map-based cloning strategies. Unless a particular gene is specific to a conifer species, it will be more practical to clone it from a well characterized plant with a small genome such as Arabidopsis (or poplar) and then use this to probe a conifer genomic or cDNA library. If the situation arises in which
very high resolution maps are required, ISH methods developed in human cytogenetics boast 50 kb - 1 Mb resolution on interphase chromatin (Lawrence et al., 1988; Trask 1991) and 1 - 400 kb resolution using extended chromatin fibers (Parra and Windle 1993; Florijn et al., 1995). Both techniques rely on the decondensed state of chromatin either at interphase or using extended single fibers of DNA released from nuclei fixed on slides.

Prior to the broad use of molecular cytogenetics, plant and animal genomes could only be studied at the DNA sequence level, by genetic recombination, and at the chromosomal level. Only by knowing the physical position of genes along a chromosome can genetic linkage, and gene and sequence interdependence, be understood (Heslop-Harrison 1991). ISH techniques bridge the gap between the resolutions of these methods allowing the chromosomal location of linkage groups to be determined, the relationship between genetic and physical distance to be understood, and the eventual cloning of genes based on map position.

LITERATURE CITED


Dvorak, J. and Chen, K.-C. 1984. Distribution of nonstructural variation between wheat cultivars along chromosome 6Bp: evidence from the linkage map and physical map of the arm. Genetics, 113: 325-333.


**APPENDIX I**

Common fluorochromes and their excitation and emission maxima.

<table>
<thead>
<tr>
<th>Fluorochrome</th>
<th>Excitation Maximum (nm)</th>
<th>Emission Maximum (nm)</th>
<th>Fluorescent Color</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nucleotide Conjugates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coumarin (AMCA)*</td>
<td>350</td>
<td>600</td>
<td>blue</td>
</tr>
<tr>
<td>Fluorescein (FITC)*</td>
<td>494</td>
<td>520</td>
<td>green</td>
</tr>
<tr>
<td>Cyanine Cy 3</td>
<td>550</td>
<td>565</td>
<td>green</td>
</tr>
<tr>
<td>Rhodamine600 (TRITC)*</td>
<td>575</td>
<td>600</td>
<td>red</td>
</tr>
<tr>
<td>Texas Red</td>
<td>596</td>
<td>615</td>
<td>red</td>
</tr>
<tr>
<td>Cyanine Cy 5</td>
<td>650</td>
<td>670</td>
<td>red</td>
</tr>
<tr>
<td><strong>DNA Stains</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAPI*</td>
<td>359</td>
<td>461</td>
<td>blue</td>
</tr>
<tr>
<td>Propidium iodide</td>
<td>340, 530</td>
<td>617</td>
<td>red</td>
</tr>
<tr>
<td>Chromomycin A&lt;sub&gt;s&lt;/sub&gt;</td>
<td>458</td>
<td>590</td>
<td>yellow</td>
</tr>
</tbody>
</table>

* 7-amino-4-methly-coumarin-3-acetic acid;  
<sup>b</sup> fluorescein isothiocyanate;  
<sup>c</sup> tetramethyl rhodamine isothiocyanate;  
<sup>d</sup> 4′,6-diamidino-2-phenyliodole.