Durham University

Durham E-Theses

Origins of Forest Patch Structure in Mountain Birch (Betula pubescens ssp. czerapanovii) in Sub-Arctic Northern Sweden; The Effects of Introgressive Hybridisation, Small Scale Genetic Selection and Individual Responses to Freezing Shock.

ROWLAND, ALICE,VICTORIA

How to cite:

ROWLAND, ALICE,VICTORIA (2013) *Origins of Forest Patch Structure in Mountain Birch (Betula pubescens ssp. czerapanovii) in Sub-Arctic Northern Sweden; The Effects of Introgressive Hybridisation, Small Scale Genetic Selection and Individual Responses to Freezing Shock.* , Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/7776/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, Durham University, University Office, Old Elvet, Durham DH1 3HP
e-mail: e-theses.admin@dur.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Origins of Forest Patch Structure in Mountain Birch (*Betula pubescens* ssp. *czerapanovii*) in Sub-Arctic Northern Sweden; The Effects of Introgressive Hybridisation, Small Scale Genetic Selection and Individual Responses to Freezing Shock.

Alice Victoria Rowland

School of Biological and Biomedical Sciences

Durham University

Submitted for the Degree Masters by Research (2013)

Abstract

The origins of observed morphological differences between windward, central and leeward sites within forest patches of *Betula pubescens* ssp. *czerapanovii* in Abisko, Northern Sweden were investigated. Samples from each of these sites from three forest patches of *Betula pubescens* ssp. *czerapanovii* and surrounding shrubs of *Betula nana* were analysed genetically. The findings showed no overall genetic differences between sites within a patch but did show differences between patches. Studies of an individual gene did however show evidence that selection may occur on individual genes at different sites. *B. nana* and *B. pubescens* ssp. *czerapanovii* were shown to be highly introgressively hybridised to the extent that *B. nana* were more closely related to neighbouring patches of *B. pubescens* ssp. *czerapnovii* than to *B. nana* adjacent to more distant patches This indicates that the hybridisation in Abisko is so great that separating these species is no longer justified from a molecular standpoint. Laboratory grown seedlings showed that morphological differences seen in *B. pubescens* ssp. *czerapanovii* have a genetic basis. Seedlings were exposed to different cold acclimation periods under summer photoperiod conditions before a freezing shock. Acclimation time had no effect on survival. Polycormic individuals showed evidence of slightly greater survival after freezing. Otherwise morphology had no effect on survival chances. The potential significance of these findings on understanding forest patches responses to global warming are discussed along with potential future experimental procedures required to further elucidate the origin of the observed morphological differences between sites in a patch. This includes whether the degree of polycormy is being genetically selected for at these sites, why this selection may be occurring and if developmental plasticity or mechanical damage are also affecting observed morphologies.

Contents

Abstract……………………………………………………………………………... i

Contents……………………………………………………………………….…..... ii

List of Figures…………………………………………………………………….. vii

List of Tables…………………………………………………………………...…. xii

List of Abbreviations…………………………………………………………...... xiii

Declaration of Copyright……………………………………………………...… xiv

Acknowledgements……………………………………………………………..… xiv

Chapter One: Introduction…………………………………………………….... 1

Chapter Two: Materials and Methods………………………………………….. 12

2.1 Sample Collection.……………………………………………………………….. 12

2.1.1 Leaves……………………………………………………………………….... 12

2.1.2 Buds………………………………………………………………………….. 13

2.1.3 Seeds…………………………………………………………………………. 13

2.1.4 Transport and Storage……………………………………………………..... 14

2.2 Growth of Test Plants…………………………………………………………….. 14

2.2.1 Germination………………………………………………………………….. 14

2.2.2 Tracking Individuals………………………………………………………… 15

2.2.3 Light Response……………………………………………………………… 15

2.2.4 Growth Conditions………………………………………………………….... 15

2.2.5 Mycorrhizal Inoculations…………………………………………………….. 16

2.3 Freezing Experiments…………………………………………………………….. 16

2.3.1 Selection of Individuals……………………………………………………… 16

2.3.2 Chamber Conditions………………………………………………………… 17

2.3.3 Morphology Data……………………………………………………………….. 17

2.3.4 Survival Tests………………………………………………………………... 18

2.4 Molecular Analysis……………………………………………………………….. 18

2.4.1 Buffers and Solutions…………………………………………………... 18

2.4.1.1 Edwards Buffer……………………………………………………….. 18

2.4.1.2 1M TE Buffer……………………………………………………...…… 18

2.4.1.3 Dellaporta Extraction Buffer………………………………….………… 19

2.4.1.4 TBE Buffer…………………………………………………………… 19

2.4.1.5 TAE buffer……………………………………………………………..19

2.4.1.6 CTAB Buffer…………………………………………………………. 20

2.4.2 Primers.………………………………………………………………… 20

2.4.2.1 Primer Sequences……………………………………………….…… 20

2.4.2.2 Primers for RAPD Analysis…………………………………………..… 21

2.4.3 DNA Extraction…………………………………………………………… 22

2.4.3.1 Edwards……………………………………………………………..… 22

2.4.3.2 Dellaporta……………………………………………………………..… 23

2.4.3.3 Phenol-chloroform Purification of DNA……………………………..… 23

2.4.3.4 CTAB……………………………………………………………… 24

2.4.3.5 Powerplant ® Pro DNA Isolation……………………………..…… 25

2.4.3.6 Serial Dilutions……………………………………………………….. 26

2.4.4 PCR Conditions…………………………………………………………… 27

2.4.4.1 PCR Mix One……………………………………………………… 27

2.4.4.2 PCR Mix Two – KAPA3G………………………………………………… 27

2.4.4.3 PCR Mix Three – KAPA3G for RAPDs……………………………...… 27

2.4.4.4 PCR Program One……………………………………………….… 28

2.4.4.5 PCR Two (KAPA3G)…………………………………………………… 28

2.4.4.6 PCR Three (KAPA3G) Modified……………………………………..… 29

2.4.4.7 PCR Four (KAPA3G) for RAPDs…………………………………….… 29

2.4.5 Gels…………………………………………………………………... 29

2.4.5.1 TBE Gels……………………………………………………...… 30

2.4.5.2 TAE Gels……………………………………………………...… 30

2.4.5.3 Imaging Gels…………………………………………………..… 30

2.4.6 Gel Band Extraction………………………………………………. 30

2.4.7 Sequencing…………………………………………………………... 32

2.4.8 Comparison ADH Sequences……………………………………... 32

2.4.9 RAPDs……………………………………………………………….. 32

2.5 Analysis of Results………………………………………………….. 33

2.5.1 ADH Sequences…………………………………………………..… 33

2.5.2 RAPDs……………………………………………………………….. 34

Chapter Three: Results and Discussion.………………………........................... 36

3.1 DNA Extraction…………………………………………………………... 36

3.1.1 Background…………………………………………………………36

3.1.2 Edwards Extraction………………………………………………...37

3.1.3 Dellaporta Extraction……………………………………………… 40

3.1.4 Phenol-Chloroform Purification of DNA………………………………. 41

3.1.5 ADH Primers……………………………………………………… 42

3.1.6 CTAB……………………………………………………………... 43

3.1.7 KAPA3G Enzyme Kit……………………………………………….. 45

3.1.8 MoBio PowerPlant® Pro DNA Isolation Kit………………………….... 49

3.1.9 Conclusions and Summary…………………………………………52

3.2 ADH Alleles………………………………………………………….54

3.2.1 Background……………………………………………………………………54

3.2.2 Gel Bands…………………………………………………………………. 55

3.2.3 Sequence Data……………………………………………………………….. 68

3.2.4 Insertions, Deletions and Misalignments……………………………………..93

3.2.5 The Middle Band……………………………………………………………100

3.2.6 Dendrograms of Similarity…………………………………………………..102

3.2.7 Conclusions…………………………………………………………………112

3.2.8 Summary……………………………………………………………………..120

3.3 RAPD Analysis………………………………………………………………..121

3.3.1 RAPD Primers………………………………………………………….....121

3.3.2 Sample Morphologies………………………………………………...…...123

3.3.3 Peak Groupings…………………………………………………………..… 124

3.3.4 Patch One………………………………………………………………… 136

3.3.5 Patch Two…………………………………………………………………140

3.3.6 Patch Three………………………………………………………………. 143

3.3.7 All Patches……………………………………………………………….. 147

3.3.8 Conclusions and Summary…………………………………………………..153

3.4 Physiology and Freezing Shock……………………………………………… 156

3.4.1 Light Response Curves……………………………………………………... 156

3.4.2 Morphology and Survival…………………………………………………... 157

Chapter Four: Conclusion………………...………………………….................... 169

4.1 Conclusions………………………………………………………………..…..169

4.1.1 Origins of Forest Patch Structure…………………………………………... 169

4.1.2 Potential Consequences of Global Warming on Forest Patches…………… 172

4.1.3 Other Findings of Particular Interest……………………………………….. 173

Bibliography………………………………………... 174

List of Figures

Fig. 2.1.1.1 Location of sites sampled within three forest patches……………….. 12

Fig. 3.1.2.1 PCR Products of DNA from Edwards extraction…………………….. 37

Fig. 3.1.4.1 Phenol-chloroform purification of Dellaporta DNA extraction………. 41

Fig. 3.1.5.1 PCR products from serial dilutions of Edwards and Dellaporta DNA extractions with ADH primers……………………………………………………43

Fig. 3.1.5.2 PCR products of serial dilutions of phenol-chloroform purified Dellaporta DNA extraction with ADH primers……………………………………. 43

Fig. 3.1.6.1 CTAB DNA extraction…………………………………………….. 44

Fig. 3.1.6.2 PCR products from CTAB DNA extraction………………………….. 44

Fig. 3.1.6.3 PCR products from serial dilutions of CTAB DNA extraction………. 45

Fig. 3.1.7.1 PCR products from CTAB DNA extraction with KAPA3G enzyme… 46

Fig. 3.1.7.2 PCR products from phenol-chloroform purified Dellaporta extraction with KAPA3G enzyme kit……………………………………………………. 47

Fig. 3.1.7.3 PCR amplification of CTAB extracted DNA and sections of leaf tissue with varying amounts of plant PCR enhancer using the KAPA3G enzyme kit…... 48

Fig. 3.1.8.1 DNA extraction using the MoBio PowerPlant® Pro DNA Isolation kit comparing a pre grinding step to normal protocol procedure…………………... 50

Fig. 3.1.8.2 DNA extraction using the MoBio PowerPlant® Pro DNA Isolation kit with a pre grinding step comparing with and without bead-beating step………….. 51

Fig. 3.2.2.1 PCR products amplified with the ADH primer pair from DNA from samples (1) 5.34C (2) 5.11A……………………………………………………. 56

Fig. 3.2.2.2 PCR products amplified with the ADH primer pair from DNA from samples (1) 6.18F (2) 6.29A. (3) 6.12C (4) 6.35E (5) 6.22A…………………… 57

Fig. 3.2.2.3 PCR products amplified with the ADH primer pair from DNA from samples (1) DW4. (2) DW5 (3) DW6………………………………………….. 57

Fig. 3.2.2.4 PCR products amplified with the ADH primer pair from DNA from samples (1) BPS1. (2) BPS2. (3) BPS3. (4) BPS4. (5) BPS5. (6) 3.3A. (7) 3.5C. (8) 3.7A. (9) 3.18E. (10) 5.32A. (11) 5.28B. (12) BIR2. (13) BIR3. (14) BIR4. (15) BIR5……………………………………………………………………….. 57

Fig. 3.2.2.5 PCR products amplified with the ADH primer pair from DNA from samples (1) BIR6. (2) N3.1. (3) N3.2. (4) N3.3. (5) N3.5. (6) N5.1. (7) N5.2. (8) N5.4. (9) N5.5. (10) DW1. (11) DW3………………………………………... 58

Fig. 3.2.2.6 PCR products amplified with the ADH primer pair from DNA from samples (1) 3.12G. (2) N5.3……………………………………………………. 58

Fig. 3.2.2.7 PCR products amplified with the ADH primer pair from DNA from sample (1) BIR1……………………………………………………………… 58

Fig. 3.2.2.6 PCR products amplified with the ADH primer pair from DNA from sample (1) N3.4……………………………………………………………….. 59

Fig. 3.2.3.1 Example from sequence of 3.3A (short) of an 'N' base consisting of three different bases……………………………………………………………... 69

Fig. 3.2.3.2 Sections of ADH sequence chromatograms showing sudden misalignment of sequence………………………………………………………. 69

Fig. 3.2.3.3 – 3.2.3.13 Full sequence of short alleles aligned with published long alleles and short allele…………………………………………………….. 71-81

Fig. 3.2.3.14-24 Full sequence of long alleles aligned with published long alleles and short allele……………………………………………………………………… 82-92

Fig. 3.2.6.1 Dendrogram of relatedness based on partial ADH genes for a section of the forward sequence of the short allele…………………………………………104

Fig. 3.2.6.2 Dendrogram of relatedness based on partial ADH genes for a section of the reverse sequence of the short allele……………………………………………105

Fig. 3.2.6.3 Dendrogram of relatedness based on partial ADH genes for a section of the forward sequence of the long allele……………………………………………106

Fig. 3.2.6.4 Dendrogram of relatedness based on partial ADH genes for a section of the reverse sequence of the long allele……………………………………………107

Fig. 3.2.6.5 Dendrogram of relatedness based on partial ADH genes for full sequences…………………………………………………………………..110

Fig. 3.3.1.1 PCR products of DNA from sample BIR4 with RAPD primers (1)ATC33. (2) OPB17. (3)OPB14. (4) OPB11. (5) OPB12. (6) OPB18. (7) ATC71. (8) ATC51. (9) RAPDa. (10) RAPDb. (11) RAPDc. (12) RAPDd. (13) RAPDe. (14) OPB08…………………………………………………………………..121

Fig. 3.3.1.2 Electropherogram readouts from the bioanalyser for DNA from sample 3.7A run with various RAPD primers……………………………………………..122

Fig. 3.3.3.1 Example of degree of error for the same sample from the same PCR run and primers on three different bioanalyser runs…………………………………...124

Fig. 3.3.3.2 Electropherogram showing division of peak groups for Patch one samples with primer ATC51……………………………………………………125

Fig. 3.3.3.3 Electropherogram showing division of peak groups for Patch one samples with primer OPB11……………………………………………………126

Fig. 3.3.3.4 Electropherogram showing division of peak groups for Patch one samples with primer OPB12……………………………………………………127

Fig. 3.3.3.5 Electropherogram showing division of peak groups for Patch one samples with primer OPB14……………………………………………………128

Fig. 3.3.3.6 Electropherogram showing division of peak groups for Patch two samples with primer ATC51……………………………………………………129

Fig. 3.3.3.7 Electropherogram showing division of peak groups for Patch two samples with primer OPB11……………………………………………………130

Fig. 3.3.3.8 Electropherogram showing division of peak groups for Patch two samples with primer OPB12……………………………………………………131

Fig. 3.3.3.9 Electropherogram showing division of peak groups for Patch two samples with primer OPB14……………………………………………………132

Fig. 3.3.3.10 Electropherogram showing division of peak groups for Patch three samples with primer ATC51……………………………………………………133

Fig. 3.3.3.11 Electropherogram showing division of peak groups for Patch three samples with primer OPB11……………………………………………………134

Fig. 3.3.3.12 Electropherogram showing division of peak groups for Patch three samples with primer OPB12……………………………………………………135

Fig. 3.3.3.13 Electropherogram showing division of peak groups for Patch three samples with primer OPB14……………………………………………………136

Fig. 3.3.4.1 Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples for patch one………………………………138

Fig. 3.3.5.1 Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples for patch two………………………………143

Fig. 3.3.6.1 Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples for patch three…………………………...145

Fig. 3.3.7.1 Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples across all patches………………………… 149

Fig. 3.4.1.1 Light response curves of photosynthesis for old leaves……………... 156

Fig. 3.4.1.2 Light response curves of photosynthesis for young leaves………..… 157

List of Tables

Table 3.3.2.1 degree of polycormy in patch one and two individuals……………..123

Table 3.3.4.1 Dice’s similarity coefficient between samples from patch one……..137

Table 3.3.5.1 Dice’s similarity coefficient between samples from patch two……..142

Table 3.3.6.1 Dice’s similarity coefficient between samples from patch three…... 146

Table 3.4.2.1 Treatment types, morphologies and survival of individual saplings………………………………………………………………………. 163-165

Table 3.4.2.2 details of branches of individuals characterised in (Table 3.4.2.1)………………………………………………………………………. 166-168

List of Abbreviations

ADh	Alcohol dehydrogenase
bp	Base pairs

Declaration of Copyright

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

I would like to thank my supervisors Marc Knight and Bob Baxter for their endless patience, support and kindness not only throughout the project but further into my academic career. I would also like to thank all members of the plant stress and signalling lab and members of Bob Baxter's research group with special thanks Rachael Oakenfull for collecting samples, caring for the saplings after my time at Durham had ended, her patience in teaching me various laboratory techniques including techniques learned through her own research, pioneering molecular arctic research at Durham and finally for her support and friendship, Rob Holden for kindly collecting samples and such short notice and under adverse conditions without which this project could not have been completed and Stephanie Johnson for her endless entertainment and camaraderie. I would also like to thank Michael Bone for his patience and rapid help with fixing the frequent problems that occurred with the growth chambers and willingness to take the time to teach me how to fix problems that may occur in future. Last but not least I would like to thank all my other friends and family for their understanding and support throughout this project.

Chapter One: Introduction

Plant life in the arctic has to persist and thrive under a variety of stresses. In addition to bouts of extreme cold, stresses such as short summers, long dark winters and long droughts followed by rapid availability of water are encountered. How species have adapted to survive in these extreme and changeable conditions and how they continue to adapt are matters of great scientific interest. With global warming, the intensity and regularity of these stresses are expected to, and currently are, changing rapidly (ACIA, 2005). In response, species are likely to change primarily by altering their distribution (Parmesan &Yohe, 2003). Changes in vegetation distribution may also affect global warming by altering the albedo (i.e. altering the reflectivity of the earth's surface) and hence how much of the incident solar ration is reflected back or absorbed, changing carbon storage and uptake, soil-vegetation interactions and impacting upon other ecological processes (ACIA, 2005). Hence it is important to attempt to understand how these climatic changes will affect the arctic ecosystem.

To different people the term 'arctic' denotes different areas (Nuttal & Callaghan, 2000). From a biological standpoint the flora and fauna are the key factors, which are influenced by a variety of abiotic and biotic factors. A good general reference for this is the treeline, where one dominant vegetation type (forest) changes to another (tundra). For the purpose of this project, areas clearly north of the treeline are termed arctic, and the often broad boreal forest-tundra ecotone to the south as subarctic.

The tree line is the northernmost limit where trees greater than 2 m in height cease to grow (Hofgaard, Dalen & Hytteborn, 2009) and tundra environment dominates. This is rarely a well defined line of trees, and is usually preceded by a gradual breakup of the forest cover leading to fragmented patches of woodland. Beyond the tree line shrubs and prostrate trees less than 2 m in height can still be found in isolated patches. In Abisko, Northern Sweden, the treeline is formed predominantly by *Betula pubescens subsp. czerapanovii* (also known as *subsp. tortuosa*). This species provides a relatively open canopy allowing a variety of understory plants to grow including the closely related dwarf shrub *Betula nana* which expands its range out of the forest patches and into the tundra. This treeline boundary, and others globally, tend to roughly follow the 10°C mean July isotherm (Stonehouse, 1989) which indicates one of the potential influencing factors in the creation of this boundary. One theory suggests that cold-induced loss in photosynthetic rate makes the respiration cost of large amounts of non-photosynthetic material greater than any benefits of greater stature (Stevens & Fox, 1991). Other potential explanations include exposure of taller plants to the harsher conditions above the snowpack, which can result in costs such as drought, induced by extreme cold and strong winds, ice blast which physically damages exposed parts, and browsing by herbivores (Stevens & Fox, 1991). Another theory suggests that nutrient limitation occurs in cold environments, with small localised patches of nutrients favouring plants with smaller more localised root systems, resulting in reduced height (Stevens & Fox, 1991). It is likely that a variety of these factors affect treeline to varying degrees and the amount of influence is likely to alter with location and species.

Soil temperature has been shown to affect nitrogen uptake in *B. pubescens subsp. czerapanovii*, with temperatures lower than 5°C resulting in negligible nitrogen uptake and growth (Karlsson & Weih, 1996). As temperature increases growth rate and nitrogen uptake increase rapidly up to 15°C, above which there is little change in uptake (Karlsson & Weih, 1996). Nitrogen availability has also been found to decrease as the treeline is approached, and growth rates can be increased by nitrogen fertilization of those trees in the transition zone and at the treeline (Sveinbjörnsson, Kauhanen & Nordell, 1996). A faster growth rate, which can be increased by high (15°C) soil temperatures or fertilization, increases the chance of overwinter survival (Weih & Karlsson, 1999). Based on these factors Karlsson and Weih (2001) calculated the minimum mean growing season soil temperature for germination and establishment of *B. pubescens subsp. czerapanovii*. On disturbed ground with no competition, a mean temperature of 9.8^{o}C was required, increasing to 12°C or higher for disturbed ground with some competition (Karlsson & Weih, 2001). This correlates well with the general localization of the treeline to the 10°C July isotherm. As calculated by Karlsson and Weih (2001) nitrogen uptake, and as a result the temperature required for establishment, can be influenced by competition. Competition resulting in soil shading reduces nitrogen uptake in *B. pubescens subsp. czerapanovii* by the reduction of soil temperature (Weih & Karlsson, 1999). Allelopathic plants such as *Empetrum nigrum* (which produces biochemicals that suppresses growth of other species), also reduce nitrogen uptake in *B. pubescens subsp. czerapanovii* (Weih & Karlsson, 1999). Shading and allelopathy can therefore greatly impede seedling establishment, growth rate and consequently survival, resulting in an alteration of the range of the treeline.

The ability of seedlings to establish, survive and grow to a height above 2 m defines the location of the treeline. Expansion of the treeline would occur if conditions favoured growth of seedlings to a height greater than 2 m beyond the current limit. Conversely if seedlings were unable to grow at the treeline to replace old individuals there would be a steady retreat from the current limit. Germination of *B. pubescens subsp. czerapanovii* occurs approximately one week after snowmelt (Kullman, 1986) and the rate of germination increases with increasing temperature (Milbau et al., 2009). General studies of the Arctic and sub-Arctic show that snow-melt is occurring earlier in the year (Dye, 2002, ACIA, 2005). A more localised study in the Abisko area shows no overall change in the duration of snow cover between 1913 and 2004 (Kohler et al, 2006). However over the next several decades it is expected that duration of snow cover in Abisko will also decrease (Kohler et al, 2006). Since seedling mortality is linked to size, increased temperatures and earlier snow melt would result in a longer growing period potentially allowing the treeline to advance (Kullman, 1986 and Milbau et al. 2009).

Establishment of seedlings after germination can be affected by the microtopography of the site, with sites with varied microtopography usually having higher treelines (Kjällgren & Kullman, 1998). Exposed sites impede seedling establishment due to low moisture and nutrient availability (Anschlag, Broll & Holtmeier, 2008). These sites, unprotected by snow, can also be subjected to extreme winter cold and frost-heave. Disruption caused by frost-heave can result in loose rooting of young seedlings making them more susceptible to drought in summer (Kullman, 1986). Summer drought also affects seedlings in less exposed areas that are in competition

for water and is a common cause of seedling death in established forests (Kullman, 1986).

Once established, seedlings can persist with little change in stature in harsh environments (Kullman, 1984). It is common for saplings, less than 2 m in height (and therefore not regarded as markers of treeline position) to be established up to 100 m above the treeline in currently unfavourable conditions (Hofgaard, Dalen & Håkan, 2009). Saplings beyond the treeline were found to be between 45 and 66 years old in 2001 (Kullman, 2001) and some 95 year old saplings have been found in sheltered areas (Hofgaard, Dalen & Håkan, 2009). However, the age of saplings declines with altitude (Hofgaard, Dalen & Håkan, 2009).This suggests that either due to harsher conditions further from the treeline, persistence to a greater age is not possible, or that saplings have more recently been able to establish further from the treeline than was previously the case to due to environmental changes. In 2002 thirty-five, 8-12 year old saplings, of *Betula pubescens subsp. czerapanovii*, were found in snow bed vegetation 435 m above the treeline limit (Kullman, 2002). The age of these saplings corresponds with germination and establishment during local warming in the 1990s (Kullman, 2001) suggesting that saplings have been able to establish further from the treeline in recent decades due to a warmer climate. Hofgaard, Dalen and Håkan (2009) suggested that the advanced age of the saplings implies that a long time is required for saplings to reach tree size and hence treeline advance in response to warming may be slow. However Kullman (2001) observed that these saplings respond to environmental change with a change in growth rate. During the warmer 1990s growth rate increased and numerous 1-1.5 m saplings reached 2 m or greater, rapidly advancing the treeline by tens of meters (Kullman,

2001). Many of these saplings (45-66 years old in 2001) would have established in the warmer period around the 1930s, persisted through the local cooler following periods and resumed more rapid growth in the warmer 1990s (Kullman, 2001), resulting in a rapid change from stunted seedlings to established treeline trees.

Established trees, as well as saplings, can also suffer from mechanical damage. Heavy snow can break main stems resulting in basal re-sprouting or even death, as occurred to the treeline of *B. pubescens subsp. czerapanovii* in the Scandes mountains, Norway during the 1970s-1980s (Kullman, 1989). Snow depth in Abisko has been increasing since the 1930s/1940s, despite temperature fluctuations (Kohler et al, 2006). Hence a greater amount of snow damage resulting in tree loss or reduction in height could be expected in the short term, which could result in regression of treeline in the Abisko region. However the advantages of warming outweigh that of mechanical damage and recent warming has caused an advance in the treeline in Abisko and surrounding areas (Truong, Palmé & Felber, 2007). Tree density has also increased over the past 34 years both at the treeline and deeper in the forest (Rundqvist, Hedenås & Sandström, 2011).

Temperatures in the arctic and sub-arctic are currently increasing (ACIA, 2005) and expected to increase by between 4°C and 7°C by the end of the century (ACIA, 2005). This change is not equal throughout the seasons; winter shows the greatest temperature change. (ACIA, 2005). The conservative B2 SRES scenario of climate change estimates an increase of 5°C on land in the winter relative to 1-2°C increase on land in summer by the end of the century (ACIA, 2005). This warming is likely to result in continued treeline advancement in Abisko and elsewhere. Despite an

increase in temperature (ACIA, 2005), seedling deaths due to summer drought (Kullman, 1986) are unlikely to greatly impede treeline advancement. Precipitation is expected to increase in the arctic and subarctic with the warming climate (ACIA, 2005) and earlier snow melt (Kohler et al, 2006) with increased winter temperatures (ACIA, 2005) is likely to extend the growing period, growth and germination rate (Kullman, 1986 & Milbau et al. 2009), providing greater time for establishment of a root system and greater tolerance of summer drought. Advancement and increased forest density, likely in response to current warming, has also already been observed in Abisko (Truong, Palmé, & Felber, 2007; Rundqvist, Hedenås & Sandström, 2011), despite warmer summers (ACIA, 2005) which could otherwise increase the risk of summer drought (Kullman,1986).

In Abisko, forest patches approaching the treeline may be expected to respond to climate warming via expansion and merging. However the patches have a particular structure which may affect how this expansion occurs. These patches are made up of *B. pubescens subsp. czerapanovii* with nearby and understory *B. nana*. The Southwest edge consists of stunted highly polycormic (many stemmed) trees which increase in size and decrease in polycormy as the Northeast edge is approached (unpublished data: Holden, R & Huntley, B). In Abisko the dominant wind direction is from the south west (personal communication: Baxter, R). Wind speed drops upon entering the patches, so in winter this results in a build up of snow a short distance into the patches (Unpublished PhD data: Holden, R, University of Durham). The extreme edges of the forest patches are therefore less protected by snow cover. The windward (Southwest) edge trees are also not protected from the wind by other trees and being highly exposed they may suffer from iceblast, cold-induced drought and

more extreme temperatures. It is unknown how this pattern occurs and is maintained. The stunting and increased polycormy close to the windward edge could be a case of mechanical damage, phenotypic plasticity, genetic selection possibly via hybridization with the hardier *B. nana*, which can survive beyond the treeline (personal observation), or a combination of these factors.

Mechanical damage could cause the observed polycormy on the windward edge and hence a reduction in height (Verwijst, 1988), via the loss of main meristems to extreme frost, iceblast or grazing. However this is unlikely to be the only factor. Verwijst (1988) found that many polycormic stands of *B. pubescens subsp. czerapanovii* showed no sign of mechanical damage. *B. pubescens subsp. czerapanovii* already shows a degree of phenotypic plasticity. Verwijst (1988) found a correlation between soil pH and polycormy as well as a negative correlation between polycormy and snow depth. However neither of these are likely to be the cause of the observed patch structure in Abisko; pH is unlikely to vary to the same degree and across the same gradient and direction across every forest patch in Abisko and snow, deposited by the wind, builds up a short distance into the patch and then snow depth drops off across the patch gradient (Unpublished data: Holden, R). Although protected from the worst winds by other trees the leeward edge has little snow coverage and hence polycormy would be expected at the leeward edge also if mechanical damage was the sole cause, but leeward individuals are the least polycormic (unpublished data: Holden, R & Huntley, B). Under stressful conditions above the treeline, seedlings appear to slow their growth rate and become stunted, resuming a faster growth rate with more favourable conditions (Kullman, 2001). These seedlings can persist for long periods of time, and 95 year old individuals have

been recorded (Kullman, 2001; Hofgaard, Dalen & Håkan, 2009). This could be a direct response to harsh growing conditions or a plastic response, and a similar change in growth rate could be occurring at the windward edge. Growth rate is negatively correlated with freezing tolerance, for example oak families from temperate and tropical locations (Koehler, Center & Cavender-Bares, 2011), hence it is possible that the change in growth rate in *B. pubescens subsp. czerapanovii* in response to temperature may also be due to an alteration of freezing tolerance. Polycormy may also be a phenotypic response to cold, reducing the risk associated with one central main stem (Stevens & Fox, 1991) and the reduced height associated with polycormy (Verwijst, 1988) may allow exploitation of the slightly warmer temperatures at the boundary layer (Stevens & Fox, 1991). However the suckering ability (and hence ability to become polycormic) is linked to genetic differences between individuals (Vaarama & Valannae, 1973 in Verwijst, 1988) and hence the degree of polycormy may be due to hybridization with *B. nana* (Vaarama & Valannae, 1973 in Verwijst, 1988).

Hybridization is known to occur between *B. nana* and *B. pubescens subsp. czerapanovii* in Iceland (Elkington, 1968; Anamthawat-Jónsson & Tomasson, 1990; Anamthawat-Jónsson et al., 2010; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2003). Initially this was inferred morphologically via the comparison of presumed pure UK *B. nana* and *B. pubescens* ssp. *czerapanovii* plants with Icelandic plants which showed a gradient of intermediate morphologies, suggesting gene flow was occurring between these two species via introgressive hybridization (Elkington, 1968). Later studies confirmed this by actively crossing diploid *B. nana* (2n=28) and tetraploid *B. pubescens* ssp. *czerapanovii* (2n=56)

producing triploid progeny (n=42) (Anamthawat-Jónsson & Tomasson, 1990). These triploids were not sterile and could be backcrossed with their parent species to produce triploid (n=42), tetraploid (n=56) and diploid (2n=28) progeny, suggesting that the triploid hybrids can produce both diploid and anuploid pollen (Anamthawat-Jónsson & Tomasson, 1990). This was later tested *in-situ* with *B. nana* seed parents producing both triploid and diploid progeny (Anamthawat-Jónsson & Tomasson, 1999). The ability of hybrids to backcross and integrate genes back into parent populations means that a degree of gene flow can likely occur between the two species. Using Southern blotting techniques to probe for genes from each species Thórsson, Salmela and Anamthawat-Jónsson (2001) showed the occurrence of introgressive hybridization between icelandic *B. nana* and *B. pubescens* ssp. *czerapanovii* in both directions. Some of these swapped genes appear to have become stabilized in the population; Several phylogenetic trees based on the ADH gene in *Betula* always show that the long ADH allele of *B. pubescens* is very closely related to the ADH allele of *B. nana*, whilst the short ADH allele of *B. pubescens* is always located elsewhere in the tree (Järvinen et al., 2004). Triploids can therefore act as a corridor for introgressive hybridization between the two species and are relatively common; approximately 10% of randomly trees sampled by Anamthawat-Jónsson et al. (2003) in Iceland were triploid. These triploids were mostly found on the edges of *B. pubescens* patches leading Anamthawat-Jónsson et al. (2003) to suggest that triploids may not be able to develop in established woodland. Another factor may be that these triploids are better adapted to survive on the edges of birch patches than *B. pubescens* due to larger numbers of genes from the hardier *B. nana.*

This project aims to investigate the hypothesis that, due to differing stresses across the patches, genetic selection via hybridization and introgression is occurring across small (~100 m diameter) forest patches in Northern Sweden and resulting in the observed forest patch structures.

Chapter Two: Materials and Methods

2.1 Collection of Samples

2.1.1 Leaves

Leaf samples were collected from forest patches in Abisko, Northern Sweden. Three patches were sampled, Patch one (N68°19'26" E18°49'49"), Patch two (N68°19'11" E18°49'33") and Patch three (N68° 19'12" E18°50'16").

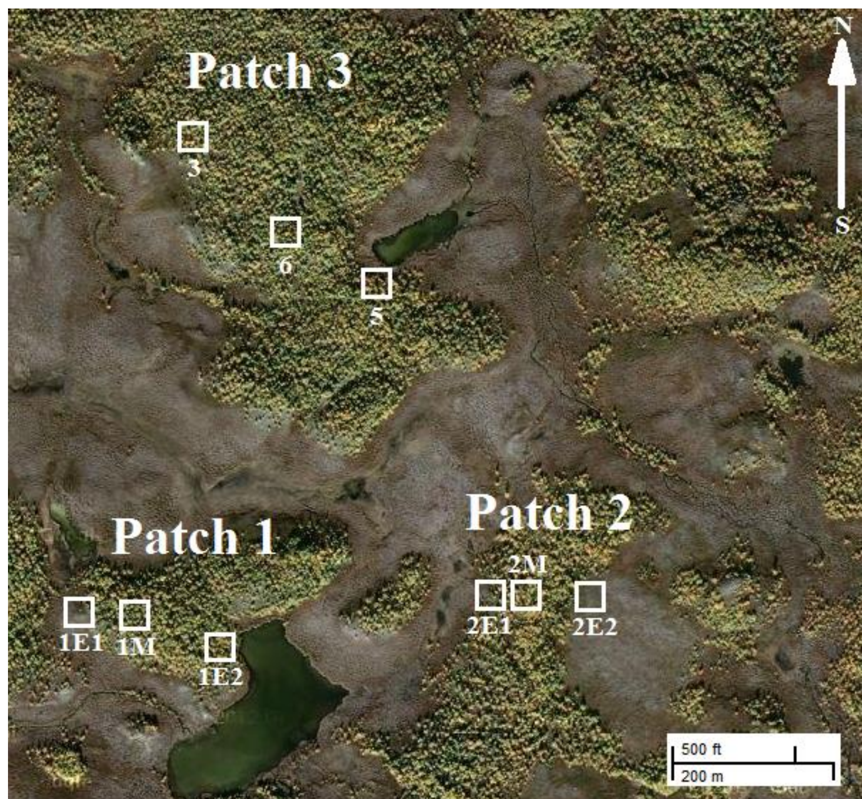

Fig. 2.1.1.1 Location of forest patches and sites sampled. Each site showing prefix code for each sample from that site. (1E1) Patch one windward site. (1M) Patch one central site. (1E2) Patch one leeward site. (2E1) Patch two windward site. (2M) Patch two central site. (2E2) Patch two leeward site. (3) Patch three windward site. (6) Patch three central site. (5) Patch three leeward site

Patch three was sampled by R. Baxter and R. Oakenfull (Durham University, UK) on 18/8/12. Three plots within this patch were sampled: a leeward site plot 5, a central site plot six and a windward site plot three (Fig 2.1.1.1) Leaves from surrounding *Betula nana* type plants were also collected from plots five and three. Saplings just windward of plot three were also sampled.

Samples from one and two were collected by R. Oakenfull (Durham University, UK) on 4/7/12 and 5/7/12 respectively. Again three plots within each new patch were sampled; a leeward site plot 1E2 and 2E2, a central site plot 1M and 2M and a windward site plot 1E1 and 2E1 (Fig 2.1.1.1). Leaves from surrounding *Betula nana* type plants were also collected from plots 1E2, 2E2, 1E1 and 2E1.

2.1.2 Buds

Bud samples were collected by R. Holden (Durham University, UK) on 12/10/11. Six morphologically *Betula nana* type plants and six morphologically *Betula pubescens subsp. czerapanovii* type plants along a gradient between GPS co-ordinates: N68^{o}21’19” E18^{o}48’56” and N68^{o}21’28” E18^{o}48’59” were sampled.

2.1.3 Seeds

Seeds were collected in the second and third week of September 2011 by R. Holden (Durham University, UK) from *Betula pubescens subsp. czerapanovii* type trees near patch three in Abisko, Northern Sweden.

2.1.4 Transport and Storage of Samples

All leaf and bud samples were kept chilled until return to the U.K. where they were stored at -80°C for later DNA analysis.

2.2 Growth of Test Plants

2.2.1 Germination

Seeds were cold stratified at 4°C for 16 days before planting on 16/12/11 and 41 days before planting on 10/1/12. 2000 seeds were sown in groups of 200 seeds onto labelled trays (A-J) on 16/12/11 and another 1000 seeds in groups of 200 (trays K-O) were planted on 10/1/12. Seeds sown in trays A-T were sown on Tray-Substrat white and black sphagnum peat and trays S & T were sown on J Arthur Bowers' traditional potting compost containing peat with added nutrients and wetting agent. Seeds were evenly spaced on the surface of their respective compost and the trays were then covered with plastic bags to keep humidity high. Trays were then placed in a growth chamber (Weiss Gallenkamp, Loughborough, UK) set at 18°C, 85% humidity, 400 μmol $m^{-2}s^{-1}$ light (with a one hour sunrise and one hour sunset of 300μmol $m^{-2}s^{-1}$), 20 h days and with additional tungsten (providing additional lighting in the red part of the spectrum) during set daylight hours. Once the first true leaf was showing seedlings were pricked out into individual pots and potted on when needed.

2.2.2 Tracking Individuals

Each individual was given a code (relating to seed tray in which they germinated and the order in which they were pricked out from that tray). A database was constructed on Microsoft Office Excel 2010 into which planting dates, potting up dates, pot sizes, mycorrhizal inoculation dates, mycorrhizal collection date and other notes (such as treatments, deaths etc were entered).

2.2.3 Light Response

68 days after the initial seed sowing date, four individuals (which had been potted up at least once) were selected randomly from the initial batch of seed sown. From each of these individuals, light response curves of photosynthesis were constructed for the youngest fully expanded leaf and oldest non-necrotic leaf using infra red gas analysis (LiCOR 6400, LiCOR, Lincoln, Nebraska, USA).

2.2.4 Growth Conditions

After pricking out plants were transferred to two growth chambers (Weiss Gallenkamp, Loughborough, UK) set at 16°C, 75% humidity, initially 400 µmol m^{-2} s^{-1} light with a one hour sunrise and one hour sunset of 300µmol $m^{-2}s^{-1}$), 20h day length and with supplemental red light from tungsten lamps during programmed daylight hours. Following results of light saturation determination, derived from light response curves, light levels were raised to 600µmol $m^{-2}s^{-1}$, and all other conditions remained the same.

2.2.5 Mycorrhizal Inoculations

Soil from a field located in Durham, UK was collected on several occasions to provide mycorrhizal inoculations for the germinated *Betula* seeds. This field was assumed to have a high density of mycorrhizal hyphae and spores due to the recent natural colonisation of *Betula sp.* saplings and the high density of Dactylorhiza purpurella X fuchsia, both of which benefit from or require mycorrhizal symbiosis. The samples were collected from around the roots of vigorously growing Betula sp. saplings to increase chances of a good mix of Mycorrhizal hyphae and spores.

Soil samples were thoroughly suspended in water and allowed to settle for several hours. 10ml of the supernatant was added to recently pricked-out seedlings and then repeated approximately one week later with another fresh suspension of soil sample.

2.3 Freezing Experiments

2.3.1 Selection of Individuals

Individuals to be tested were selected with the use of a random number generator. Fifteen individuals from all trays bar S and T were selected for each treatment type (-10°C shock, one week at 3°C prior to -10°C shock or two weeks at 3°C prior to -10°C shock).

2.3.2. Chamber Conditions

Test subjects were placed into boxes of polystyrene packed with vermiculite so that the top of the polystyrene box and the level of vermiculite was level with the top of the pot. Test subjects were then moved from ambient chambers and placed in a growth chamber. For three days they experienced the same conditions as their ambient counterparts. The temperature was then dropped from 16°C to -10°C and humidity from 75% to 0% over a period of two hours for the sudden shock samples or from 16°C to 3°C and 75% to 20% humidity over a period of two hours for the samples to undergo acclimation. Acclimation samples remained at 3°C for periods of one or two weeks and then dropped to -10°C and 0% humidity over the space of two hours. All samples remained at -10°C for 24 hours before returning to 16°C over the space of 13 hours. All other growing conditions remained the same.

2.3.3 Morphology Data

For final study samples, before experimentation, the following measurements were made of each individual tested: The height of the sapling, the number, width and length of branches and the height at which branches diverged from the main stem. Heights were measured using a measuring tape and widths and lengths were measured using digital calipers. After the experimental period samples were placed back into a growth chamber under normal holding environmental conditions.

2.3.4 Survival Tests

Experimental plants were then checked for evidence of bud burst on 2/11/12 by R. Oakenfull (Durham University).

2.4 Molecular Analysis

2.4.1 Buffers and Solutions

Chemicals supplied by Sigma-Aldritch Ltd. (Dorset, UK)

2.4.1.1 Edwards buffer

for 5 ml buffer:

1ml of 1M Tris-HCl pH 7.5,

1.25ml of 1M NaCl,

0.25ml of 0.5M EDTA pH 8.0,

0.25 ml of 10% SDS,

2.25ml of H_2O.

2.4.1.2 1M TE buffer

For 10ml:

100µl 10mM Tris-HCl PH8,

20µl 1mM EDTA pH8,

9.8ml H_2O.

2.4.1.3 Dellaporta Extraction Buffer

For 10ml:

0.5ml, 1M tris pH 8,

0.2ml 0.5M EDTA, pH8

1ml, 1M NaCl,

1ml, 10% SDS,

7ul, 14.3M β-Mercaptoethanol,

7.3ml, sdH_2O.

2.4.1.4 TBE Buffer

For 1L of 5X TBE stock:

54g Tris base,

27.5g Boric acid,

20ml 0.5M pH8 EDTA,

Make up to 1L with H_2O.

2.4.1.5 TAE Buffer

For 500ml of 50X stock:

121g Tris base,

28.55ml Glacial acetic acid,

50ml EDTA (ph8 0.5M),

400ml H_2O (milliQ).

Mix once dissolved to make 500ml with H_2O

2.4.1.6 CTAB buffer

For 50ml:

1g CTAB,

14.2ml 5M NaCl,

2ml 0.5M EDTA,

5ml 1M Tris HCl pH8,

1g PVP 40,

0.04g Ascorbic acid ,

0.045g DIECA.

Made up to 50ml with H_2O.

2.4.2 Primers

Primers were ordered from FisherScientific (Loughborough, Leicestershire, UK). Primers were supplied dehydrated and were then rehydrated in H_2O to form a 50M stock.

2.4.2.1 Primer sequences

(Designed by M. Knight (Durham University)) Primers based upon sections of sequence from Betula pubescens subsp. tortuosa clone Bo.4 microsatellite sequence (Genbank accession number: AY423612) (Bp1), Betula pubescens subsp. tortuosa clone Bo.F394 microsatellite sequence (Genbank accession number: AY423608) (Bp2) and Betula nana partial ADH gene for putative alcohol dehydrogenase, exons 2-6 (Genbank accession number: AJ535649) (Bn1 & Bn2)

Bp1 F: AGCCATTTAAAACACGTCACC

Bp1 R: TATGGGGCATGATGTTGATG

Bp2 F: TCATTTCTGCTCGTTCATGC

Bp2 R: CACGTGGCTCCTACTTCCTC

Bn1 F: ATACCGCTCTGTGCCATACC

Bn1 R: CCTCATGGCCAAGGATTCTA

Bn2 F: TAGAATCCTTGGCCCATGAGG

Bn2 R: TTTCGGTTTTGCAACATTCA

(From Järvinen et al, 2004) to amplify partial ADH sequences

ADH fwd: GCACCACCACAAGTAGGTGAAG

ADH rev: AATCTTGAAGCCCCAGCAATCC

2.4.2.2 Primers for RAPD analysis

(From Howland, Oliver and Davy, 1995)

ATC33: CGGTAGCCGC

ATC71: GTCGTTCGGG

ATC51: ATGTGGCGAC

(From Dharmar and John De Britto, 2011)

OPB17: AGGGAACGAG

OPB14: TCCGCTCTGG

OPB11: GTAGACCCGT

OPB12: CCTTGACGCA

OPB18: GGAGGGTGTT

(From Arif et al., 2010)

RAPDa: GGTGCGGGAA

RAPDb: GTTTCGCTCC

RAPDc: GTAGACCCGT

RAPDd: AACGCGCAAC

RAPDe: CCCGTCAGCA

2.4.3 DNA Extraction

2.4.3.1 Edwards

(From Edwards, Johnstone and Thompson, 1991)

Samples frozen in liquid nitrogen were ground for 5-10s with a plastic micropestle in a 1.5ml eppendorf tube. 400µl of Edwards extraction buffer was then added to the eppendorf and vortexed using a Labnet vortex mixer VX100 for 5 seconds. The samples were then spun down in a microcentrifuge (eppendorf AG centrifuge 5415D) at full speed for one minute. 300µl of the supernatant was then transferred to a fresh 1.5ml eppendorf and 300µl of isopropanol mixed in and left at room temperature for 2 minutes. The samples were then spun at full speed in a microcentrifuge (eppendorf AG centrifuge 5415D) for 5 minutes and the supernatant was aspirated and the remaining sample was respun for one minute and re-aspirated. The remaining pellet was then spun to dryness using a spin vacuum desiccator (Eppendorf concentrator 5301) and then left with the eppendorf lid open on paper for

~30 minutes to allow further drying. The pellet was then resuspended in 50 µl TE a left to dissolve at 4°C overnight.

2.4.3.2 Dellaporta

(Modified from Dellaporta, Wood and Hicks, 1983)

Samples were ground using liquid nitrogen cooled mortar and pestles then added to a 1.5ml eppendorf. 750µl of Delaporta extraction buffer was added and samples were heated at 65°C using an Labnet Accublock digital dry bath for 10 minutes. 200µl of 5M potassium acetate was then added and vortexed using a Labnet vortex mixer VX100 for 5 seconds before incubating on ice for 20 minutes. Samples were then spun down in a microcentrifuge (eppendorf AG centrifuge 5415D) for 10 minutes at 12,000 RPM and the supernatant pipetted into a fresh eppendorf. Supernatant volume was estimated using a pipette and an equal amount of isopropanol was added to the supernatant, mixed and then centrifuged and 12,000RPM for 10 minutes. The isopropanol was then removed from the resulting pellet. 80% ice cold ethanol was then added to the pellet, spun for 3 minutes at 12000RPM and removed from the pellet. The pellet was then left to air dry for 20 minutes at room temperature. The pellet was then slowly rehydrated on ice for 30 minutes in 50µl TE.

2.4.3.3 Phenol-Chloroform purification of DNA.

(Modified from Sambrook and Russel, 2001)

Solutions of DNA extracted by the Dellaporta method were made up to 400µl with TE buffer. 400µl of Phenol-Chloroform was then added and mixed using a Labnet

vortex mixer VX100. The solution was then spun down at 13000RPM (using a eppendorf AG centrifuge 5415D) for 5 minutes, the top aqueous layer was then removed and transferred to a new 1.5ml eppendorf. Spinning at 13000RPM and removal of the top aqueous layer was repeated until no upper solid layer remained. The volume of final upper aqueous layer was estimated using a micropipette and $1/10^{th}$ of the volume of the upper aqueous layer of 3M pH5.2 sodium acetate and 2X the volume of the upper aqueous layer of 100% ethanol were added and mixed. The solution was incubated at -20°C for 2 hours and then spun down at 13000RPM for 10 minutes and the supernatant discarded. The remaining pellet was then washed in 1ml of 80% ethanol and spun down at 13000RPM for 5 minutes. The supernatant was then removed and the pellet spun to dryness using an Eppendorf concentrator 5301. The pellet was then slowly rehydrated on ice for 30 minutes in 50µl TE.

2.4.3.4 CTAB

(Modified from Doylle and Doyle, 1987 and Strange, Prehn and Arce-Johnson, 1998)

~100mg of tissue was ground in a liquid nitrogen cooled pestle and mortar. 800µl of CTAB buffer at 60°C was then added to the ground leaf and the resulting mix transferred to a 1.5ml eppendorf tube. 3µl of β-mercaptoethanol was added and mixed by inversion before heating the mixture at 60°C for 30 minutes. 500µl of chloroform isoamyl-alcohol (24:1) was added and the mixture was spun at 13000RPM in a eppendorf AG centrifuge 5415D for 5 minutes and the resulting upper aqueous layer transferred to a new 1.5ml eppendorf. The addition of 500µl of chloroform isoamyl-alcohol to the removed upper aqueous layer followed by

spinning at 13000RPM for 5 minutes was repeated until the aqueous layer removed was clean and no longer cloudy. 500µl of -20°C isopropanol was then added, mixed by inversion and incubated at -20°C for 30 minutes before spinning at 13000RPM for 5 minutes. The supernatant was discarded and the pellet air dried in hot block (Labnet Accublock digital dry bath) at 37°C. The pellet was then resuspended in 100µl TE buffer, RNAse was added to concentration of 10µg/µl and the mixture was incubated at 37°C for 30 minutes. 500µl isopropanol was then added and the mixture allowed to precipitate at -20°C for 30 minutes. The mixture was then spun at 13000RPM for 10 minutes and the resulting supernatant discarded. The pellet was then washed in 800µl 70% ethanol for 20 minutes and then spun down at 13000RPM 5 minutes and the resulting supernatant discarded. The pellet was then air dried at 37°C for ~30 minutes then resuspended in 100µl TE buffer and leave to dissolve overnight at 4°C

2.4.3.5 Powerplant® Pro DNA Isolation.

(Modified from protocol supplied with PowerPlant ® Pro DNA Isolation Kit from Mo Bio Laboratories Inc.)

To labelled bead beating tubes 410µl of PD1 and 40µl of Phenolic separation solution were added. Samples were ground in a liquid nitrogen chilled mortar and pestle and added to the labelled bead beating tubes (to a maximum amount of ground tissue of ~1cm^{3}). 50µl of PD2 and 3µl of RNase A solution were added to each tube and each tube was briefly vortexed on a Labnet vortex mixer VX100. Samples were then heated at 65°C for 10 minutes then attached to a flat bed vortex (Labnet VX100) and shook at maximum speed for 10 minutes. Tubes were then spun at 13000g in a

eppendorf AG centrifuge 5415D for 2 minutes and the resulting supernatant removed to 1.5ml eppendorf tubes. To the supernatant 250µl of PD3 was added, vortexed for 5 seconds and then incubated at 4°C for 5 minutes. After incubation the samples were spun at 13000g for 2 minutes and a maximum of 600µl supernatant, from each sample, was removed to new 1.5ml eppendorf tubes and 600µl of PD4 and 600µl of PD6 were added and mixed on a vortex for 5 seconds. ~600µl of the mix was loaded onto a spin filter and spun at 10,000g for 30 seconds, flow through was discarded and ~600µl more of the mix added to the spin filter and spun until all the mix was used. 500µl of PD5 was then added to each spin filter and they were spun at 10,000g for 30 seconds and flow through discarded. Then 500µl of PD6 was added to each spin filter and they were spun at 10,000g for 30 seconds and the flow through discarded. Filters were spun again at 16,000g for 2 minutes, to remove residual PD6 and then placed into new collection tubes. 100µl of PD7 was then added to the centre of each spin filter membrane and the spin filters incubated at room temperature for 2 minutes. After incubation spin filters were spun at 10,000g for 30 seconds the flow through was then reloaded and respun. 2 µl of the flow through, mixed with 6 µl H_2O and 2 µl dye, was then run on a 8% agarose gel to check for DNA presence. DNA was then stored at -20°C

2.4.3.6 Serial Dilutions

Diluted samples were diluted with PCR H_2O to produce a series of eight 2X serial dilutions.

2.4.4 PCR Conditions

2.4.4.1 PCR Mix One

38.5 µl, PCR H_2O

5ul, Bioline 10X buffer

1.5, 50µM $MgCl_2$

1ul, 50µM fwd primer

1ul, 50µM rev primer

1uL, 10µM DNTPs

1ul, DNA

1ul, BioTaq RED Polymerase (Bioline, London, UK)

2.4.4.2 PCR Mix Two - KAPA3G.

20.6µl, PCR H_2O

25µl, KAPA plant PCR Buffer 2X (KAPA Biosystems Inc. Woburn, USA)

1.5µl, 10µM fwd primer

1µl, 10µM rev primer

1µl, DNA

0.4 µl, KAPA3G plant DNA polymerase (2.5U/µl) (KAPA Biosystems Inc. Woburn, USA)

2.4.4.3 PCR Mix Three -KAPA3G, for RAPDs.

22.1µl, PCR H_2O

25µl, KAPA plant PCR Buffer 2X (KAPA Biosystems Inc. Woburn, USA)

1.5µl, 10µM Primer

1µl, DNA

0.4 µl, KAPA3G plant DNA polymerase (2.5U/µl) (KAPA Biosystems Inc. Woburn, USA)

2.4.4.4 PCR Program One

X1:

5 mins 95°C

5 mins 55°C

5 mins 72°C

X40:

1 min 95°C

1 min 55°C

1 min 72°C

X1:

15 mins 72°C

2.4.4.5 PCR Two (KAPA 3G)

X1:

5 mins 95°C

X40:

30 secs 95°C

15 secs 55°C

45 secs 72°C

X1:

45 secs 72°C

2.4.4.6 PCR Three (KAPA 3G) modified:

1X:

10 mins 95°C

X40:

30 secs 95°C

15 secs 55°C

45 secs 72°C

X1:

45 secs 72°C

2.4.4.7 PCR Four (KAPA 3G) for RAPDs:

1X:

10 mins 95°C

X45:

30 secs 95°C

15 secs 35°C

45 secs 72°C

X1:

45 secs 72°C

2.4.5 Gels

For testing presence of DNA to each gel well a mixture of 2µl DNA 6µl Water and 2µl dye was added.

For testing PCR products using non dye containing enzymes a mixture of 8 µl DNA and 2 µl dye was added to each well.

2.4.5.1 TBE Gels

0.5g agarose (for PCR products) or 0.4g agarose (for DNA extractions) was dissolved in 500ml 0.5X TBE in a 500ml conical flask using a Kenwood 850W microwave on full power. Once cool 2.5µl 10mM ethidium bromide was mixed into the solution and then poured into a 8cm long gel tank and left to set. Gels were run on a Consort EV243 powerpack at 100V and 35mAmps for 1 hr.

2.4.5.2 TAE Gels

1% TAE gels were used for RAPD products. 1.2g of agarose was added to 120ml 1X TAE buffer in a 500ml conical flask and dissolved using a Kenwood 850W microwave on full power. Once cool 5µl 10mM ethidium bromide was mixed into the solution and then poured into a 15.5cm long gel tank and left to set. Gels were run on a Consort EV243 powerpack at 100V and 40mAmps for 6 hours.

2.4.5.3 Imaging Gels

Gels were imaged using a Fotodyne Inc. UV transilluminator.

2.4.6 Gel Band Extraction

Modified from the QIAquick Gel Extraction Kit Protocol for PCR products to be sequenced (Qiagen, Sussex, UK).

Bands for extracting were run on 8% agarose gels. After suffient time to allow easy distiguation between the bands to be cut out, ~3 hours, gels were dabbed dry with tissue to remove excess TBE. Bands were cut out on a Fotodyne UV transilluminator using scalpel blades. The weights of each gel were calculated by deducting the weight of the 1.5ml eppendorf tube before the gel slice was placed in from the weight of the 1.5ml eppendorf with the gel slice inside.

The bands were extracted with a QIAquick Gel Extraction kit using a modified version of the protocol. 3 1/2 volumes of buffer QG were added to one volume of Gel (100mg ~ 100 µl). The gel slices were placed on a Stuart roller mixer SRT6 at room temperature for 15 minutes until the gel had dissolved. Add one gel volume of ice cold IPA was added and vortexed using an eppendorf centrifuge 5415D for 5 seconds. The sample was then applied to a Promega (Wisconsin, USA) spin column and spun at full speed on the centrifuge for 30 seconds. The flow through was then reapplied and respun at full speed for 30 seconds. Flow through was discarded and 500µl of QG buffer added to column and spun at full speed for 1 minute. Flow through was discarded and 750µl of buffer PE was then added to the column, stood for 2-5 minutes then spun at 13000RPM for 1 minute. Flow through was again discarded and the column spun at full speed for an additional 2 minutes. The column was placed in a clean 1.5ml eppendorf and DNA and 50µl of warm PCR H_2O applied to the centre of the column membrane. The column and eppendorf were left to stand for 2 minutes, then spun at 13000RPM for 1 minute. DNA concentration of the flow through was then calculated by running 2µl of flow through 6µl H_2O and 2 µl dye on a 8% gel and calculating concentration from ladder relative to band brightness. If necessary elutions were then dried down to at least 25ng/µl using a

spin vacuum desiccator (eppendorf concentrator 5301) in order to be sent for sequencing.

2.4.7 Sequencing

DNA was sent to Durham University Sequencing Department for Sequenceing.

2.4.8 Comparison ADH Sequences

The following partial ADH sequences (Järvinen et al. 2004) were used for comparison with the partial ADH sequences from Abisko:

Betula pubescens partial ADH gene for putative alcohol dehydrogenase, exons 2-6, (short allele) GenBank accession number: AJ535645.1.

Betula pubescens partial ADH gene for putative alcohol dehydrogenase, exons 2-6, (long allele) GenBank accession number: AJ535646.1

Betula nana partial ADH gene for putative alcohol dehydrogenase, exons 2-6 GenBank accession number: AJ535649.1

Corylus avellana partial ADH gene for putative alcohol dehydrogenase, exons 2-6 GenBank accession number: AJ535656.1

2.4.9 RAPDs

PCR products from RAPD primers were run on an Agilent 2100 bioanalyzer, as per the Agilent DNA 7500 and DNA 12000 Assay Protocol. Using the DNA 7500 kit and settings.

2.5 Analysis of Results

2.5.1 ADH Sequences

Partial ADH sequences were visually checked in Chromas LITE version 2.1 for sequence read out accuracy, altered if necessary and exported as a FASTA file. The section of each exported FASTA file that showed clear peaks in Chromas LITE was then imported into BioEdit Sequence Alignment Editor version 7.1.7 (Hall, 1999). Sequences were then manually aligned by eye in BioEdit and where forward and reverse sequences for a particular sample overlapped sufficiently they were combined. Published partial ADH sequences for *B. nana*, *B. pubescens* (long) and *B. pubescens* (short) were imported from GenBank to BioEdit for comparison with the partial ADH sequences from samples from Abisko.

Phylogenetic trees of partial ADH sequences were produced using the BioEdit accessory application, "DNA distance nearest neighbour". The matching section of *Corylus avellana* partial ADH gene was imported from GenBank to be used as the outgroup. Since the length of forward and reverse sequences for each sample varied, aligned sequences were cut to where the majority of samples had data for a section of forward and reverse sequence, and phylogenetic trees produced. A tree was also produced for those samples where forward and reverse sequences had been joined. BioEdit graphic view was used to produce images of the aligned sequences which were then edited in Microsoft Windows paint version 6.1.

2.5.2 RAPDs

Overlays of repeat samples were produced to check accuracy of peak locations then for each patch and primer, overlays of sample electropherograms were produced using Agilent bioanalyzer 2100 expert. This image was inspected visually for "peak groups". Groups were based on obvious groupings that appeared upon overlay of all electropherograms for that patch and on accuracy of peak locations, i.e. two apparent groups that were sufficiently close to run the risk of being the same peak due to inaccuracy in peak locations were grouped into one group. Each peak group was then given a label. Peak groups were named according to the primer used and numbered by size. For peaks groups that could be split into smaller groups in different patches these split peaks were labelled by their bigger peak group name but given a separate final letter (a, b, c etc). Electropherograms for each sample for each RAPD were then visually inspected for any sign of a peak in each peak group and the number of peaks that appeared in each peak group recorded. This data was then converted to binary data (1= present 0= absent) for analysis with SPSS Statistics 17.0.

Dice's coefficient of similarity was used to judge the similarity of individuals run in SPSS Statistics 17.0. From this index, dendrograms of nearest neighbour were created in SPSS Statistics 17.0 and modified for ease of reading using Microsoft Paint version 6.1 Patches were analysed both separately and all together. Due to peak location inaccuracy changing in different patches, for the analysis of all patches together peak grouping were brought down to the level of the most inaccurate patch. For example two separate peak groups in one patch may not be considered separate in another patch due to peak location inaccuracy and hence the patch with two

separate peaks groups would have the results for those peak groups merged and labelled as in the more inaccurate patch.

Chapter Three: Results and Discussion

3.1 DNA Extraction

3.1.1 Background

In order for studies on the genetic identity of individuals to be undertaken, genomic DNA had to be isolated from the bud and leaf samples. Molecular studies on *Betula* species have not previously been undertaken at Durham University and literature on DNA extraction methods for these particular species of *Betula* is limited. The main methods used are modifications on the CTAB method: (E.g. Järvinen et al. 2004, Palme, Palsson & Lascoux, 2004 and Maliouchenko et al. 2007). The CTAB method is a particularly time consuming method, not suitable for a high throughput of DNA extractions. Although other methods have been used for this species they are similarly time consuming (E.g. Welling et al, 2004). Even using the CTAB method the use of expensive DNA kits was sometimes required to extract DNA from all samples (Järvinen et al. 2004). Since no comparative study of various DNA extraction techniques had been conducted on these species the most cost and time effective method of extraction was investigated.

3.1.2 Edwards Extraction

The initial DNA extraction used Edwards' method, one of the simplest and quickest DNA extraction techniques (Edwards, Johnstone & Thompson, 1991). The Edwards method of DNA extraction of samples BIR1 and DW1, did not produce bands when extracts were run on a gel (data not shown) suggesting that DNA extraction was unsuccessful. Since it is possible that low yields of DNA may not form obvious bands on gels, these DNA extracts were used for PCR amplification with BP1, BP2, BN1 and BN2 primer pairs (Fig. 3.1.2.1.). DNA from sample DW1 with primer pairs BP2 and BN1 shows some signs of banding (Fig. 3.1.2.1. (4 & 6)) however these bands are not well defined. There were no sign of bands for DW1 DNA with primer pairs BP1 or BN2 or for BIR1 with any primer pairs and all samples show primer dimerization (Fig 3.1.2.1).

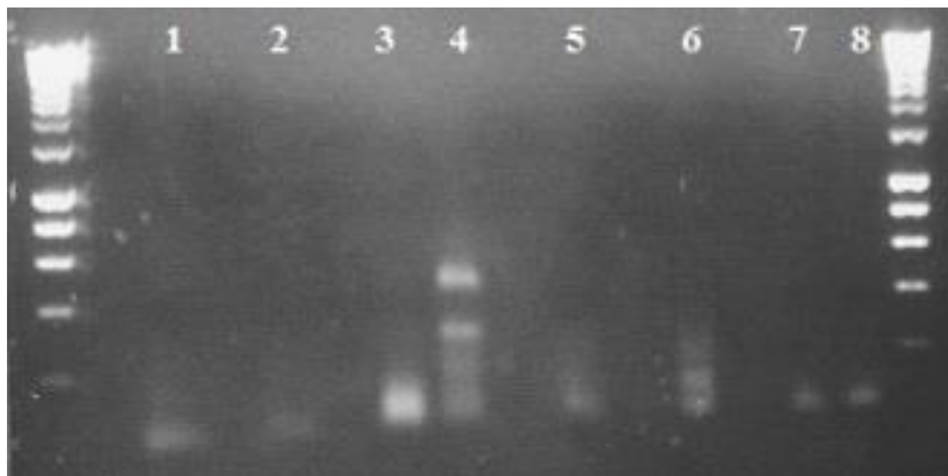

Fig. 3.1.2.1. TBE Gel image of PCR products (PCR mix1 and PCR program 1) using DNA from Edwards DNA extraction (fig X). (1) Bir1 DNA with primer pair BP1. (2) DW1 DNA with primer pair BP1 showing a weak banding pattern. (3) Bir1 DNA with primer pair BP2. (4) DW1 DNA with primer pair BP2. (5) Bir1 DNA with primer pair BN1. (6) DW1 DNA with primer pair BN1 showing a weak banding pattern. (7) Bir1 DNA with primer pair BN2 (8) DW1 DNA with primer pair BN2.

If DNA from sample DW1 was successfully extracted and PCR was not inhibited, bands for the other primer pairs would be expected. Potential explanations for the appearance of some bands but not others include; DNA extraction was limited, primers are not fully functional, the sample was contaminated, or that PCR was inhibited and primers BP2 and BN1 are better at overcoming these problems than the other primers. If DNA extraction was limited there is no obvious reason as to why bands should be formed by BP2 and BN1 but not by other primers used on the same DNA extraction. The band formed by BP2 with DW1 although not well defined is bright and hence it seems unlikely that other bands are present but extremely faint. Another explanation is that the other primers are not fully functional. All primers show strong primer dimers (Fig. 3.1.2.1.) so this is possible, however later studies (Fig. 3.1.7.1) proved that all primers were functional. Contamination is therefore a likely explanation. Comparison with later studies using the same primers (Fig. 3.1.7.1) shows that the potential band for primer pair BN1 with DW1 (Fig. 3.1.2.1. (6)) is of too low a molecular weight to be correct (Fig. 3.1.7.1 (3 & 7)). It is therefore more likely that the band is due to further primer dimerization or a low molecular weight contaminant. The band produced with BP2 primers (Fig. 3.1.2.1 (4)) however matches the band produced with BP2 primers for DNA of 3.12G (Fig. 3.1.7.1 (2)), but not that of DNA from N5.3 with BP2 primers (Fig. 3.1.7.1 (7)). DW1 being *B. nana* in form would be expected to show a banding pattern more similar to N5.3 (a *B. nana* form) than 3.12G (a *B. pubescens* ssp. *czerapanovii* form) but the capability of these species to hybridise must be considered (e.g. Elkington, 1968, Anamthawat-Jónsson & Tomasson, 1990, Anamthawat-Jónsson et al., 2010, Thórsson, Salmela & Anamthawat-Jónsson, 2001 and Anamthawat-Jónsson et al., 2003). Hence a banding pattern more similar to a *B. pubescens* displayed in a *B.*

nana type shrub is wholly possible. Hence it is debatable whether this is contamination from another source such as other species commonly used in the lab e.g. *Arabidopsis thaliana* or *Sorghum* sp.. Although not impossible, it seems unlikely that such distantly related plants as *A. thaliana* or *Sorghum spp*. would co-incidentally produce the same banding pattern as two *Betula sp*, when another sample from the same area of the same species did not. The remaining explanation is that PCR inhibition was overcome in that particular sample. There is no obvious reason as to why primer pair BP2 would be more resistant than others, later studies showed that this is unlikely since in other situations such as (Fig. 3.1.6.2 (4)) only the BN2 primer produced a band. It is also unlikely that PCR inhibition can explain the lack of some bands and presence of others considering that all DNA from DW1 came from the same stock and hence should have similar levels of inhibitors in all PCR tubes. However due to small fluctuations in pipetting technique and sample homogeneity it is possible that by pure luck in that particular PCR tube inhibitors were sufficiently low to not inhibit PCR. This may seem unlikely since later methods with greater inhibitor removal did not produce results (see Dellaporta and phenol-chloroform purification sections) however it is possible that this technique does not release as many inhibitors in the first instance. In order to address the cause of this DW1 DNA extracted using a more reliable method and run with primer pair BP2 and compared to other species used in the lab with primer pair BP2 would have to be run. However it is evident that this method is unreliable since PCR with Bir1 DNA (Fig. 3.1.2.1. (1, 3, 5 & 7)) did not show any bands at all. Regardless of whether DNA extraction was partially successful with one sample and primer set, the method does not produce reliable high quality results and hence more intensive techniques were assessed.

3.1.3 Dellaporta Extraction

Due to the unreliability of Edwards extraction for the *Betula* samples, a modification of the Dellaporta (Dellaporta, Wood & Hicks, 1983) method of DNA extraction was used. This method has been reported to be effective at extracting DNA from a range of species (Dellaporta, Wood & Hicks, 1983). The addition of β-mercaptoethanol, a reductant, (not present in Edwards extraction buffer) is used to reduce oxidation of polyphenolics which once oxidised covalently bond with the proteins and nucleic acids in the samples inhibiting PCR (Paterson, Brubaker & Wendel, 1993, Couch & Fritz, 1990, Maltas, Vural & Yildiz, 2011, Guillemaut & Mardchal-Drouard, 1992). Phenolics will be present in high concentrations in these samples (Haukioja, Niemelä & Sirén, 1985, Nurmi et al. 1996) and hence the addition of β-Mercaptoethanol should aid successful extraction, therefore Dellaporta extraction is more likely to be successful than Edwards extraction. No sign of banding is seen in DNA extracts of BIR3 and DW3 using this method of extraction and PCR amplified with BP1, BP2, BN1 and BN2 primer pairs (data not shown.) indicating that DNA extraction was unsuccessful. Extracts were deep brown in colouration, which indicated the presence of oxidised phenolics (Taylor & Clydesdale, 1987) which would inhibit PCR reactions and a deep brown colour of extracts is known to indicate that DNA is unusable (Guillemaut, 1992). Despite the addition of β-Mercaptoethanol to prevent oxidation of polyphenolics there is evidently some oxidation occurring and likely inhibiting PCR, so further removal of phenolics is necessary.

3.1.4 Phenol-Chloroform Purification of DNA

All DNA extractions up to this point had shown strong brown colouration suggesting high levels of oxidised phenolics (Taylor & Clydesdale, 1987 and Guillemaut, Mardchal-Drouard, 1992). Consequently, phenol-chloroform purification of DNA extracts was used to increase the removal of phenolics. Samples 3.3A and BIR3 were extracted using the Dellaporta extraction method and then purified with phenol-chloroform. Strong banding is evident (Fig. 3.1.4.1.), however these bands are of low molecular weight.

3.3A (Fig. 3.1.4.1.), (1)) shows a band of just low molecular weight and BIR3 shows a range of molecular weights with the predominant fragments being of low molecular weight (Fig. 3.1.4.1. (2)).

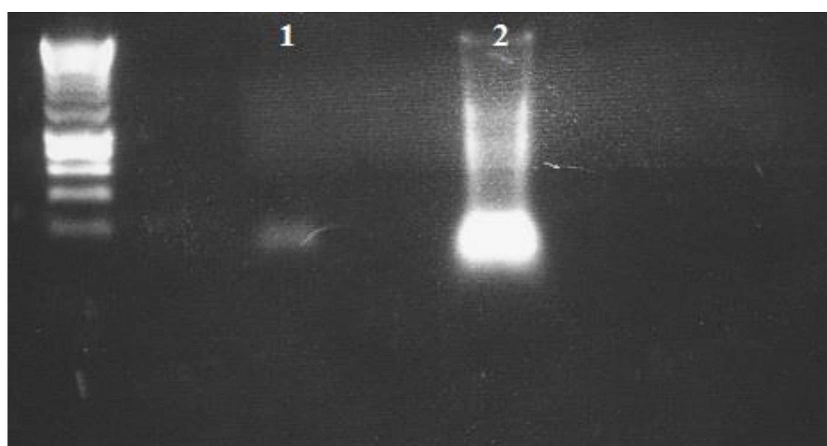

Fig. 3.1.4.1. TBE Gel image of phenol-chloroform purification of Dellaporta DNA extractions showing strong predominantly low molecular weight bands (1) 3.3A DNA (2) Bir3 DNA

Unfragmented genomic DNA should have a very large molecular weight. The most likely explanation is that since DNA extracts already exhibited high levels of oxidised phenolics as evident from the brown colouration of extracts (Taylor & Clydesdale, 1987) irreversible binding between oxidised phenolics and nucleic acids

(Couch & Fritz, 1990 and Maltas, Vural & Yildiz, 2011) had already occurred prior to the phenol-chloroform step and hence only fragments of unbound DNA are purified. Full genomic DNA or large fragments may still have been present at concentrations too low to show well on the gel (Fig. 3.1.4.1.) hence extracts were PCR amplified using primers BP1, BP2, BN1 and BN2 primer pairs. There were no signs of banding (data not shown) suggesting that either the PCR was still being inhibited by phenolics, that isolation of genomic DNA was not initially successful or, most likely given sample colouration, that oxidised phenolics had already bound to the majority of the DNA.

3.1.5 ADH Primers

Another potential problem could have been with the primers. The primers used were produced simply by looking at DNA sequences from *Betula* (Pers. Comms. Prof. M. Knight, Durham University). It is evident that these primers have a high level of dimerisation (e.g. Fig 3.1.2.1 & Fig. 3.1.5.2) and hence may be binding to themselves in preference to any DNA present. Järvinen et al., 2004 produced ADH primers that worked both with *Betula nana* and *Betula pubescens* with different banding patterns produced between the two species. Using the same ADH primers as Järvinen et al., 2004 there was still no sign of any banding produced with these primers using DNA extracted via the Edwards or Dellaporta methods of extraction (data not shown.), indicating that the previous primers were not the main problem. Serial dilutions were produced for Edwards and Dellaporta DNA extractions (Fig. 3.1.5.1.) and phenol-chloroform purified Dellaporta extractions (Fig. 3.1.5.2.), with the aim of reducing potential inhibiting compounds, such as phenolics, sufficiently to

reduce PCR inhibition whilst having a sufficiently high concentration of DNA for PCR. Neither set of dilutions showed any evidence of success (Fig. 3.1.5.1. & Fig. 3.1.5.2.) suggesting that either the extraction methods are unsuccessful or that inhibitory compounds are at very high levels.

Fig. 3.1.5.1. TBE gel image of ADH primer pair PCR products (PCR mix 1, Program 1) from serial dilutions of Edwards and Dellaporta extractions. (1-8) 2X serial dilution of edwards extraction Bir3 DNA 1/2 - 1/256 concentration. (9-16) 2X serial dilution of Dellaporta extraction 3.3A DNA 1/2 - 1/256 concentration.

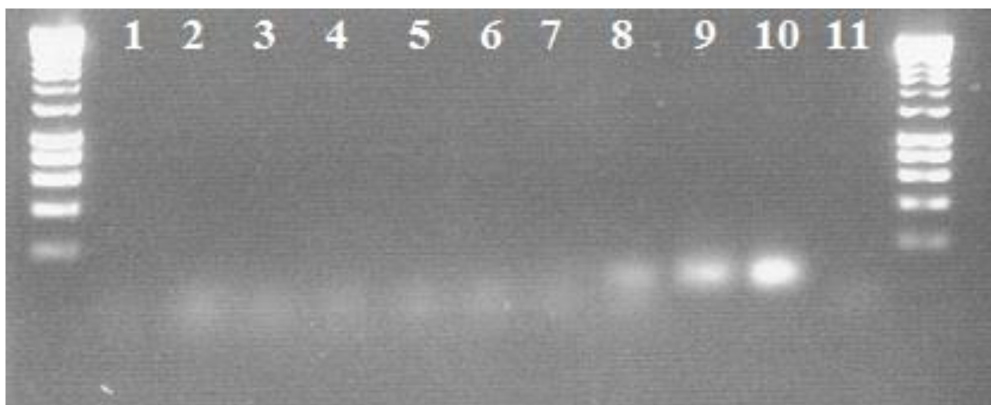

Fig. 3.1.5.2. TBE Gel image of ADH primer pair PCR products (PCR mix 1, Program 1) from Phenol-Chloroform purified Dellaporta DNA extractions. (1-9) 2X serial dilution of Bir3 DNA 1/2 - 1/512. (10) undiluted 3.3A DNA (11) undiluted Bir3 DNA

3.1.6 CTAB

This method of extraction and modifications thereof, is often used to overcome difficulties with high levels of phenolics in samples (Guillemaut & Mardchal-Drouard, 1992). Variations on the CTAB method have also been used for extraction

of *Betula* DNA (e.g. Järvinen et al., 2004, Palme, Palsson & Lascoux, 2004 and Maliouchenko et al., 2007). DNA extractions of 3.12G and N5.3 showed promising bands of high molecular weight when run on a gel (Fig. 3.1.6.1.).

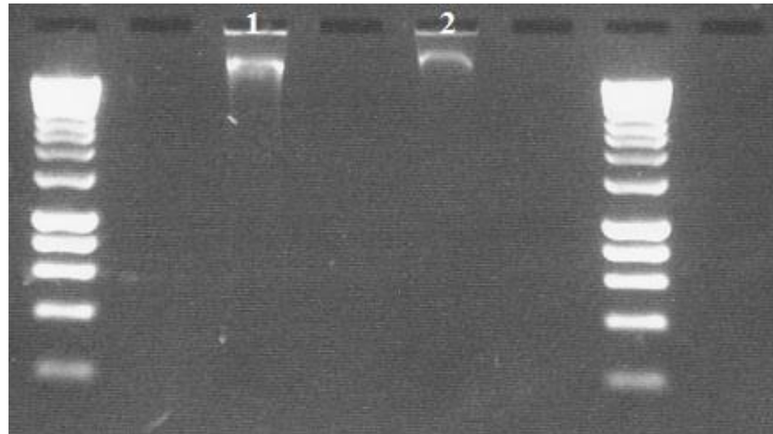

Fig. 3.1.6.1. TBE Gel image showing strong high molecular weight bands of CTAB DNA extractions of (1) 3.12G and (2) N5.3.

PCR amplification of these extractions with primer pairs BP1, BP2, BN1, BN2 and ADH produced one band for 3.12G with primer pairs BN2 (Fig. 3.1.6.2 (4)) but no bands for any of the other primer pairs (Fig. 3.1.6.2 (1-3 & 5-10).

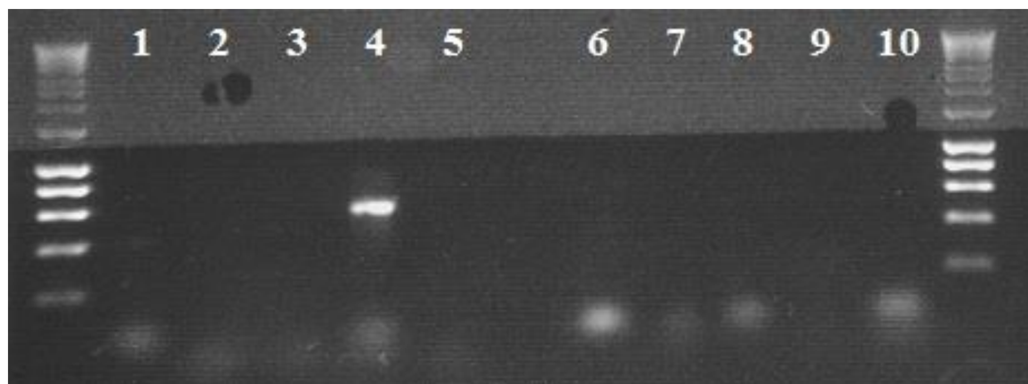

Fig. 3.1.6.2. TBE gel image PCR products (PCR mix 1, Program 1) from CTAB DNA extractions. (1) 3.12G with primer pair BP1. (2) 3.12G with primer pair BP2. (3) 3.12G with primer pair BN1. (4) 3.12G with primer pair BN2 showing a strong band. (5) 3.12G ADH primer pair. (6)N5.3 with primer pair BP1. (7) N5.3 with primer pair BP2. (8) N5.3 with primer pair BN1. (9) N5.3 with primer pair BN2. (10) N5.3 with ADH primer pair

Considering the strong promising bands shown in (Fig. 3.1.6.1.) it is likely that the PCR reaction is being inhibited, most likely by remaining phenolics since the extractions were still brown in colour indicating the presence of oxidised phenolics (Guillemaut & Mardchal-Drouard, 1992). Serial dilutions of CTAB-extracted 3.12g DNA were produced and PCR amplified with the ADH primer pair, with the aim of lowering the inhibitory compounds concentrations (Fig. 3.1.6.3.).

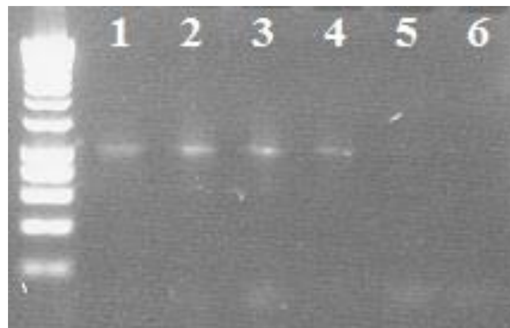

Fig. 3.1.6.3. TBE gel image of ADH primer pair PCR products (PCR mix 1, Program 1) from serial dilutions of CTAB DNA extracted 3.12G. 2X serial dilution (1-4) 1/2 - 1/16 showing bands. (5-6) 1/32 -1/64 concentration showing no evidence of bands.

Dilutions of 1/2 to 1/16 original concentration (Fig. 3.1.6.3 (1-4)) produced bands, indicating that although the level of inhibitory compounds has been lowered with the CTAB extraction, the DNA extracts are still high in inhibitors producing intermittent unreliable success based on the fluctuation of inhibitors relative to DNA concentration.

3.1.7 KAPA3G Enzyme Kit.

Kit from: (KAPA Biosystems Inc. Woburn, USA)

Since phenolics were strongly indicated as inhibiting PCR in CTAB DNA extracts a way of overcoming or removing the inhibitors was required. KAPA3G Plant PCR Kit claims to contain a novel DNA polymerase with improved tolerance to PCR inhibitors such as polyphenolics and polysaccharides and a plant PCR enhancer which claims to improve PCR performance (KAPA3G Plant PCR Kit, Technical Data Sheet, version 1.11). This new kit with the recommended PCR program (Chapter 2.4.4 PCR 2) using CTAB DNA extracts, 3.12G and N5.3 with primer pairs, BP1, BP2, BN1, BN2 and ADH produced clear well defined bands for all samples and primers (Fig. 3.1.7.1). Indicating that all primers were fully functional and the problems encountered with CTAB extracted samples lay in inhibition of PCR.

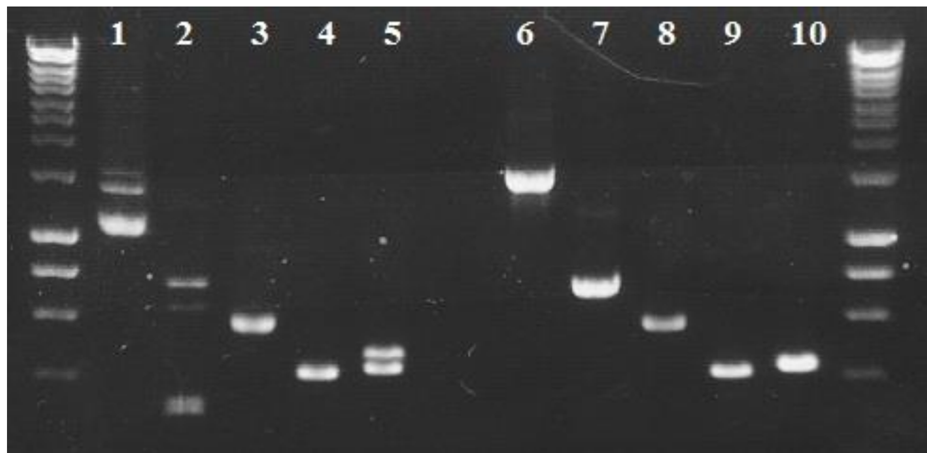

Fig. 3.1.7.1. TBE gel image of PCR products (PCR mix 2, Program 2) from CTAB DNA extractions all showing strong banding patterns. (1) 3.12G with ADH primer pair. (2) 3.12G with primer pair BN1. (3) 3.12G with primer pair BN2. (4) 3.12G with primer pair BP1. (5) 3.12G with primer pair BP2. (6) N5.3 with ADH primer pair. (7) N5.3 with primer pair BN1. (8) N5.3 with primer pair BN2. (9)N5.3 with primer pair BP1. (10) N5.3 with primer pair BP2.

Since CTAB extracted samples could be successfully amplified using the KAPA3G enzyme and protocol (Fig. 3.1.7.1), dilutions of the phenol-chloroform purifications

of the Dellaporta extracted DNA BIR3 and DW3 were PCR amplified with the ADH primer pair using the new kit (Fig. 3.1.7.2). Dellaporta extraction of DNA of BIR3 at 1/2 concentration produced a band (Fig. 3.1.7.2 (1)) where previously without the KAPA enzyme and protocol it had not (Fig. 3.1.5.2 (1)), indicating that in the previous attempts inhibitors were too high preventing successful PCR. However the band shown (Fig. 3.1.7.2 (1)) is still fairly weak and DW3 shows no bands (Fig. 3.1.7.2 (3 & 4)). This suggests either that the Dellaporta method of extraction is somewhat unreliable and not very effective or, more likely, that the KAPA3G enzyme and protocol, although more tolerant of PCR inhibitors, is limited in the concentration of inhibitors that it can tolerate.

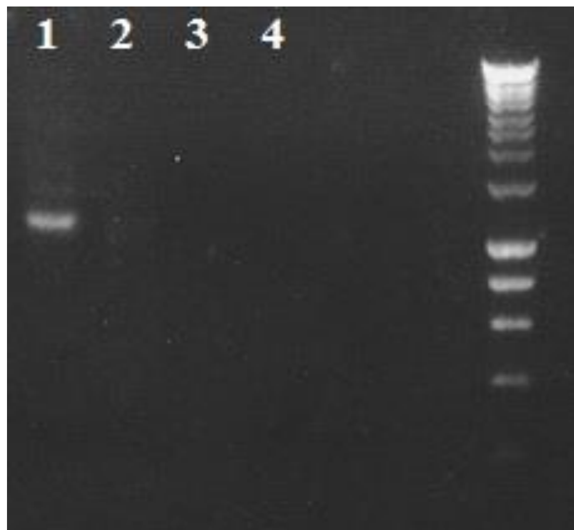

Fig. 3.1.7.2 TBE gel image of PCR products with ADH primer pair. PCR mix two program 2. (1) Bir3 1/2 concentration from Phenol-Chloroform purified**,** Dellaporta DNA extraction showing weak bands. (2) Bir3 1/4 concentration from Phenol-Chloroform purified**,** Dellaporta DNA extraction. (3) DW3 1/2 concentration from Dellaporta DNA extraction. (4) DW3 1/4 concentration from Dellaporta DNA extraction.

The KAPA3G protocol suggests using a section of leaf added directly to the PCR tube with no prior extraction (KAPA3G Plant PCR Kit, Technical Data Sheet, version 1.11). It could be possible that with minimal disruption of cells fewer inhibitory compounds would be released and by extracting DNA whilst PCR is occurring phenolics have less chance to bind to the DNA and inhibit PCR. Hence this was tried with the ADH primer pair and leaf sample 3.12G added directly to the PCR tube as per the KAPA3G protocol (KAPA3G Plant PCR Kit, Technical Data Sheet, version 1.11), with various levels of KAPA3G plant PCR enhancer (0µl, 0.25µl, 0.5µl) and compared to CTAB extracted 3.12G DNA(Fig 3.1.7.3).

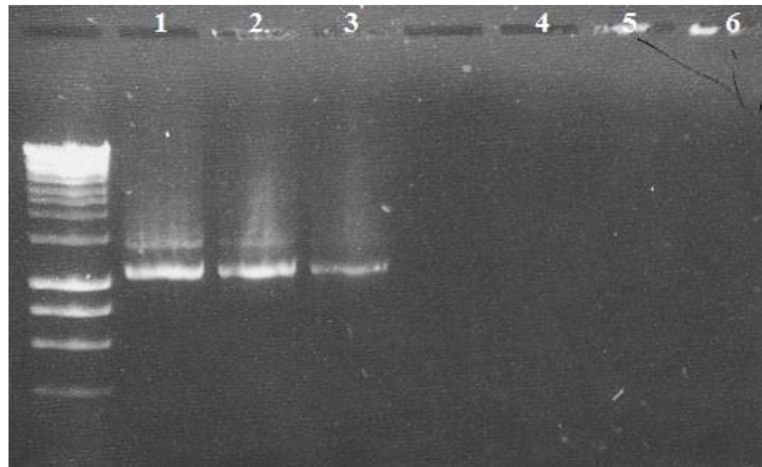

Fig. 3.1.7.3. TBE gel image of PCR products with ADH primer pair. using PCR mix 2, PCR program three. (1-3) 3.12G from CTAB DNA extraction showing strong bands with (1) 0µl (2) 0.25 µl (3) 0.5 µl KAPA plant PCR enhancer (4-6) 3.12G using a fragment of leaf in PCR tube a per protocol with KAPA3G Plant PCR kit showing no evidence of bands with (4) 0 µl (5) 0.25 µl (6) 0.5 µl KAPA plant PCR enhancer.

PCR directly from the leaf was unsuccessful and did not produce any bands irrespective of enhancer concentration (Fig 3.1.7.3 (4-6)). This could be due to insufficient release of DNA from the leaf tissue. PCR amplification with the ADH primer pair of DNA extracted via CTAB from sample 3.12G with the same levels of plant PCR enhancer as the leaf samples (0µl, 0.25µl, 0.5µl) produced banding for all

concentrations of plant PCR enhancer (Fig. 3.1.7.3 (1-3)). A plant PCR enhancer concentration of 0.5µl lowered DNA quality producing a weak band (Fig. 3.1.7.3 (3)) perhaps due to the plant PCR enhancer inhibiting the reaction itself. There was no obvious difference when 2.5µl was added (Fig. 3.1.7.3 (2)) compared to use without the plant PCR enhancer (Fig. 3.1.7.3 (1)).

This enzyme continued to be used in further PCR reactions since it is evident that it improves reliability of PCR in mildly contaminated samples, but the plant PCR enhancer was never included due to no evidence being found of it being useful to the reaction (Fig. 3.1.7.3). Although CTAB extraction combined with PCR with the KAPA3G enzyme produced strong results, CTAB extractions are extremely time consuming and not really feasible for a high throughput of samples and so a more rapid method of extraction was investigated.

3.1.8 MoBio PowerPlant® Pro DNA Isolation Kit

Kit from: MoBio Laboratories Inc. Calafornia, USA

MoBio PowerPlant® Pro DNA Isolation kit advertises its ability to replace CTAB techniques and rapidly isolate inhibitor free DNA from tough samples such as pine needles (MoBio PowerPlant® Pro DNA Isolation Kit Description, 2010). DNA extraction of 3.12G was attempted following the protocol outlined in (MoBio PowerPlant® Pro DNA Isolation kit instruction Manual. version 01132012) and including the advised heating step for troublesome samples. Running the DNA extracts on a gel showed no evidence of successful extraction (data not shown). This

extraction method relies on releasing DNA from cells using bead beating tubes, since *Betula* leaves are tough yet springy, once thawed it was possible that the bead beating was insufficient to release DNA from the cells. A comparison of extraction between pre-ground tissue of sample 3.12G and unground tissue of sample 3.12G was therefore run (Fig. 3.1.8.1).

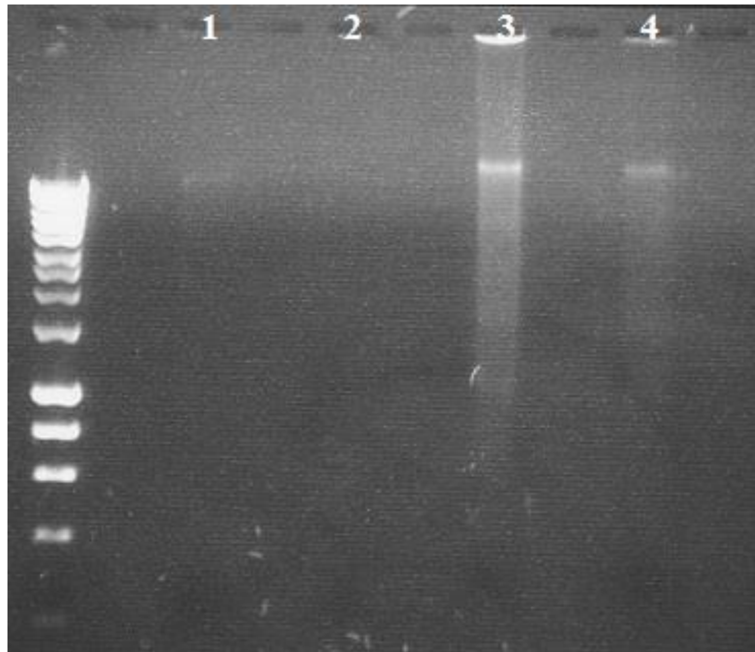

Fig. 3.1.8.1. TBE gel image of DNA extraction from sample 3.12G using MoBio PowerPlant® Pro DNA Isolation kit. (1) Elution one using a section of leaf showing weak bands. (2) Elution two using a section of leaf not showing any bands. (3) Elution one grinding the leaf tissue before addition to bead beating tubes showing strong bands. (4) elution two grinding the leaf tissue before addition to bead beating tubes showing fairly strong bands.

Pre-grinding of tissue of sample 3.12G before addition to the bead beating tubes using the MoBio kit produced strong bands for both DNA elutions (Fig. 3.1.8.1). Elutions of extraction of 3.12G, without pre-grinding, run at the same time and with the same conditions as the pre-ground sample, did not produce such strong bands as the pre-ground sample (Fig. 3.1.8.1). It is therefore evident that elutions of the

ground sample (Fig. 3.1.8.1 (3 & 4)) had a much higher yield than the unground samples (Fig. 3.1.8.1 (1 & 2)). Since with the exception of pre-grinding of the ground sample in liquid nitrogen, all other conditions remained the same and samples were run at the same time, the pre-grinding of samples is obviously necessary to allow access to the DNA. A grinding step prior to bead beating was therefore used for all further DNA extractions with this kit.

Since prior grinding may suggest that a bead beating step would not be required a test comparing pre ground samples with and without bead beating step (but still shaken for the bead beating time) was run with sample N3.4 (Fig. 3.1.8.2).

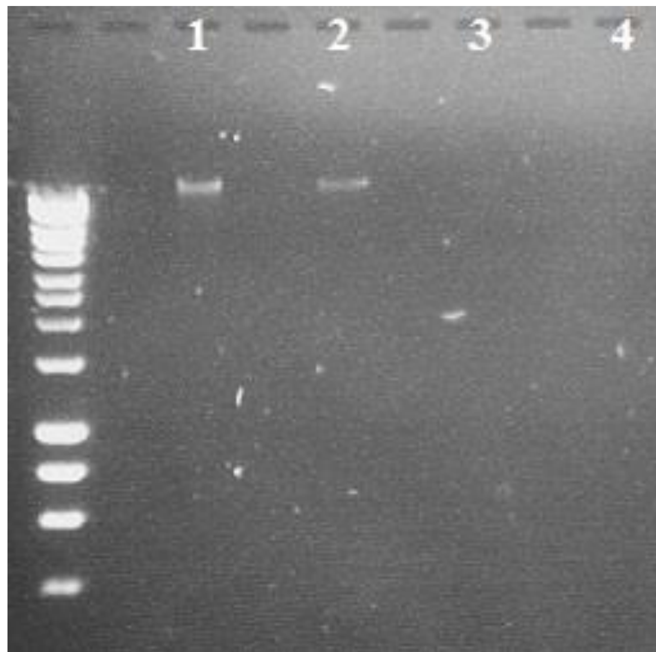

Fig. 3.1.8.2. TBE gel image of DNA extraction from sample N3.4 using MoBio PowerPlant® Pro DNA Isolation kit with ground tissue. (1) With bead beating step, elution one showing a strong band. (2) With bead beating step, elution two showing a strong band. (3) Without bead beating step, elution one showing no evidence of a band. (4) Without bead beating step, elution two showing no evidence of a band.

Elutions of the extraction run using the bead beating step successfully extracted (Fig. 3.1.8.2 (1 & 2) whilst the elutions from the sample without a bead beating step either did not successfully extract or the DNA concentration was too low to be evident on a gel (Fig. 3.1.8.2 (3 & 4)). The bead beating step may therefore still be necessary for further thorough mixing of reagents and sample or to further beak up of the sample.

With the necessary modifications this DNA extraction kit worked reliably and quickly and was therefore used for DNA extraction from all further samples.

3.1.9 Conclusions and Summary

Phenolics were the most likely major problem affecting successful DNA extraction and PCR amplification. As well as being naturally high in phenolics (Nurmi et al., 1996) many of the leaf samples had signs of insect damage which is known to increase the level of phenolics in *Betula* (Hartley & Firn, 1989). It is also possible that some DNA degradation had occurred in samples prior to extraction. Samples were collected in the field and stored in a standard freezer (~-18°C) before transporting to Durham University where they were stored in a -80°C freezer. Buds of *B.* pubescens have been shown to have a maximum hardiness of -60°C and buds of B. nana -80°C (Stushnoff & Junttila, 1986) hence time in the -18°C freezer may have only slowed enzyme activity rather than stopping it. Stushnoff and Junttila (1986) did not study maximum cold hardiness of leaves. Since leaves are dropped prior to winter it is likely that their maximum cold hardiness is a higher temperature than that of the buds, nevertheless it is likely that -18°C may not be sufficiently low to stop enzyme activity in the leaves. Hence both bud and leaf samples may have

started to degrade prior to being moved to -80°C storage. None of the publications where CTAB was used to extract DNA from *Betula* mention the time between sampling and extraction or the state of the leaves used (e.g. degree of herbivory) (e.g. Järvinen et al., 2004., Palme, Palsson & Lascoux, 2004, Maliouchenko et al., 2007). If these samples were extracted immediately or rapidly frozen to -80°C or colder and did not suffer herbivory damage that may explain why some papers successfully extracted DNA using only CTAB such as (Palme, Palsson & Lascoux, 2004 and Maliouchenko et al., 2007). If other samples however experienced herbivory or took longer to be frozen to -80°C or colder that may explain why other samples required the use of a kit for extraction such as in Järvinen et al., 2004.

Selection of methods to use for DNA extraction depends on individual circumstances. If there is a small number of samples to be extracted and cost is of greatest priority CTAB extraction works with reasonable reliability. An extra phenol purification step after CTAB extraction can be added to improve results or extract DNA from particularly difficult samples, (Oakenfull, Baxter & Knight, 2013). However this is an extremely time consuming method. The MoBio PowerPlant® Pro DNA Isolation kit method of extraction proved extremely rapid and reliable, for these species of *Betula* making it ideal for extraction of the large number of difficult *Betula* samples studied here. Hence in this case MoBio PowerPlant® Pro DNA Isolation kit extraction was chosen as the best method. For future experiments in which this Isolation kit is used other more readily available and less expensive enzymes could be investigated rather than using KAPA3G. Although amplification with RedTAQ proved difficult and resulted in samples having a predisposition to float out of gel wells, other PCR enzymes without a combined dye may prove more

effective. Another interesting route to explore would be that of the band formed in the initial Edwards extraction (Fig. 3.1.2.1 (4)). If further experimentation showed that to be a true band, combining the gentle Edwards technique with compounds known to reduce effect of phenolics such as β-Mercaptoethanol, DDT, ascorbic acid and PVP (Lefort & Douglas, 1999) may, potentially, improve results.

3.2 Partial Alcohol Dehydrogenase (ADH) Alleles

3.2.1 Background

The ADH gene was chosen for this study for a number of reasons. Firstly the ADH gene has been partially sequenced in both *B. nana* and *B. pubescens* (Järvinen, et al., 2004), allowing for comparison between published and extracted sequences. The same primer pair can also be used for amplification of ADH alleles in both *B. nana* and *B. pubescens* (Järvinen, et al., 2004) making it ideal to use when unsure about the hereditary of samples. Finally these primers produce different banding patterns for *B. nana* and *B. pubescens. B. nana* has a short allele only and *B. pubescens* has both a long and short allele (Järvinen, et al., 2004). This results in the ability to get a rudimentary idea of hereditary purely from gel banding patterns after amplification with these primers.

The ADH gene has a variety of known functions and responses in other plants. Most of these responses are linked to dehydration. ADH mRNA is known to increase in response to anaerobiosis in both the leaves and roots (Xie & Wu, 1989; Hoeren et al.,

1998) and also helps plants survive hypoxia (Freeling & Bennett, 1985; Dolferus et al., 1994). It is also known to increase in response to dehydration (Dolferus et al., 1994) and is induced in response to low temperatures, but is not involved in freezing tolerance in *Arabidopsis thaliana* (Jarillo et al., 1993; Dolferus et al., 1994). All these responses are linked to dehydration either via hypoxia killing roots and reducing water uptake, lack of water, low temperatures reducing the availability of water and anaerobiosis resulting either from closure of stomata due to reduced water availability or unavailability of oxygen to the roots such as in anoxic or hypoxic conditions which can also result in the death of roots.

3.2.2 Gel Bands

PCR amplification with the ADH primer pair produced a variety of banding patterns for all samples. All of the *B. nana* type individuals showed the expected banding pattern, consisting of a band for the long allele only (N3.1, N3.2, N3.3, (Fig. 3.2.2.5 (1-4)) N3.4 (Fig. 3.2.2.6), N3.5, N5.1, N5.2, (Fig. 3.2.2.5 (5-7)) N5.3 (Fig. 3.2.2.6 (2)), N5.4, N5.5, DW1, DW3, (Fig. 3.2.2.5 (8-11)) DW4, DW5, DW6 (Fig. 3.2.2.3)). Since the *B. nana* type samples were selected for morphological similarity to "pure" *B. nana* and were selected from shrubs outside of the patches this is not an unexpected result. Not all of the *B. pubescens* ssp. *czerapanovii* types showed their expected banding pattern. 5.28B (Fig. 3.2.2.4 (11)) and 5.34C (Fig. 3.2.2.1 (1)) showed a *B. nana* type banding pattern, of the long allele only. Although greater amplification of the short allele occurs in the majority of samples which display both a long and short allele (BPS1, BPS4, (Fig. 3.2.2.4 (1, 4)) 3.12G (Fig. 3.2.2.6 (1)), 3.18E, 5.32A (Fig. 3.2.2.4 (9-10)), 6.29A, 6.12C, 6.35E, 6.22A, (Fig. 3.2.2.2 (2-5))

BIR1 (Fig. 3.2.2.7), BIR4, BIR5 (Fig. 3.2.2.4 (14, 15))and BIR6 (Fig. 3.2.2.5 (1))), samples BPS5, 3.3A, 3.5C, 3.7A, (Fig. 3.2.2.4 (5-8)) BIR2 and BIR3 (Fig. 3.2.2.4 (12-13)) showed an extremely weak band for the long allele. 6.18F (Fig. 3.2.2.2 (1)) showed a stronger long allele band than the short allele sample, BPS2 (Fig. 3.2.2.4 (2)) had equal intensity bands and BPS3 (Fig. 3.2.2.4 (3)) and 5.11A (Fig. 3.2.2.1 (2)) only showed a band for the short allele.

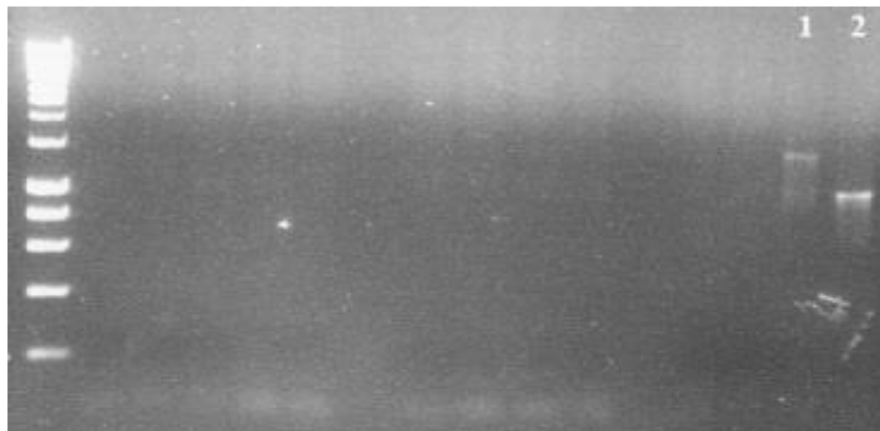

Fig. 3.2.2.1. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) 5.34C. (2) 5.11A.

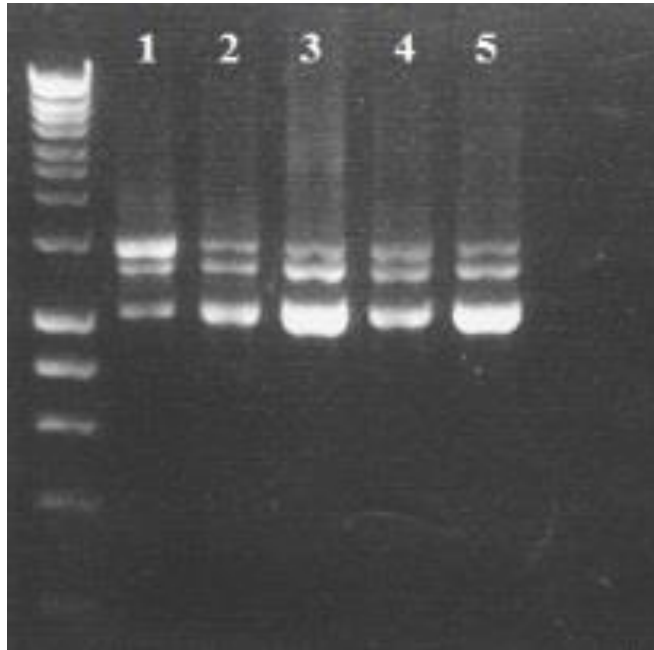

Fig. 3.2.2.2. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) 6.18F (2) 6.29A. (3) 6.12C (4) 6.35E (5) 6.22A.

Fig. 3.2.2.3. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) DW4. (2) DW5 (3) DW6.

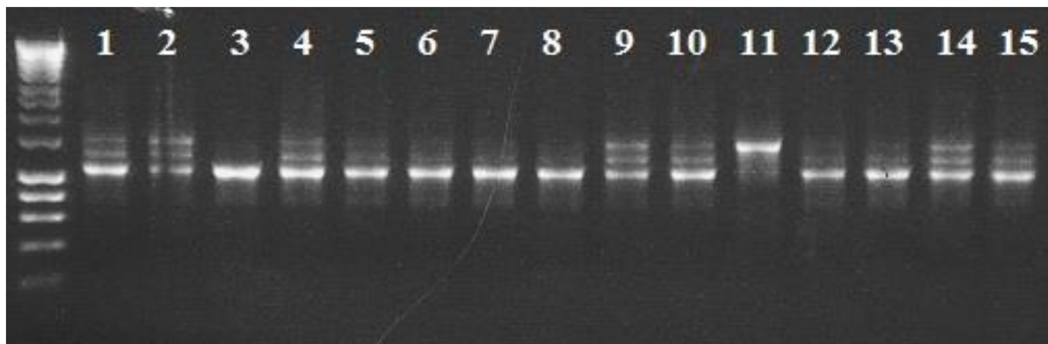

Fig. 3.2.2.4. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) BPS1. (2) BPS2. (3) BPS3. (4) BPS4. (5) BPS5. (6) 3.3A. (7) 3.5C. (8) 3.7A. (9) 3.18E. (10) 5.32A. (11) 5.28B. (12) BIR2. (13) BIR3. (14) BIR4. (15) BIR5.

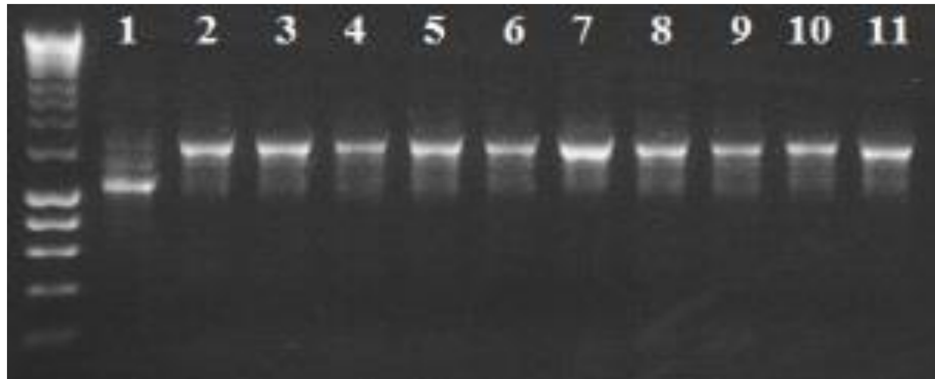

Fig. 3.2.2.5. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) BIR6. (2) N3.1. (3) N3.2. (4) N3.3. (5) N3.5. (6) N5.1. (7) N5.2. (8) N5.4. (9) N5.5. (10) DW1. (11) DW3.

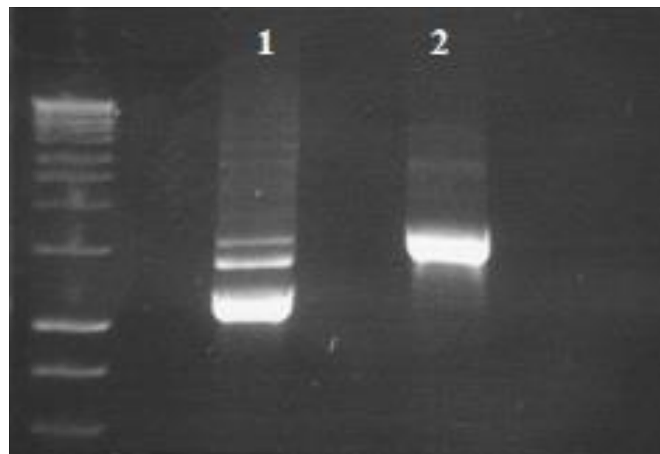

Fig. 3.2.2.6. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) 3.12G. (2) N5.3.

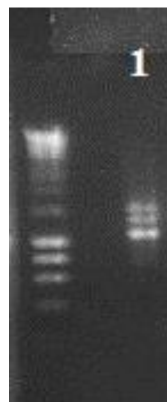

Fig. 3.2.2.7. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) BIR1.

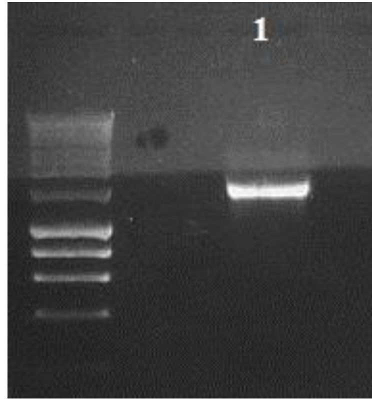

Fig. 3.2.2.6. TBE Gel image of PCR products using the ADH primer pair with DNA extracted from samples (1) N3.4.

The *B. pubescens* ssp. *czerapanovii* type samples covered trees from one edge of the patch to the other, with a large range in morphologies so variation is to be expected. Although the differences in banding pattern strength could be explained by fluctuations in PCR, since the majority of samples were amplified at the same time, (Figs. 3.2.2.4 & 3.2.2.5) and these show the majority of different banding strength patterns, it seems unlikely. Also considering the consistency in banding patterns, forming a limited variety of different pattern intensities it is more likely that the different banding pattern intensities are due to differences in allele numbers rather than PCR fluctuations.

Since *B. pubescens* is tetraploid showing evidence of being an alloploid (Walters, 1968; Howland, Oliver & Davy, 1995; Järvinen et al., 2004; Schenk et al., 2008) the majority of "pure" type samples should have two short ADH alleles and two long ADH alleles. Since the long allele forms a weak band and the short a strong band in the majority of samples with both alleles, assuming the majority of *B. pubescens* ssp. *czerapanovii* samples show a "pure" type banding pattern, this would suggest that

the short allele amplifies preferentially over the long allele. This is supported by the examples with an even weaker band for the long allele (BPS5, 3.3A, 3.5C, 3.7A, (Fig. 3.2.2.4 (5-8)) BIR2 and BIR3 (Fig. 3.2.2.4 (12-13))). These samples must have at least one long allele copy in order for the band to be present. Since no *B. pubescens, B. nana* or *B. pubescens X B. nana* hybrid has ever shown a ploidy greater than tetraploid (Anamthawat-Jónsson & Tomasson, 1990; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2010), the smallest ratio of long alleles to short alleles any individual could display would be one long to three short alleles. The second smallest ratio would be one long to two short alleles and the third smallest ratio could either be two long to two short alleles or one long allele and one short allele. It is unlikely that the majority of samples are triploid, with the ratio one long to two short alleles since only approximately 10% of randomly sampled trees by Anamthawat-Jónsson et al. (2003) in Iceland were triploid. These triploids were also mostly found on the edges of *B. pubescens* patches (Anamthawat-Jónsson et al., 2003) and all bar one central sample (site 6) show a stronger short allele band than the long allele band (Fig. 3.2.2.2) suggesting the stronger short to long allele band is not a triploid ratio. The majority of samples also cannot be one long allele to three short alleles since this is the smallest ratio of long to short alleles and therefore those with even weaker long allele bands (BPS5, 3.3A, 3.5C, 3.7A, (Fig. 3.2.2.4 (5-8)) BIR2 and BIR3 (Fig. 3.2.2.4 (12-13))) could not exist. Hence the majority of samples with both alleles must either be diploid with one short allele to one long allele or tetraploid with two short alleles and two long alleles i.e. the expected “pure” type banding pattern. Those samples with a very weak long allele band (weaker than the majority) are therefore likely to be either tetraploid with one long to three short alleles or triploid with one long to two short alleles. Since the

samples with weaker long allele to short allele bands are likely to have two long to two short alleles or one short one long allele, sample BPS2 (Fig. 3.2.2.4 (2)) with equal strength bands and sample 6.18F (Fig. 3.2.2.2 (1)) with the long allele band more intense than the short, must have either a greater number of long alleles or a smaller number of short alleles than those with two short, two long or one short one long. The remaining potential ratios are three long and one short and two long and one short. Since BPS2 has equal intensity bands (Fig. 3.2.2.4 (2)) whilst 6.18F has a stronger long allele band (Fig. 3.2.2.2 (1)) it would suggest that 6.18F has the highest ratio of long alleles (three long to one short) and BPS has the lower ratio (two long alleles and one short). Those samples with only the short allele (BPS3 (Fig. 3.2.2.4 (3)) and 5.11A (Fig. 3.2.2.1 (2))) have inherited the short alleles from *B. pubescens* only and could be either diploid, triploid or tetraploid.

There are a couple of potential origins for the wide array of banding patterns observed. It has been suggested by (Brown & Al-Dawoody, 1979) that aneuploidy plants may be formed due to abnormal chromosome separations, however no aneuploid cells or plants were found in studies by (Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson & Thórsson, 2003; Anamthawat-Jónsson et al., 2010) suggesting either that any aneuploid gametes produced are not viable or that due to difficulty with chromosome counting in *Betula* (Brown & Al-Dawoody, 1979; Anamthawat-Jónsson & Tomasson, 1990; Anamthawat-Jónsson & Tomasson, 1999; Anamthawat-Jónsson & Thórsson, 2003) that the abnormal chromosome numbers observed may be errors in counting.

Hybridisation is known to occur between these two species (Elkington, 1968; Anamthawat-Jónsson & Tomasson, 1990; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2003). Evidence strongly indicates that *B. pubescens* is an alloploid (Walters, 1968; Howland, Oliver & Davy, 1995; Järvinen et al., 2004; Schenk et al., 2008) and that the short allele is from one parent species and the long allele from another (Järvinen et al., 2004). Due to evidence for the tendency of homologous chromosomes from each original parent to pair faithfully in allopolyploids (Sybenga, 1996) gametes of "pure" *B. pubescens* should contain one chromosome with the short ADH allele (from one of its ancestral parents) and one chromosome with the long ADH allele (from its other ancestral parent). Hence first generation triploids between *B. nana* and *B. pubescens* should contain one chromosome from each parent (*B. nana* and each of *B. pubescens* ancestral parents) and therefore have two long alleles and one short. As shown earlier this would give equal sized bands for the long and the short allele as seen in BPS2 (Fig. 3.2.2.4 (2)). First generation triploids would therefore not explain the other observed banding patterns i.e. short allele band only, long allele band only in *B. pubescens* ssp. *czerapanovii* type specimens, weaker that average long allele bands for those with both short and long allele bands and a stronger long allele band than short allele band.

Introgressive hybridisation is known to occur between these species (Elkington, 1968; Anamthawat-Jónsson & Tomasson, 1990; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2003). The triploid is not sterile and can produce both haploid and diploid gametes backcrossing to parent plants and promoting gene flow between the two species (Anamthawat-Jónsson & Tomasson,

1990; Anamthawat-Jónsson & Thórsson, 2003). Since triploids can produce both haploid and diploid pollen this allows for the splitting of the different parent chromosomes of *B. pubescens* (i.e. the one carrying the long allele from the one carrying the short allele) within the triploid gametes and cross breeding back to other triploids, *B. nana* or *B. pubescens*.

Crossing over is another potential cause of the different banding patterns, but due to the high number of different banding patterns and the rarity of crossing over events, this seems unlikely to be the main cause. Since all evidence in other species so far shows that in allopolyploids homologous chromosomes from each original parent pair faithfully (Sybenga, 1996), it is unlikely that a chromosome carrying the short allele would ever pair with one carrying the long allele, therefore crossing over is unlikely to occur in pure type *B. pubescens* plants. Pure type *B. nana* plants only carry the long allele form and hence crossing over between long and short alleles could not occur here. How chromosomes pair in triploids is unknown, other than that both haploid and diploid pollen can be produced (Anamthawat-Jónsson & Tomasson, 1990; Anamthawat-Jónsson & Thórsson, 2003). However if long and short allele carrying chromosomes could not pair in the F1 triploid the gametes produced could only be haploid with one long allele (like pure *B. nana* pollen) and diploid with one long and one short allele (like pure *B. pubescens* pollen) although the long allele chromosome in each gamete could be from either *B. nana* or *B. pubescens*. No amount of backcrossing would ever produce individuals with short alleles only, two short one long, one long three short or three long one short. Since individuals with this banding pattern have been observed it is evident that long and short alleles can pair in the triploid. Since triploids are relatively rare, (in Iceland

approximately 10% of randomly sampled trees by Anamthawat-Jónsson et al. (2003) were triploid), and chromosome pairing patterns in later generation diploid or tetraploid hybrids is unknown due to the rarity of long and short alleles pairing and the rarity of crossing over events, crossing over is unlikely to be the main factor. Consequently introgressive hybridisation is likely to be the major cause of different banding patterns.

The difference in banding patterns can be used to infer relatedness between samples. All the *B. nana* samples, sample 5.28B (Fig. 3.2.2.4 (11)) and sample 5.34C (Fig. 3.2.2.1 (1)) show the same banding pattern of a long allele band only and would initially suggest similarity to each other. Despite showing a banding pattern expected for "pure" *B. nana*, there could be different levels of ploidy occurring and long alleles could be from any mixture of *B. nana* and *B. pubescens*. It seems likely that samples 5.28B and 5.34C are different from the *B. nana* samples and more similar to each other as they are from the leeward side (site 5). Leeward side trees are morphologically unlike *B. nana* being the tallest and least polycormic trees in the patch. Hence samples 5.28B and 5.34C cannot be "pure" *B. nana* and likewise due to the lack of short allele bands cannot be "pure" *B. pubescens*. It would therefore seem likely that samples 5.28B and 5.34C would contain a greater number of *B. pubescens* chromosomes (and hence *B. pubescens* type long ADH alleles) as opposed to *B. nana* chromosomes (and *B. nana* type long ADH alleles). Since 5.28B and 5.34C are not "pure" *B. nana* or "pure" *B. pubescens* they must be hybrids. As shown earlier first generation hybrids are likely to contain one long allele from *B. nana*, one long allele from one of the ancestral parents of *B. pubescens* and one short allele from the other ancestral parent of *B. pubescens*. These individuals are therefore very likely to

be highly introgressed individuals from backcrossings between hybrids, *B. pubescens* and *B. nana*.

Samples BPS1, BPS2, BPS4, (Fig. 3.2.2.4 (1, 4)) 3.12G (Fig. 3.2.2.6 (1)), 3.18E, 5.32A (Fig. 3.2.2.4 (9-10)), 6.29A, 6.12C, 6.35E, 6.22A, (Fig. 3.2.2.2 (2-5)) BIR1 (Fig. 3.2.2.7), BIR4, BIR5 (Fig. 3.2.2.4 (14, 15)) and BIR6 (Fig. 3.2.2.5 (1)) show a *B. pubescens* "pure" type banding pattern of a strong short ADH allele and a weaker long ADH allele. A wide range of morphologies are seen in these samples with saplings from outside the main patch (BPSX), windward (3.X), central (6.X) and leeward (5.X) samples as well as morphologically pure *B. pubescens* ssp. *czerapanovii* samples (BIRX). If these samples are "pure" *B. pubescens* it would indicate high morphological plasticity in *B. pubescens* ssp. *czerapanovii*. However some samples could be diploid and the long alleles could be from either *B. nana* or *B. pubescens* ssp. *czerapanovii* therefore some or all samples may be varying degrees of hybrid. Sequence data would be required for further indication of hybrid status of these samples.

BPS5, 3.3A, 3.5C, 3.7A, (Fig. 3.2.2.4 (5-8)) BIR2 and BIR3 (Fig. 3.2.2.4 (12-13)) with weaker than average long alleles could either have one long to three short alleles or one long to two short alleles, as detailed earlier. These samples therefore cannot be pure samples and must be some form of hybrid. This ratio of alleles cannot occur in an F1 hybrid (which as detailed earlier would contain two long alleles and one short allele) and hence these samples must be from introgressive hybridisation between triploids, *B. pubescens* and *B. nana*. Samples BIR2 and BIR3 were selected for being morphologically *B. pubescens* ssp. *czerapanovii*. Since these samples are

evidently introgressed hybrids it would strongly suggest that morphology cannot be used to predict genetic purity of a sample.

BPS3 (Fig. 3.2.2.4 (3)) and 5.11A (Fig. 3.2.2.1 (2)) only display a band for the short allele. *B. nana* does not have a short allele (Järvinen et al., 2004) therefore this allele can only come from *B. pubescens* ssp. *czerapanovii.* Since the homologous chromosomes from each original parent pair faithfully in allopolyploids (Sybenga, 1996) chromosomes carrying short and long alleles are highly unlikely to be separated in *B. pubescens* gametes. In order for chromosomes carrying the short allele to be separated from the chromosomes carrying the long allele, backcrossing with triploid hybrids or crossing of two F1 triploid hybrids must have occurred. These samples must have undergone a high level of introgressive hybridisation in order to show these banding patterns and are likely to contain a mix of both *B. pubescens* and *B. nana* genes.

Sample BPS2 (Fig. 3.2.2.4 (2)) shows equal strength bands likely to have arisen from two long alleles and one short allele. This sample could well be an F1 hybrid between *B. nana* and *B. pubescens* ssp. *czerapanovii* and hence be exactly two thirds *B. pubescens* ssp. *czerapanovii* to one third *B. nana*. It could also be a later generation triploid displaying F1 hybrid banding patterns. Sample 6.18F (Fig. 3.2.2.2 (1)) shows a stronger than average long allele band suggesting the allele frequency three long alleles to one short allele. This is evidence for a high level of hybridisation via intermediate triploid forms. The short allele must be from *B. pubescens* ssp. *czerapanovii* but the long alleles could be from any combination of *B. nana* and/or *B. pubescens* ssp. *czerapanovii*.

It is not possible to be certain if all the *B. nana* samples, 5.28B (Fig. 3.2.2.4 (11)) and 5.34C (Fig. 3.2.2.1 (1)) are pure *B. nana* (although as stated earlier, for reasons of morphology, it is highly unlikely that 5.28B and 5.34C are pure *B. nana*) and that BPS1, BPS2, BPS4, (Fig. 3.2.2.4 (1, 4)) 3.12G (Fig. 3.2.2.6 (1)), 3.18E, 5.32A (Fig. 3.2.2.4 (9-10)), 6.29A, 6.12C, 6.35E, 6.22A, (Fig. 3.2.2.2 (2-5)) BIR1 (Fig. 3.2.2.7), BIR4, BIR5 (Fig. 3.2.2.4 (14, 15)) and BIR6 (Fig. 3.2.2.5 (1)) are all pure *B. pubescens* ssp. *czerapanovii*. It is however evident that Samples BPS5, 3.3A, 3.5C, 3.7A, (Fig. 3.2.2.4 (5-8)) BIR2 and BIR3 (Fig. 3.2.2.4 (12-13)), 6.18F (Fig. 3.2.2.2 (1), BPS2 (Fig. 3.2.2.4 (2)), BPS3 (Fig. 3.2.2.4 (3)) and 5.11A (Fig. 3.2.2.1 (2)) are all varying levels of hybrids and that introgressive hybridisation occurs between these two species in Abisko.

In order to look at the "purity" of those apparently pure samples choromosome counts of each individual would need to be taken, these are notoriously difficult in *Betula* (Brown & Al-dawoody, 1979; Anamthawat-Jónsson & Tomasson, 1990. Anamthawat-Jónsson & Tomasson, 1999; Anamthawat-Jónsson & Thórsson, 2003) and would still not show whether long alleles are *B. pubescens* ssp. *czerapanovii* or *B. nana* in origin. Since the long allele of *B. pubescens* and *B. nana* differ slightly in sequence (Järvinen et al., 2004), sequence data for the ADH allele from each sample could help elucidate the origin of the alleles in each sample as well as relatedness between individuals. Hence sequencing and analysis of ADH alleles was therefore regarded as essential.

3.2.3. Sequence Data

None of the very weak long bands were successfully extracted from the gels, most likely due to the very low concentrations of PCR product in the gel as a result of the preferential amplification of the short allele.

The samples sequenced could be diploid, triploid or tetraploid. Some samples could have up to four copies/versions of the same length ADH allele, and the majority of samples will have at least two copies/versions of at least one of the ADH allele lengths. DNA was sequenced directly from purified gel bands, since alleles of the same length were not separated it is possible that sequence data consists of up to four different overlapping versions of the same length ADH allele. Where these overlapping alleles differ in a single base, a series of overlapping smaller peaks is formed. For example, from the banding pattern, sample 3.3A is likely to contain three short alleles, within the sequence data for the short allele of sample 3.3A three overlapping bases can be seen (Fig. 3.2.3.1) suggesting each version of the short allele has a different base at this point. These come out as N's in the sequence data. Many sequences would suddenly misalign early in the sequence. (for example Fig. 3.2.3.2 sequence mangling). These are shown as a sudden early cessation of sequence in the sequence alignments. (Fig 3.2.3.3-3.2.3.24.) Sequences could have up to four different overlapping bases at any point which show as 'N' in the sequence data (Fig 3.2.3.3-3.2.3.24.) an example of three overlapping bases is shown in Fig. 3.2.3.1.

Fig. 3.2.3.1. example from 3.3A (short) of an 'N' base consisting of three different bases.

Sudden sequence misalignments also frequently occur for example (Fig. 3.2.3.2).

Fig. 3.2.3.2. Sections of ADH sequence chromatograms showing sudden misalignment of sequence (1) 6.18F (long) (2) N5.1 (long) (3) BIR5 (long).

Although this could be due to errors in PCR, the fact that the first sections of the sequence are clear and well defined and that some individuals show perfect sequences when others misalign abruptly makes PCR error somewhat unlikely. The sudden nature of these misalignments suggest that at the point of misalignment there is an insertion or deletion in one or more version of that length allele which the other

version(s) do not have in that sample, resulting in different sequences from the same individual overlapping. This is supported by reverse sequence data often clearly overlapping the tail end of misaligned forward sequences whilst remaining well defined itself, suggesting the misalignment is due to some difference in alleles earlier in the forward sequence which affects the entire remaining forward readout. These sudden misalignments are shown in the sequence alignments (Fig. 3.2.3.14-24 & Fig. 3.2.3.3–3.2.3.13) as an abrupt end to sequence readouts at a lower than expected base count. Such abrupt misalignments often occur at certain locations and when shared, may suggest similarity between samples.

Sequences were compared to published sequences from Punkaharju, Finland (*B. pubescens*) and Halsua, Finland (*B. nana*) (Järvinen et al., 2004). As seen in (Fig. 3.2.3.3 – 3.2.3.13), the short allele is extremely similar to the long allele with bases 62-521 missing (Fig. 3.2.3.3 – 3.2.3.6). However there are still differences between the published *B. nana* long allele (Fig. 3.2.3.14-24 (*B. nana*)) and the published *B. pubescens* long allele (Fig. 3.2.3.14-24 (*B. pubescens* long)).

Sequences aligned well with published sequences with with some areas of high variability. (Fig. 3.2.3.3-3.2.3.13) Sudden cessation of sequence code signifies a sudden misalignment in the sequence (Fig. 3.2.3.14-3.2.3.24) and 'N' indicating several overlapping bases (Fig. 3.2.3.1.)

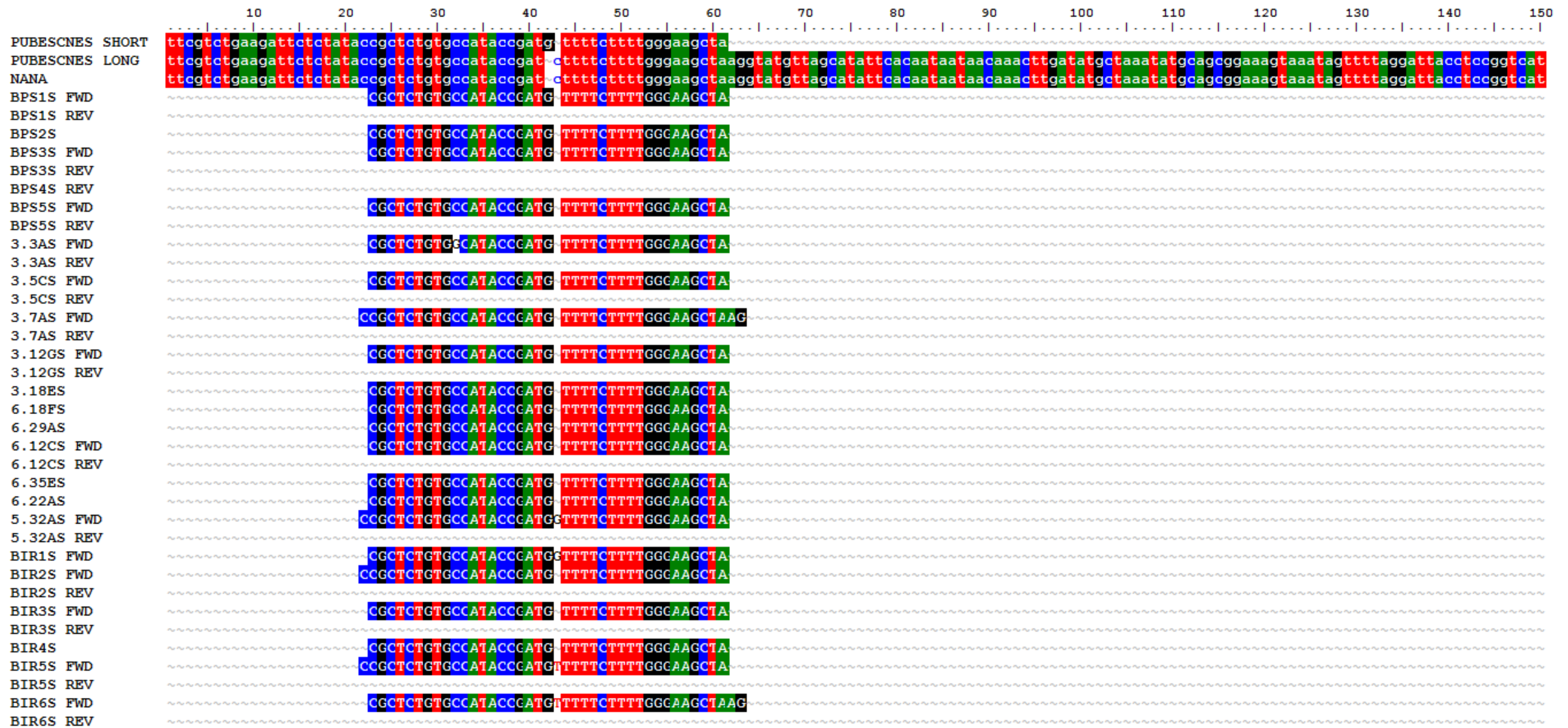

Fig. 3.2.3.3. Part 1. Full sequence of short alleles aligned with published long alleles and short allele.

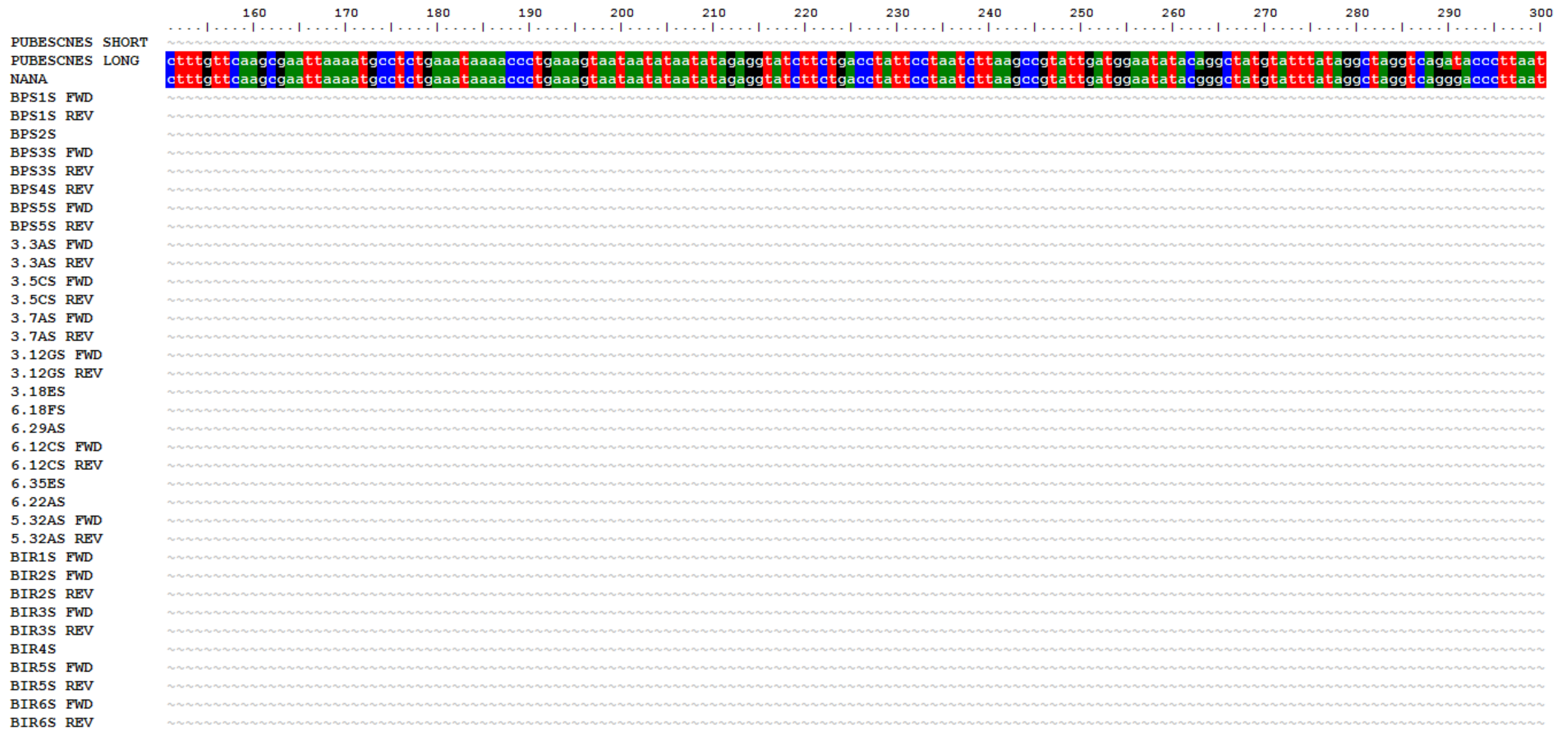

Fig. 3.2.3.4. Part 2. Full sequence of short alleles aligned with published long alleles and short allele.

Fig. 3.2.3.5. Part 3. Full sequence of short alleles aligned with published long alleles and short allele.

```
                   460       470       480       490       500       510       520       530       540       550       560       570       580       590       600
PUBESCNES SHORT                                                                       aggtaacatatttatcattgctttgtaaagttataatataaattaaatgattaa-ttacattc-atttctattagctta
PUBESCNES LONG     ttattaataaagataataaactatatacacatgtggcatataggtcataccatatagagaattttcttacgagataacatatttatcattgctttgtaaagttataatataaattaaatgattaaatt-cat-cctttt-attagctta
NANA               ttattaataaagataataaactatatacacatgtggcatataggtcataccatatagagaattttcttacgaggtaacatatttatcattgctttgtaaagttataatataaattaaatgattaaatt-cat-cctttt-attagctta
BPS1S FWD                                                                             AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA
BPS1S REV
BPS2S                                                                                 AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA-TTACATTC-ATTTCTATTAGCTTA
BPS3S FWD                                                                             AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA-TTACATTC-ATTTCTATTAGCTTA
BPS3S REV
BPS4S REV
BPS5S FWD                                                                             AGGTAACATATTTATCATTGNTTTGTAAAGTTATAATATAAATTAAATGATTAAATT
BPS5S REV
3.3AS FWD                                                                             AGGTAACATATTTATCATTGNTTTGTAAAGTTATAATATAAATTAAATGATTAAATT-CAT-CCATTTNTATTA-CTTA
3.3AS REV
3.5CS FWD                                                                             AGGTAACATATTTATCATTGTTTTGTAAAGTTATAATATAAATTAAATGATTAAATT-CAT-CCATTTTTATTA-CTTA
3.5CS REV
3.7AS FWD                                                                             AGGTAACATATTTATCATTGTTTTGTAAAGTTATAATATAAATTAAATGATTAAATT-CAT-CCATTTTTATTA-CTTA
3.7AS REV
3.12GS FWD                                                                            AGGTAACATATTTATCATTGNTTTGTAAAGTTATAATATAAATTAAATGATTAAATT-CAT-CCATTTCTATTA-CTTA
3.12GS REV
3.18ES                                                                                AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAAATT-CAT-CCATTTTTATTA-CTTA
6.18FS                                                                                AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA-TTACATTC-ATTTCTATTAGCTTA
6.29AS                                                                                AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA-TTACATTC-ATTTCTATTAGCTTA
6.12CS FWD                                                                            AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA
6.12CS REV
6.35ES                                                                                AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA-TTACATTC-ATTTCTATTAGCTTA
6.22AS                                                                                AGGTAACATATTTATCATTGTTTTGTAAAGTTATAANATAAATTAAATGATTAAATT-CAT-CCATTTTTATTA-CTTA
5.32AS FWD                                                                            AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA
5.32AS REV
BIR1S FWD                                                                             AGGTAACATATTTATCATTGTTTTGTAAAGTTATAATATAAATTAAATGATTAAATT-CAT-CCATTTTTATTAGCTTA
BIR2S FWD                                                                             AGGTAACATATTTATCATTGCTTTGTAAANTTATAATATAAATTAAATGATTAA
BIR2S REV
BIR3S FWD                                                                             AGGTAACATATTTATCATTGCTNTGTAAAGTTATAATATAAATTAAATGATTAAAT
BIR3S REV
BIR4S                                                                                 AGGTAACATATTTATCATTGTTTTGTAAAGTTATAACATAAATTAAATGATTAAATT-CAT-CCATTTTTATTA-CTTA
BIR5S FWD                                                                             AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA
BIR5S REV
BIR6S FWD                                                                             AGGTAACATATTTATCATTGCTTTGTAAAGTTATAATATAAATTAAATGATTAA
BIR6S REV
```

Fig. 3.2.3.6. Part 4. Full sequence of short alleles aligned with published long alleles and short allele.

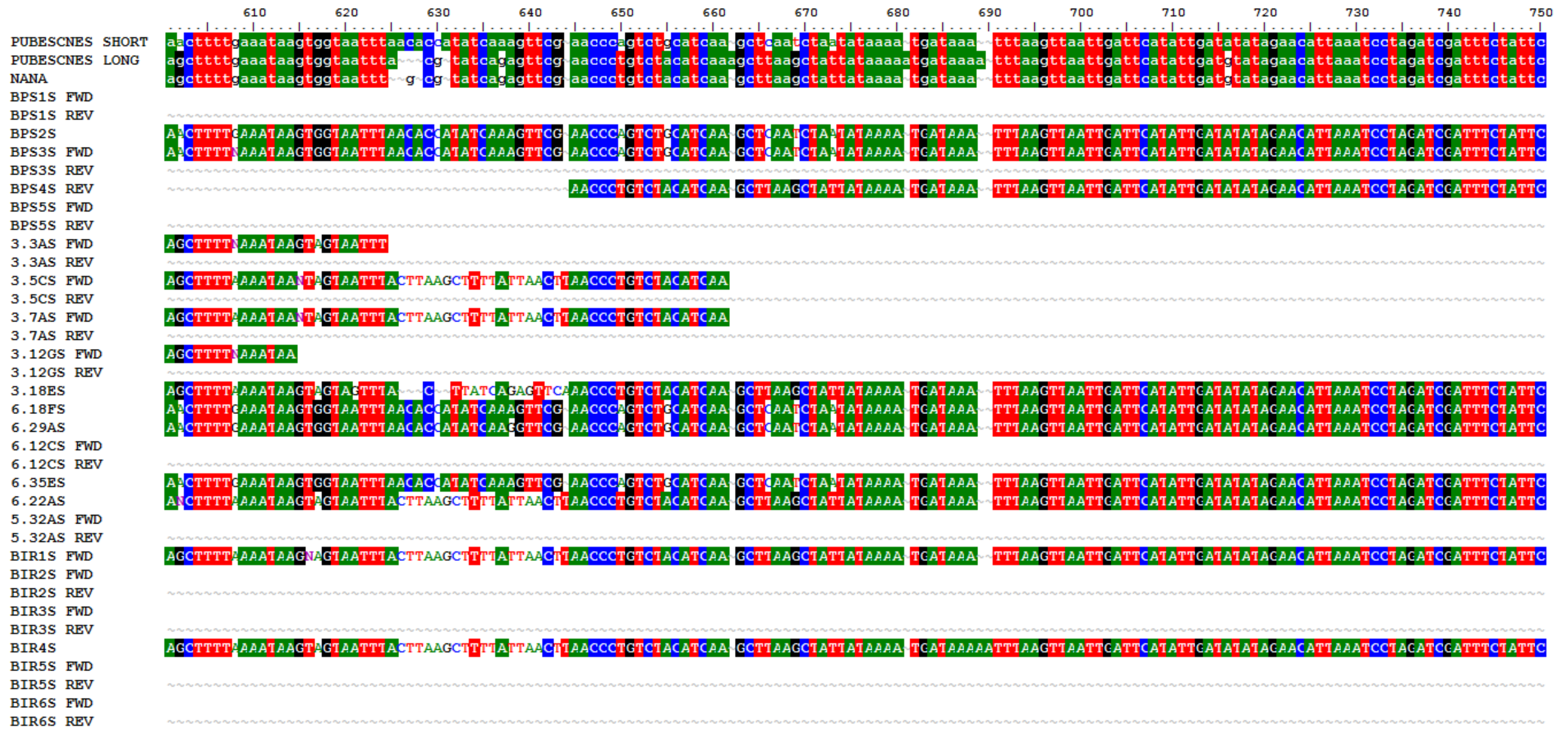

Fig. 3.2.3.7. Part 5. Full sequence of short alleles aligned with published long alleles and short allele.

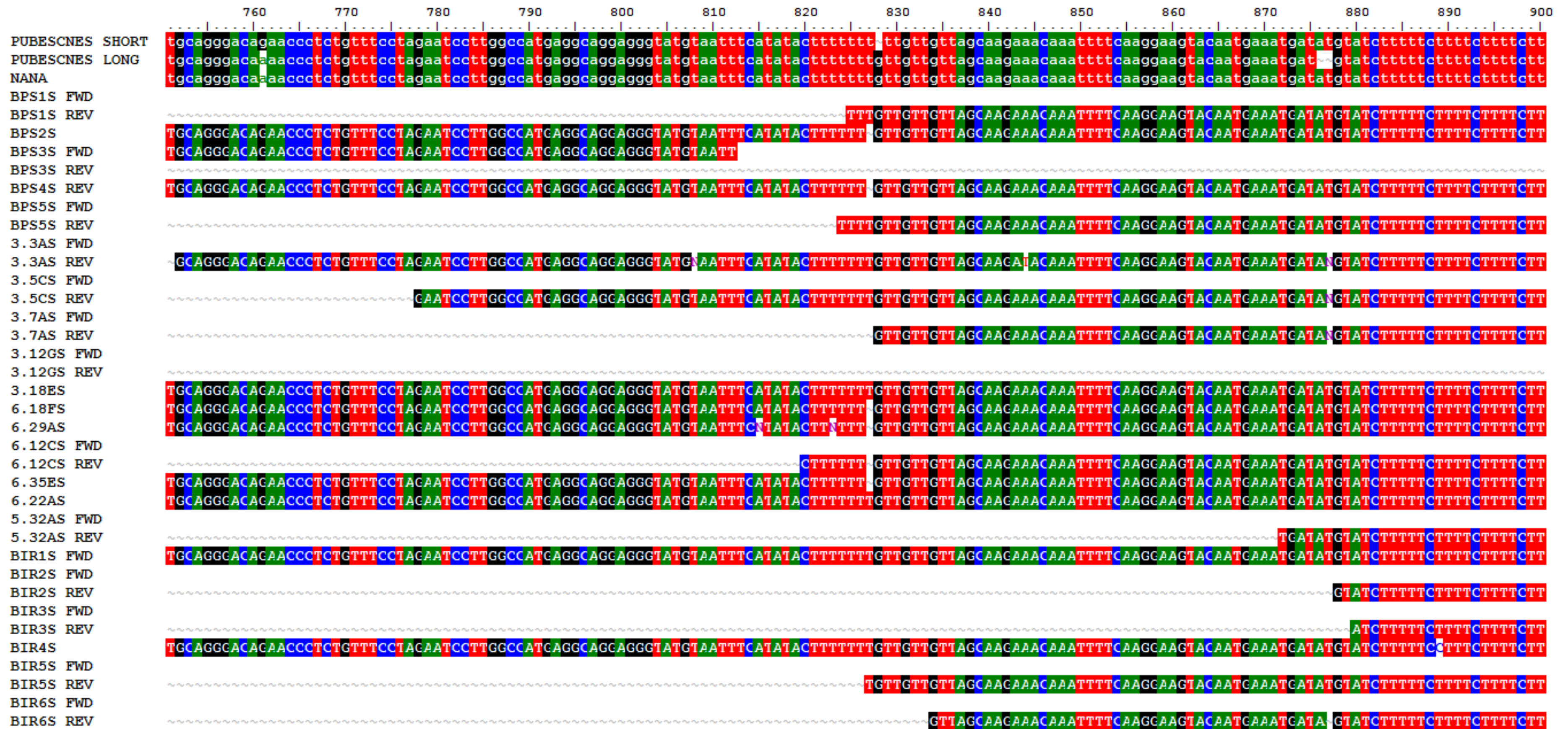

Fig. 3.2.3.8. Part 6. Full sequence of short alleles aligned with published long alleles and short allele.

```
                         910       920       930       940       950       960       970       980       990       1000      1010      1020      1030      1040      1050
PUBESCNES SHORT   ttaatgaataaaaaaggatcgttgagagcgttggtgagggcgtgacagatctggaagcaggcgatcatgtcctgccagtgttcactggggaatgcaaagaatgtagacactgtaaatcagaagagagcaacttgtgtgagctgctgaggat
PUBESCNES LONG    ttaatgaataaaaaaggatcgttgagagcgttggtgagggcgtgacagatctgaaagcaggcgatcatgtcctgccagtgttcactggggaatgcaaagaatgtagacactgtaaatcagaagagagcaacctgtgtgagctgctgaggat
NANA              ttaatgaataaaaaaggatcgttgagagcgttggtgagggcgtgacagatctgaaagcaggcgatcatgtcctgccagtgttcactggggaatgcaaagaatgtagacactgtaaatcagaagagagcaacctgtgtgagctgctgaggat
BPS1S FWD
BPS1S REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGNAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACNTGTGTGAGCTGCTGAGGAT
BPS2S             TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGGAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BPS3S FWD
BPS3S REV         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
BPS4S REV         GTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BPS5S FWD
BPS5S REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
3.3AS FWD
3.3AS REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
3.5CS FWD
3.5CS REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
3.7AS FWD
3.7AS REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
3.12GS FWD
3.12GS REV        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
3.18ES            GTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
6.18FS            TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGGAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACTTGTGTGAGCTGCTGAGGAT
6.29AS            TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACNTGTGTGAGCTGCTGAGGAT
6.12CS FWD
6.12CS REV        TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGNAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACNTGTGTGAGCTGCTGAGGAT
6.35ES            TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGGAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
6.22AS            TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
5.32AS FWD
5.32AS REV        TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BIR1S FWD         GTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BIR2S FWD
BIR2S REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGNAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACNTGTGTGAGCTGCTGAGGAT
BIR3S FWD
BIR3S REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGNAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BIR4S             TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BIR5S FWD
BIR5S REV         TTAATGAATAAAAAAGGATCGTTGAGAGCNTTGGTGAGGGCGTGACAGATCTGNAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACCTGTGTGAGCTGCTGAGGAT
BIR6S FWD
BIR6S REV         TTAATGAATAAAAAAGGATCGTTGAGAGCGTTGGTGAGGGCGTGACAGATCTGAAAGCAGGCGATCATGTCCTGCCAGTGTTCACTGGGGAATGCAAAGAATGTAGACACTGTAAATCAGAAGAGAGCAACNTGTGTGAGCTGCTGAGGAT
```

Fig. 3.2.3.9. Part 7. Full sequence of short alleles aligned with published long alleles and short allele.

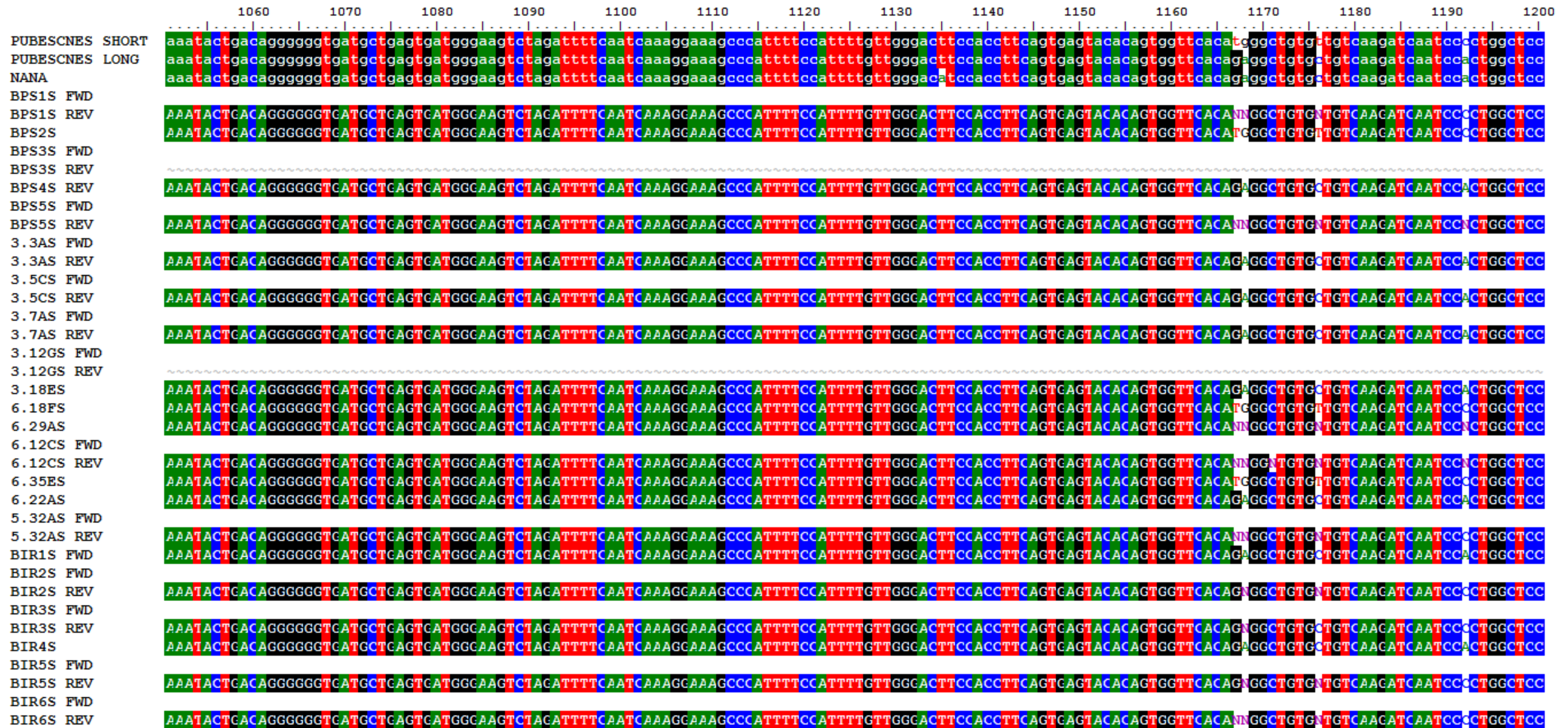

Fig. 3.2.3.10. Part 8. Full sequence of short alleles aligned with published long alleles and short allele.

```
                 1210       1220       1230       1240       1250       1260       1270       1280       1290       1300       1310       1320       1330       1340       1350
PUBESCNES SHORT  tcttgacaaagtttgtatcctcagttgtggaatctccacaggtactaagaaacatgagaattcctgagactagcttgtatattttctgttaattaataattttttatagaattgattttgagtatttgctgtctaggccttggtgccac
PUBESCNES LONG   tcttgacaaagtttgtatcctcagttgtggaatctccacaggtaccaagaaacatgagaattccagagactagcttgtatattttctgttaattaataattttttatagaattgattttgagtatttgttgtctaggccttggtgccac
NANA             tcttgacaaagtttgtatcctcagttgtggaatctccacaggtaccaagaaacatgagaattccagagactagcttgtatattttctgttaattaataattttttatagaattgattttgagtatttgttgtctaggccttggtgccac
BPS1S FWD
BPS1S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCNGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGNTGTCTAGGCCTTGGTGCCAC
BPS2S            TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACTAAGAAACATGAGAATTCCTGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGCTGTCTAGGCCTTGGTGCCAC
BPS3S FWD
BPS3S REV        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
BPS4S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
BPS5S FWD
BPS5S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
3.3AS FWD
3.3AS REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
3.5CS FWD
3.5CS REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
3.7AS FWD
3.7AS REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATNTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
3.12GS FWD
3.12GS REV       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
3.18ES           TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
6.18FS           TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACTAAGAAACATGAGAATTCCTGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGCTGTCTAGGCCTTGGTGCCAC
6.29AS           TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACNAAGAAACATGAGAATTCCNGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
6.12CS FWD
6.12CS REV       TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCNGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
6.35ES           TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACTAAGAAACATGAGAATTCCTGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGCTGTCTAGGCCTTGGTGCCAC
6.22AS           TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
5.32AS FWD
5.32AS REV       TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGNTGTCTAGGCCTTGGTGCCAC
BIR1S FWD        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATCA
BIR2S FWD
BIR2S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCNGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGANTATTTGNTGTCTAGGCCTTGGTGCCAC
BIR3S FWD
BIR3S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
BIR4S            TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
BIR5S FWD
BIR5S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCNGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGTTGTCTAGGCCTTGGTGCCAC
BIR6S FWD
BIR6S REV        TCTTGACAAAGTTTGTATCCTCAGTTGTGGAATCTCCACAGGTACCAAGAAACATGAGAATTCCAGAGACTAGCTTGTATATTTTCTGTTAATTAATAATTTTTTATAGAATTGATTTTGAGTATTTGNTGTCTAGGCCTTGGTGCCAC
```

Fig. 3.2.3.11. Part 9. Full sequence of short alleles aligned with published long alleles and short allele.

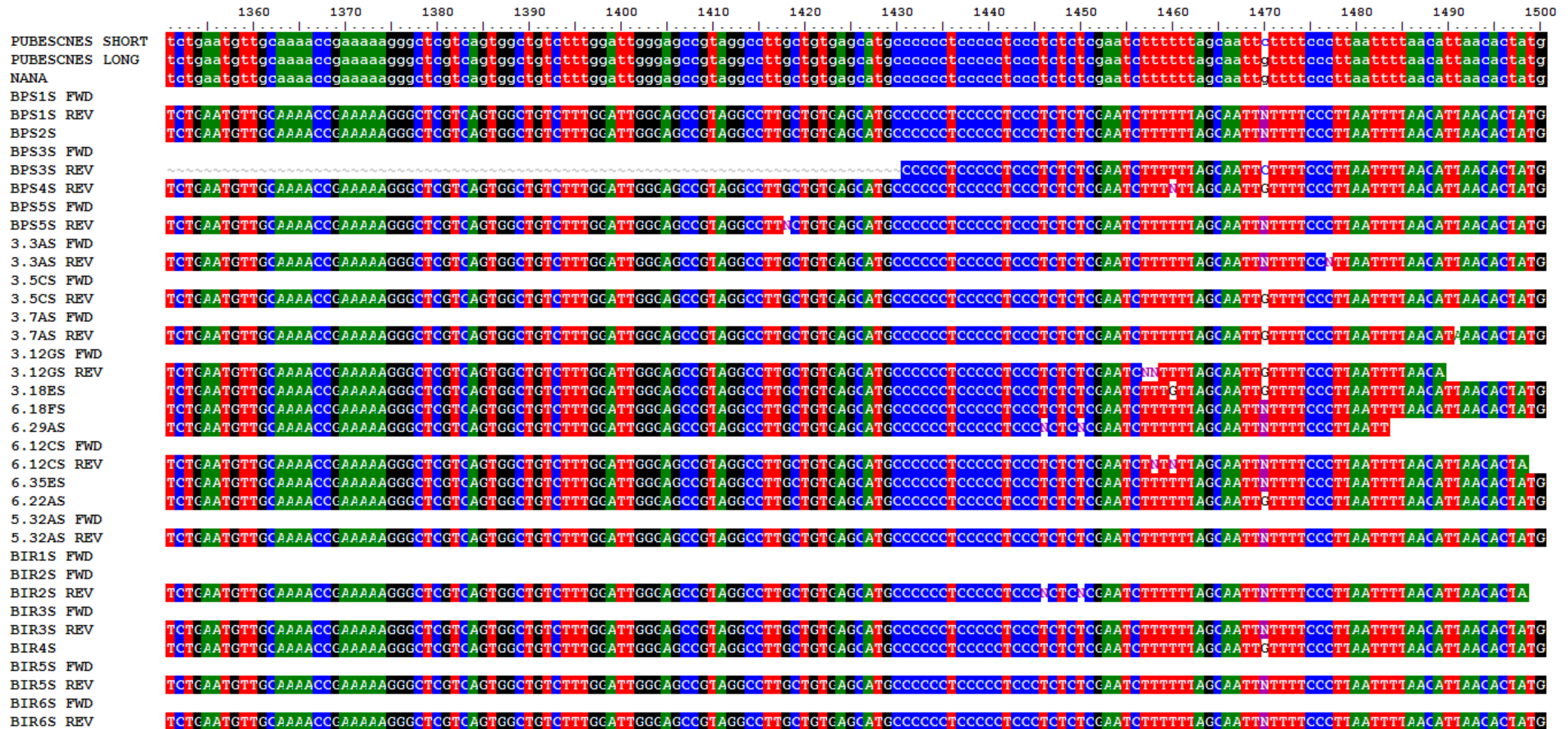

Fig. 3.2.3.12. Part 10. Full sequence of short alleles aligned with published long alleles and short allele.

```
                              1510       1520       1530
                    ....|....|....|....|....|....|....|....
PUBESCNES SHORT     gttctttgttgtgattaggctgccgagggggcca
PUBESCNES LONG      gttctttgttgtgattaggctgccgagggggcca
NANA                gttctttgttgtgattaggctgccgagggggcca
BPS1S FWD
BPS1S REV           GT
BPS2S               GTTCTT
BPS3S FWD
BPS3S REV           GTTCTT
BPS4S REV           GTTCTT
BPS5S FWD
BPS5S REV           GTTCTT
3.3AS FWD
3.3AS REV           GTTCTT
3.5CS FWD
3.5CS REV           GTTCTT
3.7AS FWD
3.7AS REV           GTTCTT
3.12GS FWD
3.12GS REV
3.18ES              GTTCTT
6.18FS              GT
6.29AS
6.12CS FWD
6.12CS REV
6.35ES              GT
6.22AS              GTTCTT
5.32AS FWD
5.32AS REV          GTTCTT
BIR1S FWD
BIR2S FWD
BIR2S REV
BIR3S FWD
BIR3S REV           GTTCTT
BIR4S               GTTCTT
BIR5S FWD
BIR5S REV           GTTCTT
BIR6S FWD
BIR6S REV           GTTCTT
```

Fig. 3.2.3.13. Part 11. Full sequence of short alleles aligned with published long alleles and short allele.

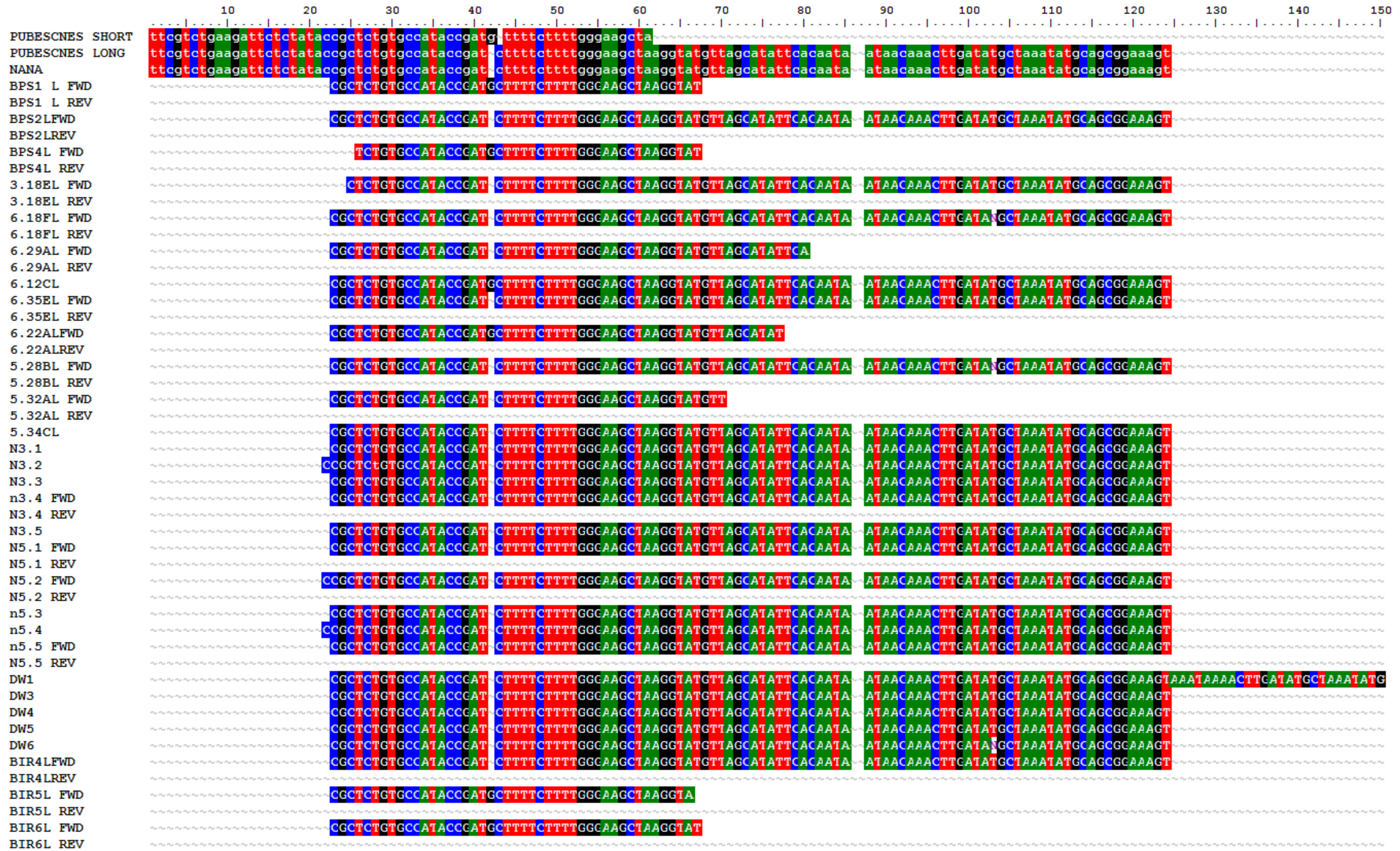

Fig. 3.2.3.14. Part 1. Full sequence of long alleles aligned with published long alleles and short allele.

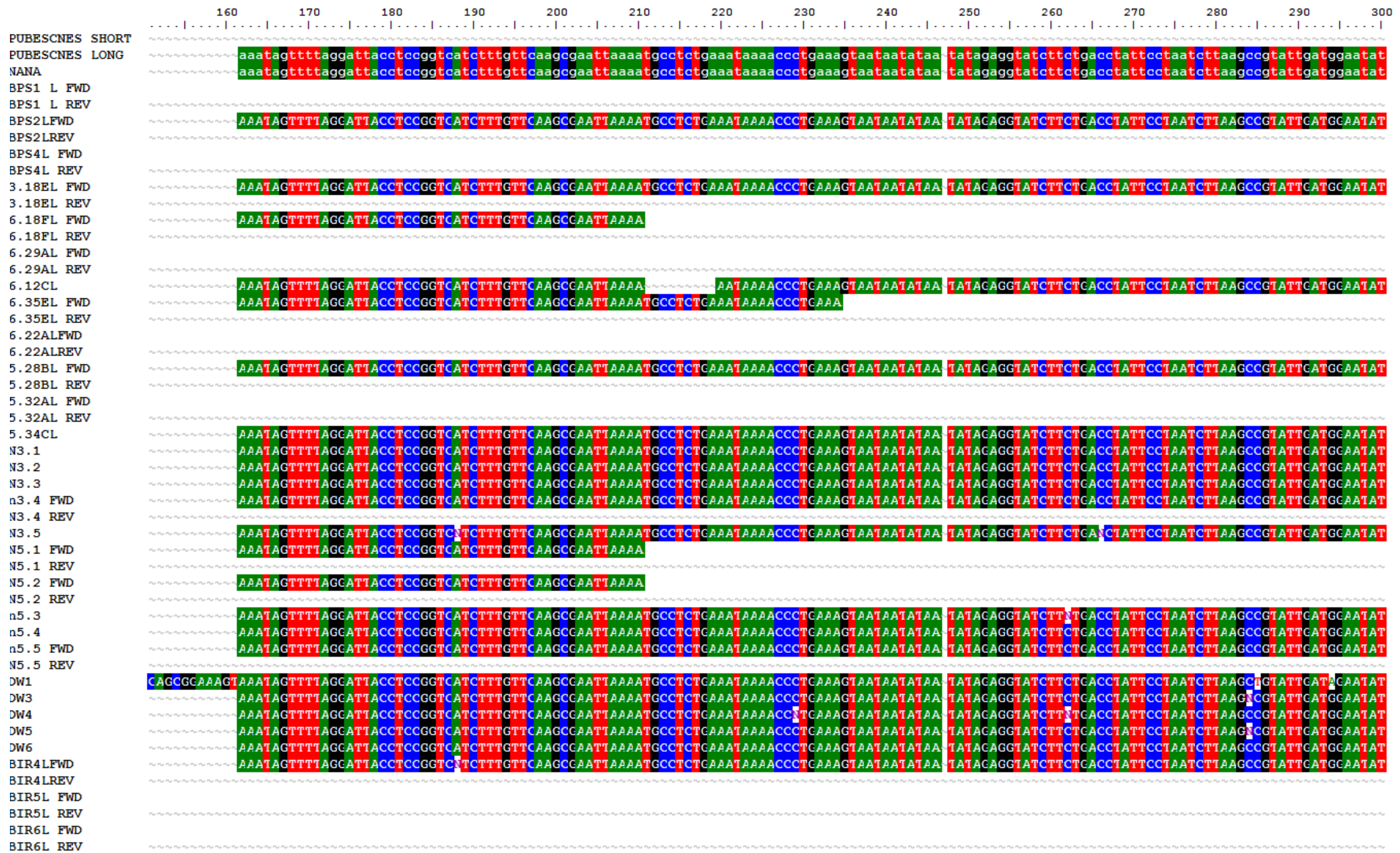

Fig. 3.2.3.15. Part 2. Full sequence of long alleles aligned with published long alleles and short allele.

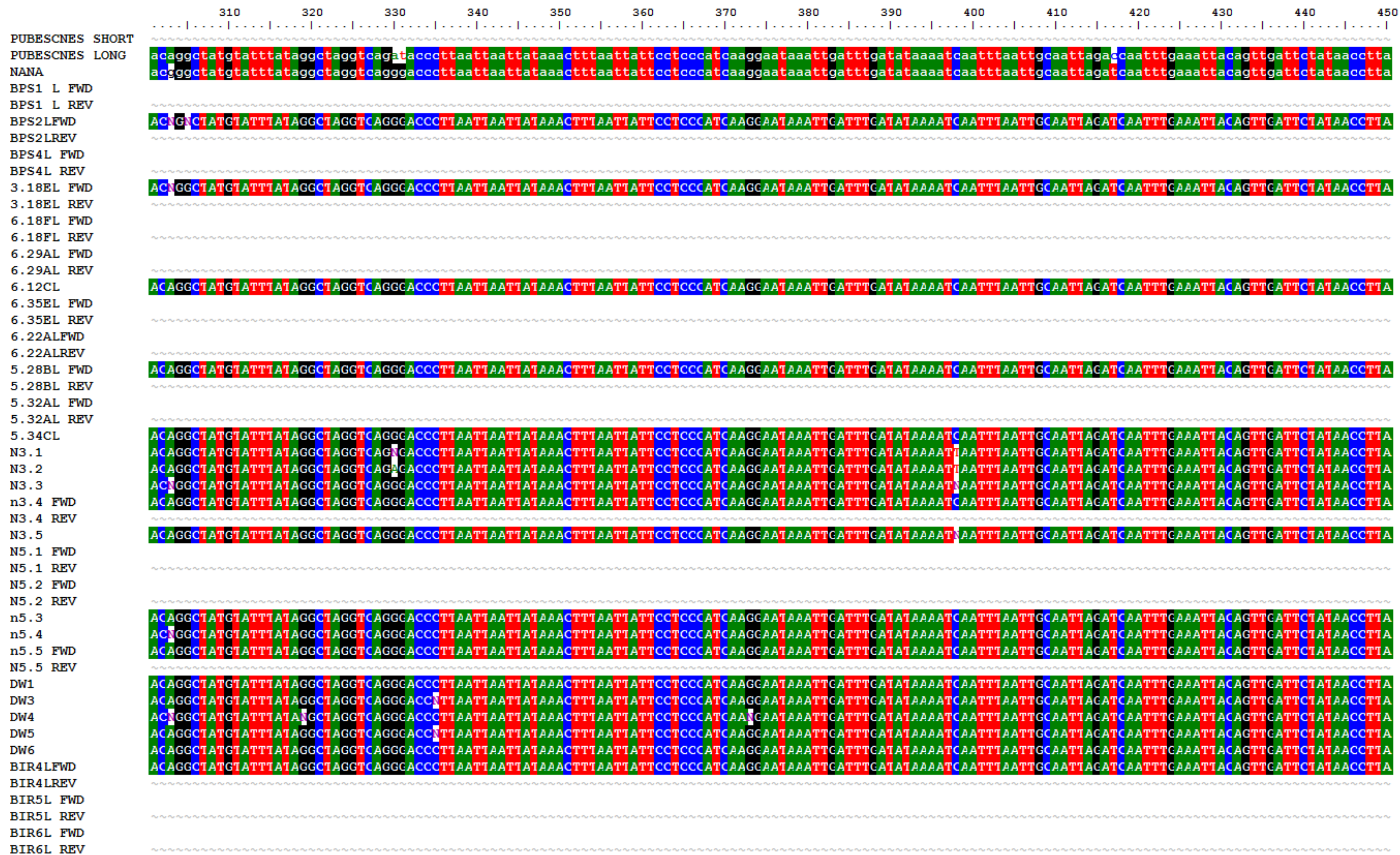

Fig. 3.2.3.16. Part 3. Full sequence of long alleles aligned with published long alleles and short allele.

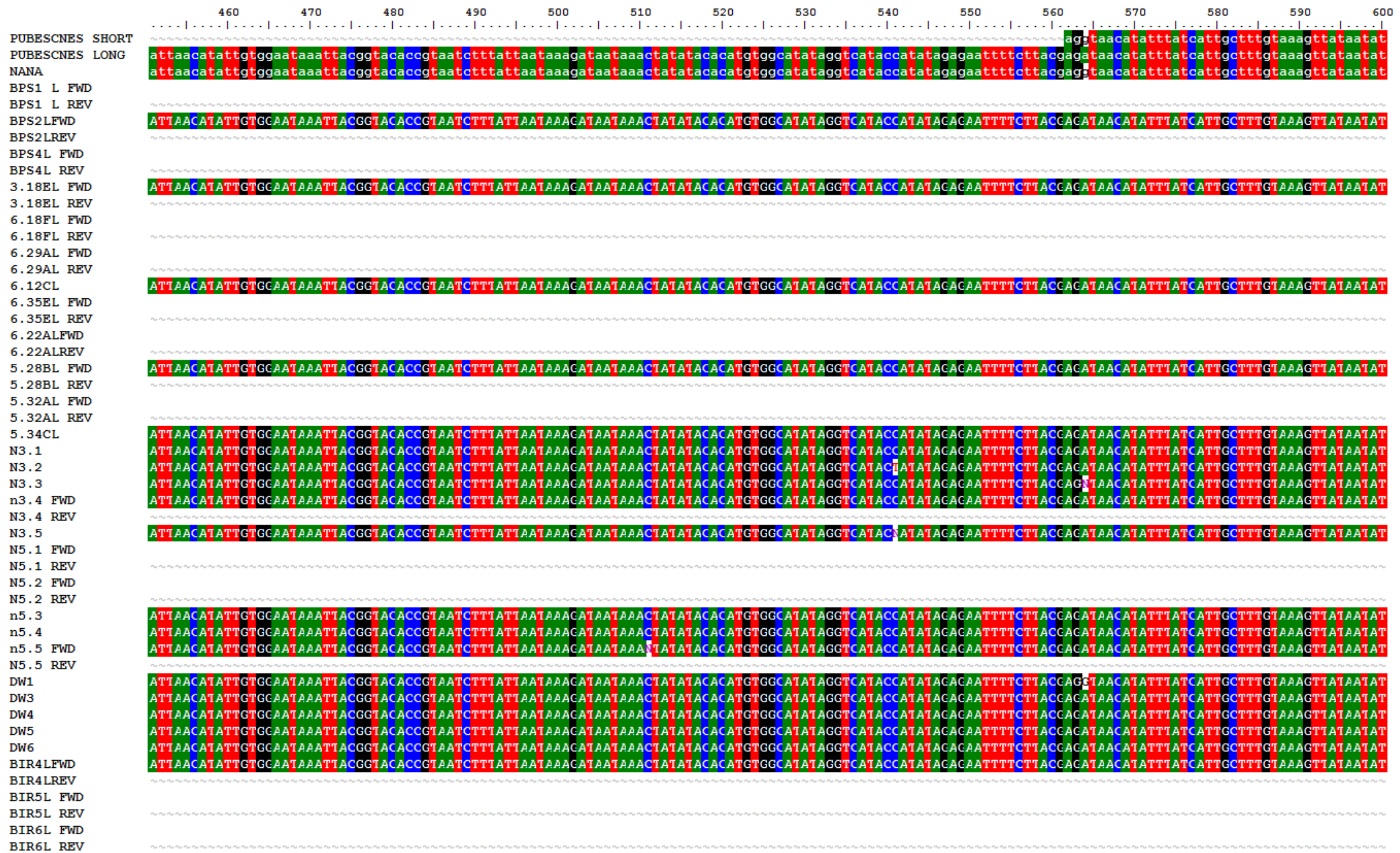

Fig. 3.2.3.17. Part 4. Full sequence of long alleles aligned with published long alleles and short allele.

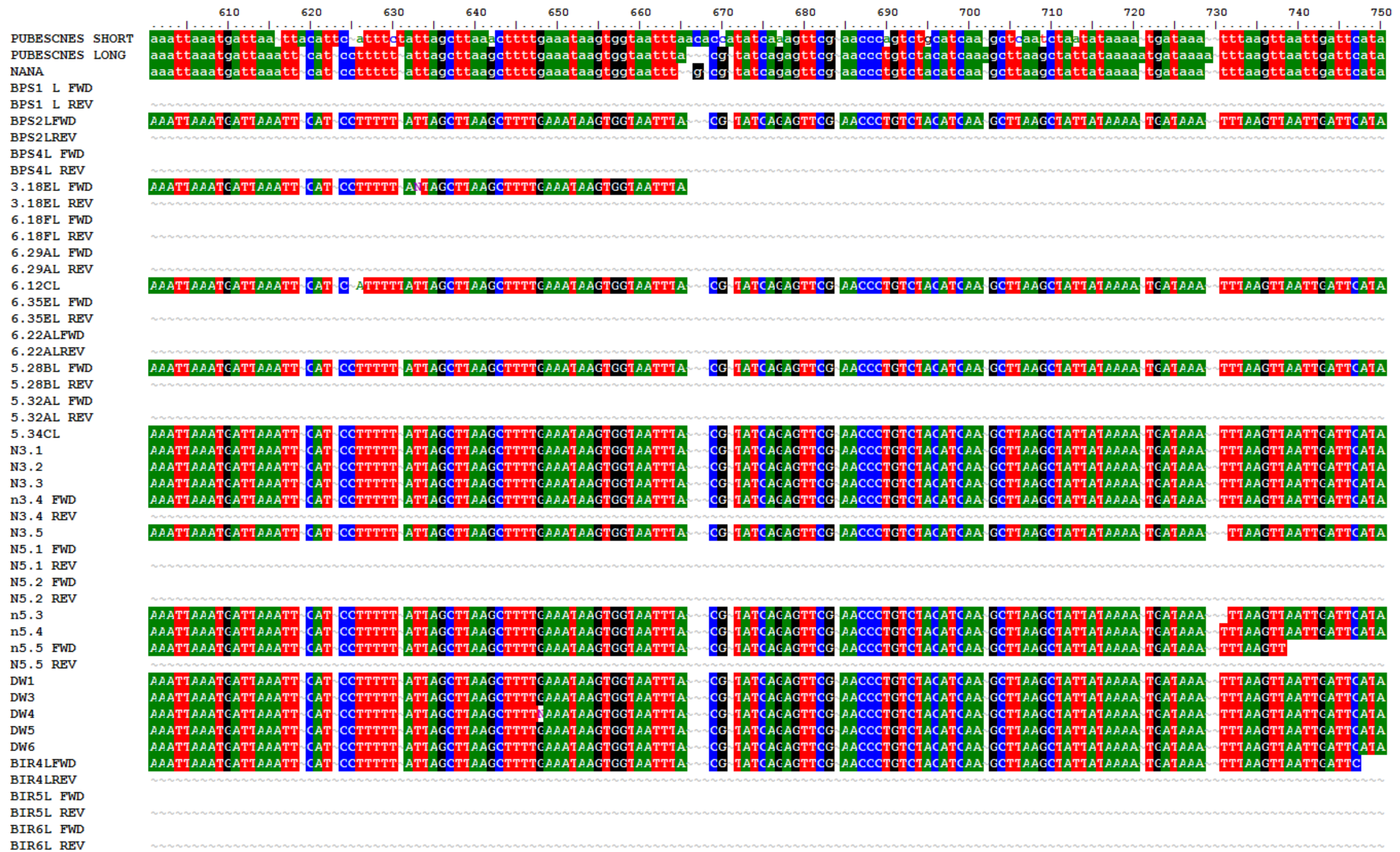

Fig. 3.2.3.18. Part 5. Full sequence of long alleles aligned with published long alleles and short allele.

```
                       760       770       780       790       800       810       820       830       840       850       860       870       880       890       900
PUBESCNES SHORT  ttgatatatagaacattaaatcctagatcgatttctattctgcagggacagaaccctctgtttcctagaatccttggccatgaggcaggagggtatgtaatttcatatactttttttt~ttgttgttagcaagaaacaaattttcaaggaa
PUBESCNES LONG   ttgatgtatagaacattaaatcctagatcgatttctattctgcagggacaaaaccctctgtttcctagaatccttggccatgaggcaggagggtatgtaatttcatatacttttttttgttgttgttagcaagaaacaaattttcaaggaa
NANA             ttgatgtatagaacattaaatcctagatcgatttctattctgcagggacaaaaccctctgtttcctagaatccttggccatgaggcaggagggtatgtaatttcatatacttttttttgttgttgttagcaagaaacaaattttcaaggaa
BPS1 L FWD
BPS1 L REV
BPS2LFWD         TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACT
BPS2LREV
BPS4L FWD
BPS4L REV        TTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
3.18EL FWD
3.18EL REV       TTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
6.18FL FWD
6.18FL REV       TTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
6.29AL FWD
6.29AL REV       TTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
6.12CL           TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
6.35EL FWD
6.35EL REV       TTTTTTGTTGTTGTTAGCAAGAAACAAATTT~CAAGGAA
6.22ALFWD
6.22ALREV        AGAAACAAATTTTCAAGGAA
5.28BL FWD       TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCA
5.28BL REV
5.32AL FWD
5.32AL REV       TTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
5.34CL           TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
N3.1             TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
N3.2             TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
N3.3             TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
n3.4 FWD         TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATG
N3.4 REV
N3.5             TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
N5.1 FWD
N5.1 REV         TTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
N5.2 FWD
N5.2 REV         TTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
n5.3             TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
n5.4             TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
n5.5 FWD
N5.5 REV
DW1              TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
DW3              TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
DW4              TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
DW5              TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
DW6              TTGATGTATAGAACATTAAATCCTAGATCGATTTCTATTCTGCAGGGACAAAACCCTCTGTTTCCTAGAATCCTTGGCCATGAGGCAGGAGGGTATGTAATTTCATATACTTTTTTTTGTTGTTGTTAGCAAGAAACAAATTTTCAAGGAA
BIR4LFWD
BIR4LREV
BIR5L FWD
BIR5L REV        GTTGTTAGCAAGAAACAAATTTTCAAGGAA
BIR6L FWD
BIR6L REV
```

Fig. 3.2.3.19. Part 6. Full sequence of long alleles aligned with published long alleles and short allele.

Fig. 3.2.3.20. Part 7. Full sequence of long alleles aligned with published long alleles and short allele.

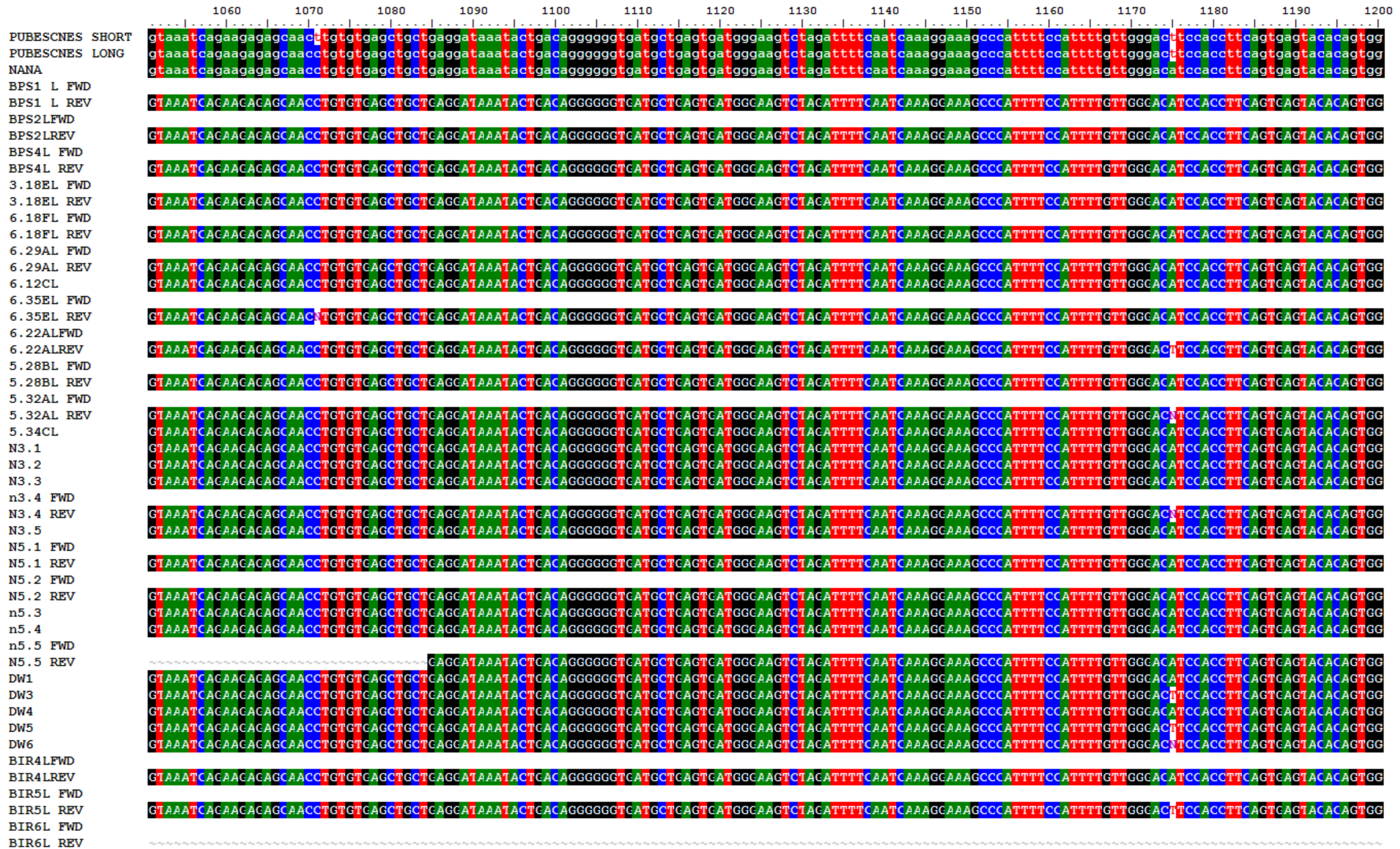

Fig. 3.2.3.21. Part 8. Full sequence of long alleles aligned with published long alleles and short allele.

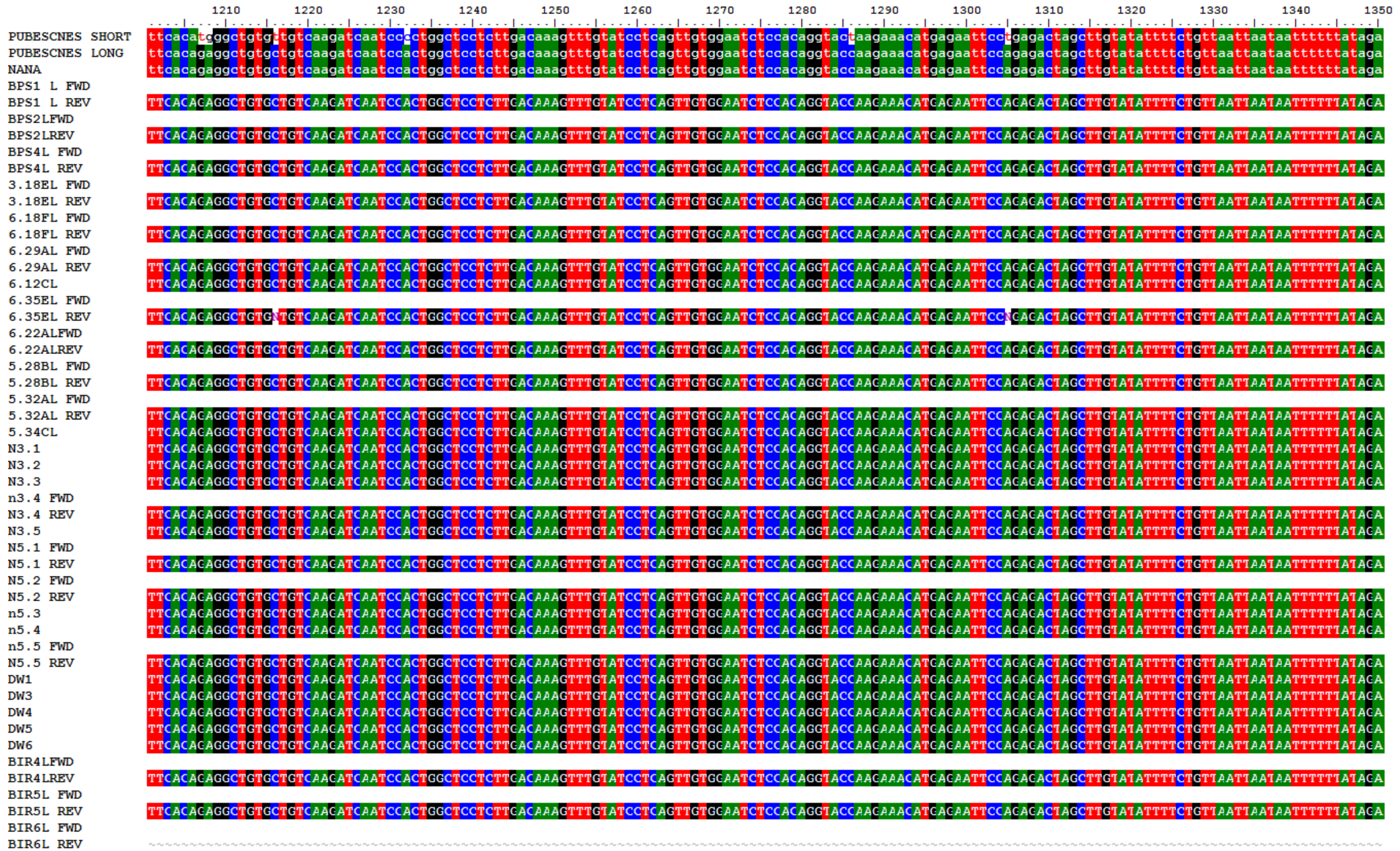

Fig. 3.2.3.22. Part 9. Full sequence of long alleles aligned with published long alleles and short allele.

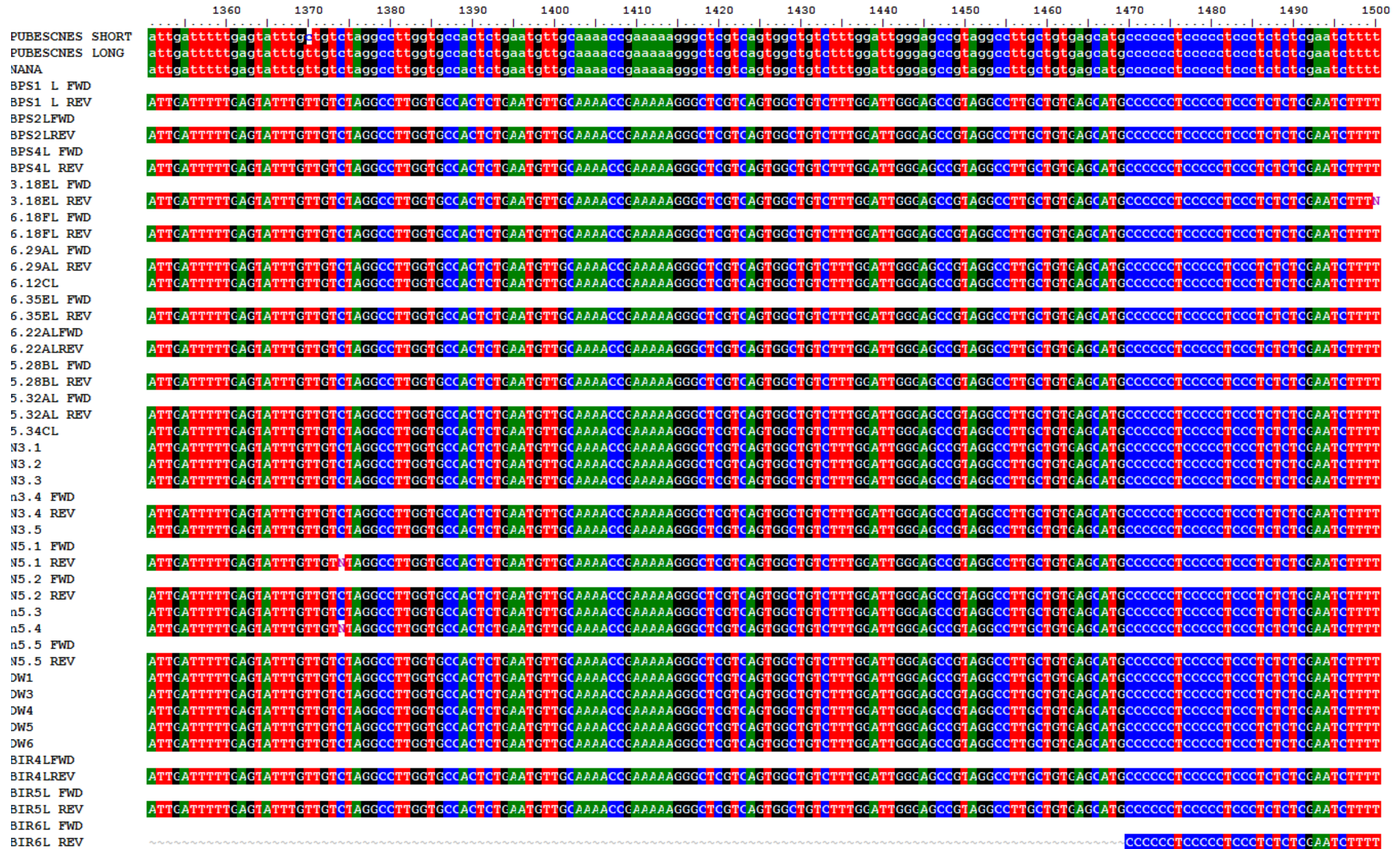

Fig. 3.2.3.23. Part 10. Full sequence of long alleles aligned with published long alleles and short allele.

```
                        1510      1520      1530      1540      1550      1560      1570
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....
PUBESCNES SHORT     ttagcaattcttttcccttaattttaacattaacactatggttctttgttgtgattaggctgccgagggggcca
PUBESCNES LONG      ttagcaattgttttcccttaattttaacattaacactatggttctttgttgtgattaggctgccgagggggcca
NANA                ttagcaattgttttcccttaattttaacattaacactatggttctttgttgtgattaggctgccgagggggcca
BPS1 L FWD
BPS1 L REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
BPS2LFWD
BPS2LREV            TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
BPS4L FWD
BPS4L REV           TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
3.18EL FWD
3.18EL REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
6.18FL FWD
6.18FL REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
6.29AL FWD
6.29AL REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
6.12CL              TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
6.35EL FWD
6.35EL REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
6.22ALFWD
6.22ALREV           TTAGCANTTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
5.28BL FWD
5.28BL REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTA
5.32AL FWD
5.32AL REV          TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
5.34CL              TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
N3.1                TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
N3.2                TTAGCAATTGTTTTCCCTTAATTTTAACATTAACAC
N3.3                TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
n3.4 FWD
N3.4 REV            TTAGCANTTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
N3.5                TTAGCAATTGTTTTCCNTTAATTTTAACATTAACACTATGGTTCTT
N5.1 FWD
N5.1 REV            TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
N5.2 FWD
N5.2 REV            TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATG
n5.3                TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
n5.4                TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCT
n5.5 FWD
N5.5 REV            TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
DW1                 TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGT
DW3                 TTAGCANTTGTTTTCCNTTAATTTTAACATTAACACTATGGTTCTT
DW4                 TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
DW5                 TTAGCANTTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
DW6                 TTAGCANTTGTTTTCCCTTAATTTTAACATTAACACTATGGT
BIR4LFWD
BIR4LREV            TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
BIR5L FWD
BIR5L REV           TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACTATGGTTCTT
BIR6L FWD
BIR6L REV           TTAGCAATTGTTTTCCCTTAATTTTAACATTAACACNATGGTTCTT
```

Fig. 3.2.3.24. Part 10. Full sequence of long alleles aligned with published long alleles and short allele.

3.2.4 Insertions, Deletions and Misalignments

For the long allele several samples show a sudden misalignment around 66-80 bp from the start: BIR5 forward (66 bp from the start) BPS1 forward (67 bp from the start) BPS4 forward (67 bp from the start) BIR6 forward (67 bp from the start) 5.32A forward (70 bp from the start) 6.22A forward (77 bp from the start) 6.29A forward (80 bp from the start) (Fig. 3.2.3.14-24). These samples also share the same banding pattern of a strong short allele band and a weaker long allele band BPS1, BPS4, (Fig. 3.2.2.4 (1, 4)), 5.32A (Fig. 3.2.2.4 (10)), 6.29A, 6.22A, (Fig. 3.2.2.2 (2,5)) BIR5 (Fig. 3.2.2.2 (1,5)) and BIR6 (Fig. 3.2.2.5 (1)), which indicates that for the long allele there are two possible alleles. Therefore it is likely that all of these samples have one allele which has an insertion or deletion around the 66-80 base pair mark and one which does not. This would suggest that around that base pair number these samples are likely to share at least one allele with an insertion or deletion, suggesting close relations between these samples, as supported by the similar banding pattern. It would also suggest these samples are more closely related to each other than those samples with the same banding patterns but without this sudden misalignment between 66-80 bp from the start. This misalignment occurs shortly after the point at which the short allele contains a deletion 61 bp from the start (Fig. 3.2.3.3). It is worth noting here that the sequence "AGGTA" is seen both at the beginning of the section which the short allele lacks in the long allele, and immediately after the section the short allele lacks in both the short allele and the published long *B. nana* sequence (the second G is replaced with an A in the published long *B. pubescens* sequence). Hence it could be proposed that the section missing in the short allele (or added in the long allele) occurs anywhere up to 66 bp

from the start depending on whether the missing section is considered to occur either before, part way through or after this sequence. The proximity of the insertion or deletion in one version of the long allele in these samples to the beginning of the section missing in the short allele is interesting. It is possible part of the short allele was also extracted and sequenced with these samples, however given similar groupings of samples with a sudden misalignment seen at other locations on the alleles this would seem unlikely. Although it could be co-incidental it could also indicate that this particular region is prone to insertions or deletions, as supported by the two base insertions seen in the short alleles of samples 3.7A and BIR6, 62 bp from the start (Fig. 3.2.3.3), the 36 base pair insertion seen 125 bp from the start with sample DW1 (long allele) (Fig. 3.2.3.14- 3.2.3.15) and the 10 base pair deletion seen in 6.12C (long allele) 210 bp from the start (Fig. 3.2.3.15). These insertions and deletions all occur within the section the short allele lacks. The fact that samples BIR5 forward (66 bp from the start) BPS1 forward (67 bp from the start) BPS4 forward (67 bp from the start) BIR6 forward (67 bp from the start) 5.32A forward (70 bp from the start) 6.22A forward (77 bp from the start) and 6.29A forward (80 bp from the start) (Fig. 3.2.3.14) do not become misaligned at exactly the same base pair number may also support this. Rather than being closely related and sharing the same insertion or deletion it is possible that BIR5 forward (66 bp from the start) (Fig. 3.2.3.14) has one version of an insertion or deletion, BPS1 forward (67 bp from the start), BPS4 forward (67 bp from the start) and BIR6 forward (67 bp from the start) (Fig. 3.2.3.14) share a second version, 5.32A forward (70 bp from the start) has a third version, 6.22A forward (77 bp from the start) a forth version and 6.29A forward (80 bp from the start) yet another version (Fig. 3.2.3.14).

Further along the long allele sequence samples 6.18F, N5.1 and N5.2 all have a sudden misalignment exactly 210 bp from the start (Fig. 3.2.3.15). This coincides with the 10 base pair deletion also seen 210 bp from the start, on the long allele of sample 6.12C (Fig. 3.2.3.15). This strongly suggests close relations between these individuals. From the banding pattern data, 6.12C is likely to have two short alleles and two long, or one long allele one short, suggesting it has one to two copies of the allele with a ten base pair deletion 210 bp from the start. 6.18F is likely to have thee long alleles and one short, suggesting it has one to two copies of the allele with the ten base pair and one to two copies of the allele without this deletion. N5.1 and N5.2 only have a long allele band and could have two to four copies of this long allele and hence contain anything from one to three copies of it with the deletion and one to three copies of it without this deletion. Due to this likely sharing of a deleted section in at least one long allele in these individuals, it is highly probable that these samples are closely related, most likely siblings or parents and offspring.

The sequence of one sample, BIR6, also shows a sequence misalignment for the reverse sequence of the long allele from base 1469 (Fig. 3.2.3.23). No other samples (including comparing long and short alleles) show any difference between one another at this point. Interestingly the sequence from sample BIR3 shows a sudden misalignment at exactly the same point but in the short allele (due to various insertions and deletions this is at base pair 1431 in the short allele sequences (Fig. 3.2.3.12)). This implies that an insertion or deletion not seen in any other sample has occurred in both a long and a short allele at the same point. This may seem unlikely however it could be possible since it occurs in an extremely repetitive region: “5’-CCCCCCTCCCCCTCCCTCTCTC-3’” and these samples may have coincidentally

developed an insertion or deletion in one allele at this point. Alternatively due to the highly repetitive nature of that section, reading errors either during sequencing or PCR may have occurred causing a misalignment. Another possibility is crossing over between the short and long allele in these samples, but this is unlikely considering the rarity of these events.

The short allele sequence of BPS1, 6.12C, 5.32A, BIR2, BIR5, BIR6 misaligns 575 bp from the start (Fig. 3.2.3.6). Although 575 bp may seem like a reasonable number after which natural degradation of sequence may occur it must be noted that short alleles have been aligned with long alleles. There is a sequence 459 bp long in the long allele that the short allele does not contain and hence 575 bp from the start is actually only approximately 116 bp from the start for a short allele. At this point of misalignment published long alleles for both *B. nana* and *B. pubescens* have an extra adenine nucleotide which the short allele lacks. However comparing all short ADH alleles from the Abisko samples it is evident that short alleles can have either two or three adenine nucleotides at this point: Samples BPS2, BPS3, 6.29A,6.18F and 6.35E have two adenine nucleotides at this point whilst BPS5, 3.3A, 3.5C, 3.7A, 3.12G, 3.18E, 6.22A, BIR1, BIR3 and BIR4 have three. It is therefore likely that either *B. pubescens* short ADH alleles from Punkaharju, Finland (Järvinen et al., 2004) (from which the published sequence originates) only ever have two adenine nucleotides at this point, or that the sample used in this publication happened to have two adenine nucleotides at this point. So the likelihood is that having either two or three adenine nucleotides at this point is not indicative of a short or long ADH allele in these species. Since samples BPS1, 6.12C, BIR5, BIR6 and 5.32A are likely to contain two versions of the short allele they are also likely to contain one version

with three adenine nucleotides and one with two. Sample BIR2 is likely to contain three short alleles and hence could contain one to two versions with three adenine nucleotides and one to two versions with two.

The short allele sequence of samples BIR3 and BPS5 (Fig. 3.2.3.6) share a misalignment 577-578 bp from the start. After this point of misalignment the published short allele has the sequence 5'-TACA-3' whereas the long alleles are lacking the adenine nucleotide and read 5'-TCA-3'. Short alleles from Abisko samples show both the 5'-TACA-3' sequence (samples BPS2, BPS3, 6.18F, 6.29A and 6.35E (Fig. 3.2.3.6)) and the TCA sequence (3.3A, 3.5C, 3.7A. 3.12G, 3.18E, 6.22A, BIR1 and BIR4 (Fig. 3.2.3.6)) indicating that this single point deletion is not indicative of the long allele but can vary between different versions of the same length allele. As shown earlier BIR3 and BPS5 are likely to have two or three copies of the short allele so it is likely that at least one copy has the sequence 5'-TACA-3' and at least one other copy has the sequence 5'-TCA-3' at this point, resulting in misalignment in sequence readouts after this base.

The area 614-650 bp from the start on the short allele consists of many insertions, deletions and substitutions (Fig. 3.2.3.7). Different versions of the short allele at this point are likely to cause mismatches due to the high levels of variance within this area. Sequence data for sample 3.12G becomes misaligned around 614 bp from the start at the point where many substitutions occur before a section of high variability at 625 bp from the start. Sequence data for sample 3.3A becomes misaligned 624 bp from the start just before the area where there is a high variance between the sequences. Over this area the published *B. pubescens* short allele sequence reads 5'-

AACACCATATCAAAGTTCG-3' the short alleles of BPS2, BPS3, 6.18F and 6.35E match this published sequence. The published *B. pubescens* long allele reads 5'-ACGTATCAGAGTTCG-3' and the published *B. nana* allele (long) reads 5'GCGTATCAGAGTTC3' none of the short allele samples match these two long allele sequences. However the samples from Abisko show a range of other sequences not seen in the published sequences at this point. 6.29A has the sequence 5'-AACACCATATCAAGGTTCG-3', 3.18E has the sequence5'-ACTTATCAGAGTTCA-3' and 3.5C, 3.7A, 6.22A, BIR1 and BIR4 have the sequence 5'-ACTTAAGCTTTTATTAACTT-3'. As shown earlier it is highly likely that sample 3.3A has either two or three short alleles and 3.12G has two short alleles. With the exception of this section of sequence for 6.29A (which only has one base substitution different from that of the published *B. pubescens* short allele) all of these potential short allele sequences are of different lengths (i.e. contain insertions or deletions). Therefore it is likely that 3.12G carries two short ADH alleles with different sequences of different lengths at this point and 3.3A carries between two to three different sequences at this point. To discover which versions they contain would however require cloning of samples into plasmids to separate the different versions of the alleles.

661 bp into the short allele sequence, sequence data from samples 3.5C and 3.7A become misaligned (Fig. 3.2.3.7). At this point the published short allele of *B. pubescens* and the ADH allele of *B. nana* (a long ADH allele) have two adenine nucleotides whilst the published long allele of *B. pubescens* has three adenine nucleotides. Although no other short allele samples from Abisko contain three adenine nucleotides at this point, at this length into the sequence most other samples

have been cut short due to misalignment leaving only ten other samples for comparison. It is likely that as with points 557 and 575 bp from the start on the short allele, the apparently distinguishing feature of an extra adenine nucleotide between different species or length alleles in published sequences is not exclusive to particular alleles/species and can vary between alleles. This misalignment cannot be due to contamination of the sequences with the long *B. pubescens* allele as misalignments would have occurred at an earlier stage if this was the case. Since these samples are likely to have between two and three copies of the short ADH allele it is likely that at least one copy in both samples contains three rather than two adenine nucleotides at this point.

As with the long allele, misalignments in the reverse sequence are rare relative to misalignments in the forward sequence. 3.12G contains a misalignment 1266 bp from the start (Fig. 3.2.3.11). This is in close proximity to many base substitutions but no insertions or deletions are seen in other samples (Fig. 3.2.3.11). 3.21G is likely to have two short alleles, it is therefore possible that one of these contains an insertion or deletion at this point. Another consideration is the high repetitiveness of this region (Fig. 3.2.3.11), which may cause problems with readability by either the PCR enzyme or sequencing techniques which may lead to this apparent misalignment.

Other insertions, deletions and substitutions have occurred throughout the sequences. Those which are not discussed here due to them being of a type which is not a cause of misalignment in sequence readouts for an individual, should be detected using a

nearest neighbour algorithm which can be used to produce dendrograms of relatedness of samples.

3.2.5 The Middle Band

A band midway between the long and short allele is seen on all gel images where both the short and long allele are present (BPS1, BPS2, BPS4, (Fig. 3.2.2.4 (1, 4)) 3.12G (Fig. 3.2.2.6 (1)), 3.18E, 5.32A (Fig. 3.2.2.4 (9-10)), 6.29A, 6.12C, 6.35E, 6.22A, (Fig. 3.2.2.2 (2-5)) BIR1 (Fig. 3.2.2.7), BIR4, BIR5 (Fig. 3.2.2.4 (14, 15)), BIR6 (Fig. 3.2.2.5 (1)), BPS5, 3.3A, 3.5C, 3.7A, (Fig. 3.2.2.4 (5-8)) BIR2, BIR3 (Fig. 3.2.2.4 (12-13)), 6.18F (Fig. 3.2.2.2 (1) and BPS2 (Fig. 3.2.2.4 (2))). This band has not previously been mentioned by individuals using the same primers in the same species (Järvinen et al., 2004) and hence the identity of this middle band was investigated.

Sequences of this middle band from twelve individuals all followed a similar pattern: A short section of the forward sequence was clear up until the point at which the long and the short allele differ significantly (i.e. the 62-521 base sequence (Fig. 3.2.3.3 – 3.2.3.6) absent in the short allele) at which point the sequence became misaligned. The reverse sequence did not show misalignment and had clear well defined peaks. This reverse sequence always matched the sequence of the respective sample for both its long and short allele. The reverse sequence is highly conserved between samples and allele lengths (Fig. 3.2.3.21-24 & Fig. 3.2.3.11 – 3.2.3.13). This band could therefore be another “middle” ADH allele, containing half of the insertion seen in the long allele rather than the full 62-521 base sequence (all short 1-

4). However if this was a properly amplified ADH allele there would need to be at least two different alleles with insertions/deletions at the point at which the short allele and long allele differ (i.e. the 62-521 base sequence (Fig. 3.2.3.3 – 3.2.3.6)) to cause the sequence misalignment. Since there is no evidence that a ploidy greater than tetraploid is possible in this species (Anamthawat-Jónsson & Tomasson, 1990; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2010) this would mean that samples with a middle allele could only have one long allele and one short allele. This cannot be the case, because the different banding patterns seen show varying intensities of short and long allele bands indicating different numbers of copies of those alleles and sequence data for many long and short alleles contain a mismatch in reading of sequences, likely due to insertions or deletions in one copy of the allele meaning that more than one long or short allele is present. There is also no evidence of this middle allele in any samples containing only one length of the ADH allele. If this was another length of allele it would be expected to also occur in samples other that just those which contain both the long and short allele.

Alternatively, it is possible that another gene entirely is being amplified by the primers. The misalignment could be due to every middle band consisting of four different sequences in each tetraploid, three in each triploid or two in each diploid. However this seems highly unlikely, if this was the case it would be expected that some of the samples showing only one allele would also show a middle band, which they do not. It would be also be highly unlikely to sit exactly between the long and the short allele. The sequence data also indicates that this is some form of ADH allele given that the reverse sequence matches the long and short allele reverse

sequences, and the forward sequence matches both long and short alleles up to the 62-521 base (Fig. 3.2.3.3 – 3.2.3.6) sequence difference between the long and short allele.

Given that this band matches the ADH alleles yet cannot itself be another ADH allele length and that it occurs in all samples containing both long and short alleles, but not in samples containing only one length allele, it would strongly suggest this band is some form of artefact due to the presence of both long and short alleles in the PCR mix. Since this band has a weight exactly half way between that of the long and short allele and the sequence becomes misaligned at the point of the large insertion/deletion difference between the long and short allele, it would strongly indicate that this band is a combination of one strand of a long ADH allele combined with one strand of a short ADH allele. There is no mention of this phenomenon in (Järvinen et al., 2004) however it is worth noting that no gel runs are shown and it is possible that either the middle band was ignored, their samples were sufficiently different from one another or that due to using a very different PCR program and enzyme this different strand length annealing did not occur. Irrespective of the origins of this middle band it is evidently some form of artefact rather than an actual ADH allele, so it will not be discussed further.

3.2.6 Dendrograms of Similarity

Due to misalignments and lack of clearly defined sequence ends, sections of sequence were used for analysis where the majority of samples had a section of forward or reverse sequence. Any gaps not present in that particular clip were

removed (e.g. the gap left when aligning short alleles with long alleles). *Corylus avellana* ADH sequence from (Järvinen et al., 2004) was used as the outgroup. The system used only produces bifurcating dendrograms therefore simple “stepping” of samples does not suggest any degree of relatedness and could be a proxy for a three point or greater node, for example (Fig. 3.2.6.3) does not suggest that DW1 is more closely related to DW6 than to 6.12C. Hence groups had to show strong evidence of grouping, with their own branch, before being considered as a group. Dendrograms in which branch length does not represent distances had to be used due to the close similarity of samples producing trees which were difficult to read due to scale effects.

A section of forward sequence from the short alleles was produced from base 23 to 275 (Fig. 3.2.3.3 – 3.2.3.6) the two published long allele sequences were excluded due to the section of forward sequence crossing the gap between bases 62-521(Fig. 3.2.3.3 – 3.2.3.6). A dendrogram based on similarity of samples produced two main groups. BIR4, 6.22A, 3.7A, BIR1 and 3.5C formed a group which formed part of a larger group with samples 3.3A, BPS5, 3.12G, 6.12C, 6.18F, BIR3, BIR5, BIR6 and 6.35E. BIR2, 5.32A, the published *B. pubescens* short allele, BPS1, BPS2, BPS3, 3.18E and 6.29A did not group (Fig. 3.2.6.1.). This would suggest close similarity between samples BIR4 6.22A, 3.7A, BIR1 and 3.5C, and in turn similarity between this group and samples 3.3A, BPS5, 3.12G, 6.12C, 6.18F, BIR3, BIR5, BIR6 and 6.35E. This forward section is a very short sequence of 95 bp with very few differences hence an analysis of similarity from such a short sequence is weak at best.

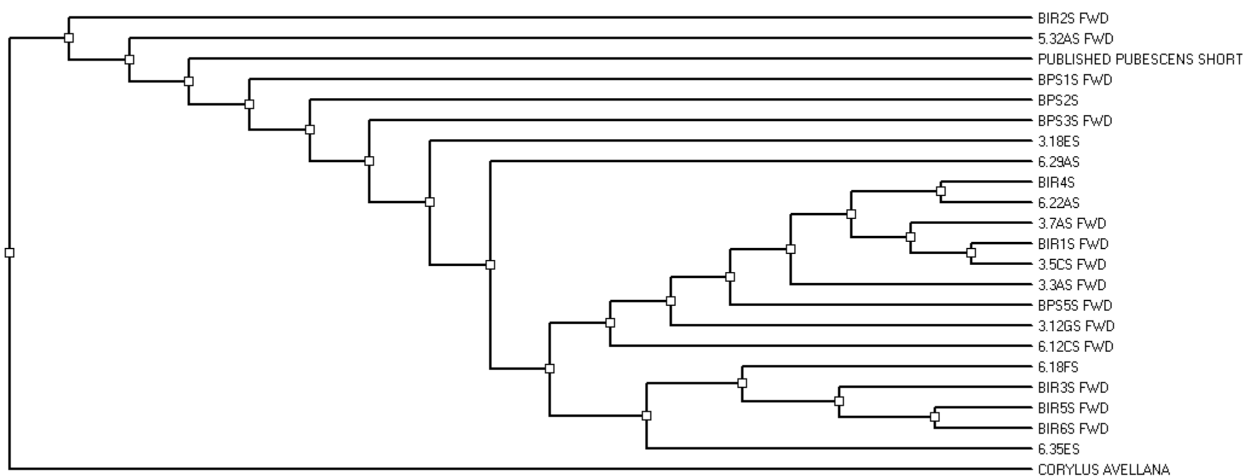

Fig. 3.2.6.1. Dendrogram of relatedness based on partial ADH genes for a section, based on where the majority of samples have data, of the forward sequence of the short allele.

A reverse section of sequence of the short alleles was produced 1268- 1483 (all short 9-10) this section of sequence is much longer than the forward sequence being 600 bp long and therefore should produce more reliable results. Differences between sequences are still however minimal (bases 1268-1483 (Fig. 3.2.3.11 – 3.2.3.12)), both published long alleles are well nested in the tree with the short alleles (Fig. 3.2.6.2). The published *B. nana* (long) allele and the short alleles of BIR4, BPS4 and 3.18E all form a closely related group which in turn groups with samples 3.7A and 3.3A to form a larger slightly less related group (Fig. 3.2.6.2). Another group, with less relation to one another than samples in the previously mentioned group, forms between samples 6.18F, 6.35E, BPS2 and the published *B. pubescens* long allele. This data suggests that there is no difference between long and short alleles in this section of the reverse sequence. Variation between different length alleles is no greater than between different versions of the same length allele and some short alleles such as BIR4, BPS4, 3.18E, 3.7A and 3.3A show greater similarity to the published section of sequence for *B. nana* ADH (a long allele) than to the *B. pubescens* short allele.

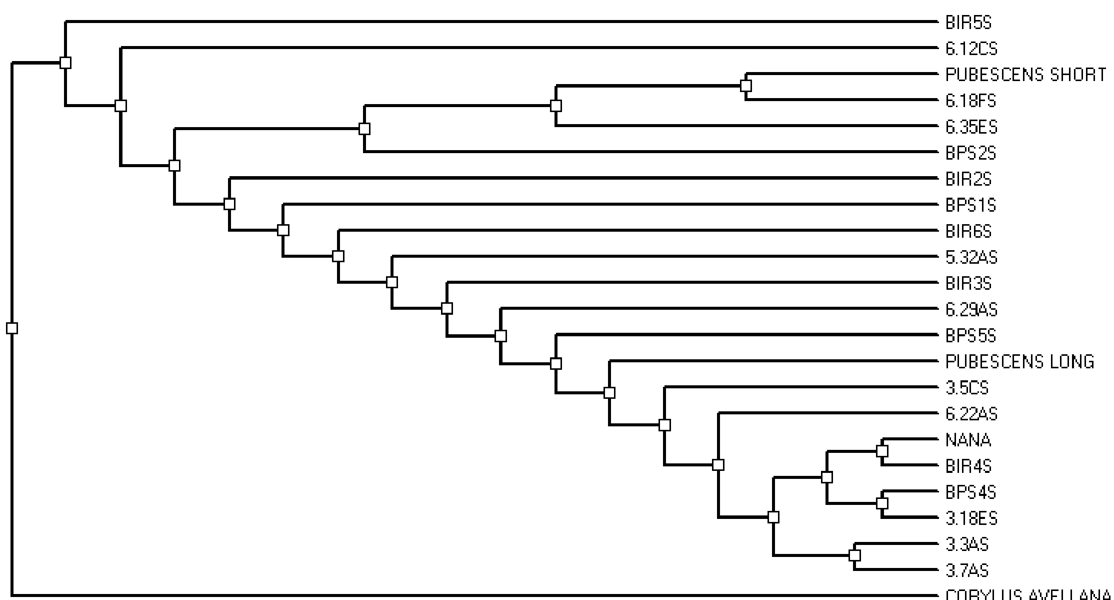

Fig. 3.2.6.2. Dendrogram of relatedness based on partial ADH genes for a section, based on where the majority of samples have data, of the reverse sequence of the short allele.

The two different sections of the sequence produce very different similarities between samples. The forward section suggests a relation between BIR4, 6.22A, 3.7A, BIR1 and 3.5C and at a weaker level with 3.3A, BPS5, 3.12G, 6.12C, 6.18F, BIR3, BIR5, BIR6 and 6.35E. The reverse section however suggests relations between samples 6.18F, 6.35E, BPS2 and the published *B. pubescens* long allele, another set of relations between BIR4, BPS4 and 3.18E and at a weaker level 3.7A and 3.3A. These are obviously very different groups of relatedness, with only samples BIR4, 3.3A and 3.7A and samples 6.18E and 6.35E consistently showing relatedness to one another. Taking this into account, only the similarity between BIR4, 3.3A and 3.7A and samples 6.18E and 6.35E can be regarded as reliable.

The nearest neighbour dendrogram for a forward section of the long allele (Fig. 3.2.6.3) bases 26-210 (Fig. 3.2.3.14-15) shows regular steps between samples. Since

the dendrogram bifurcates the only conclusion that can be drawn from this tree is that all of these sequences are equally similar. It would also suggest the algorithm does not attach high significance to insertions. DW1 contains a large insertion yet is still nested well within the tree. Again this is a relatively short section of sequence being 190 bp long and it is likely that there are insufficient differences for a reliable tree of similarity to be produced.

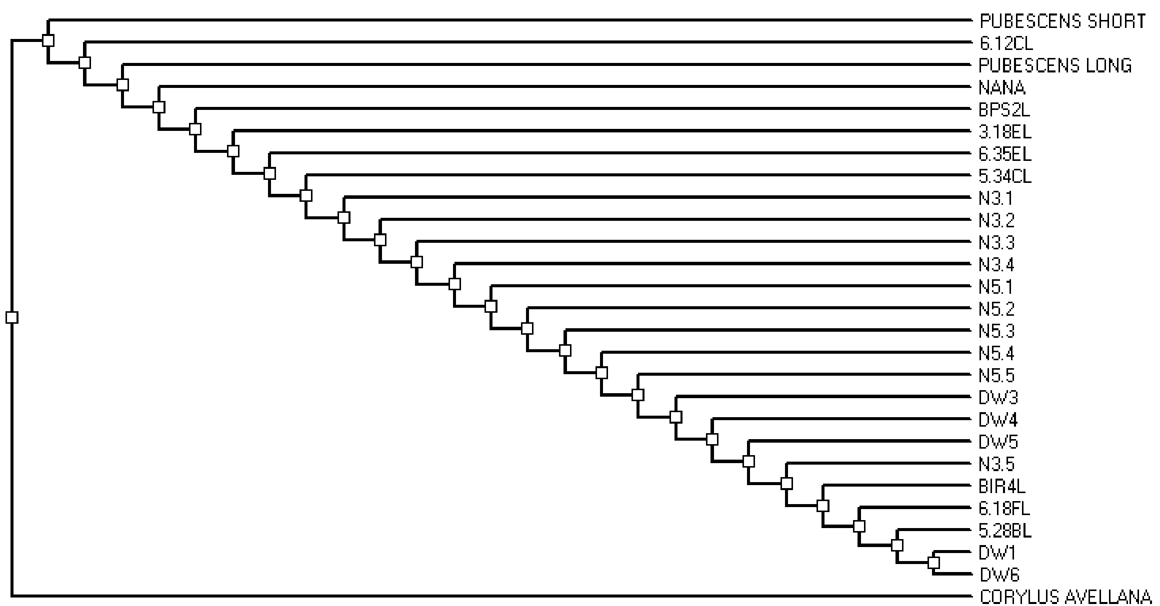

Fig. 3.2.6.3. Dendrogram of relatedness based on partial ADH genes for a section, based on where the majority of samples have data, of the forward sequence of the long allele.

A section of the reverse sequence of the long allele 450 bp from base 921 to 1536 (Fig. 3.2.3.22-24) long did however produce groups of similarity (Fig. 3.2.6.4). These were: A large group containing samples BPS1, BPS2, BPS4, 6.18F, 6.29A, 6.12C, 5.28B, 5.34C, N3.1, N3.2, N3.3, BIR4, DW4, DW1, N5.5, N5.3, N5.2 and the published *B. nana* sequence, plus a smaller group of two samples: the short allele of *B. pubescens* and the long allele of sample 6.35E (Fig. 3.2.6.4). As with the short alleles there was no evidence of a difference between different length alleles in the reverse section.

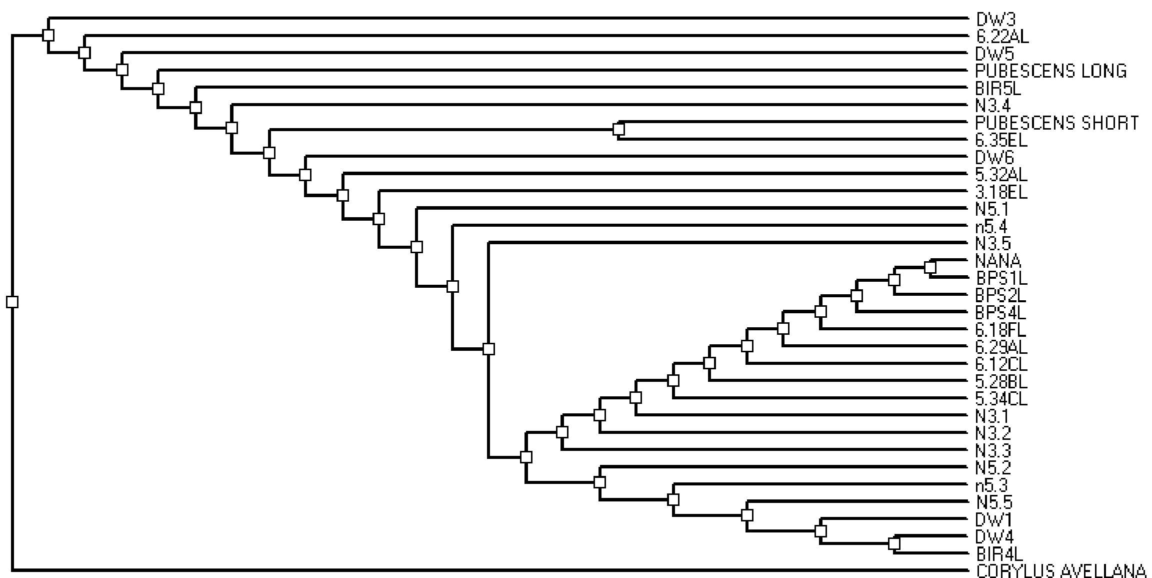

Fig. 3.2.6.4. Dendrogram of relatedness based on partial ADH genes for a section, based on where the majority of samples have data, of the reverse sequence of the long allele.

The relationship between samples do not match between different alleles, or even different sections of the same length allele and all evidence suggests that for a section of at least 450 bp from the start the reverse direction there is no difference between long and short ADH sequences. One of the potential problems is the shortness of sections used combined with relatively few differences between these sections. With very few differences groups are unlikely to form because a significant difference between samples is unlikely to occur. Any differences seen are likely to be based purely on a few base changes. Where significant differences between individuals tend to occur many samples misalign due to carrying different versions of the same length allele. Since this means sequence data for those samples cannot progress beyond that point a section of sequence containing data for the majority of samples is inherently unlikely to contain many differences therefore producing weak results. Hence those samples for which complete sequences were obtained were compared.

For complete sequence data, phylogenetic trees based on nearest neighbour algorithms resulted in the grouping of all short sequences together (published *B. pubescens* short, 6.18F(short). BPS2(short), 6.35E(short), 6.29A (short), 3.18E (short), 6.22A (short), BIR4 (short) (Fig. 3.2.6.5)) nestled among the long alleles. Another groups formed between long alleles; DW3 and DW5 and a final group between N3.1, N3.2 and N3.5. (Fig. 3.2.6.5).

The samples for which complete sequence data was obtained are the long alleles of DW4, N5.3, DW1, N3.3, 5.34, DW3, DW5, DW6, 6.12C, N3.1, N3.2, N3.5, N5.4 and the short alleles of 6.18F, BPS2, 6.35E, 6.29A, 3.18E, 6.22A, BIR4. Sample 6.12C produced a complete sequence and is the only sample known with either two short and two long alleles or one short one long allele to produce a complete long allele sequence, suggesting that most samples with a matching banding pattern to 6.12C have two long and two short alleles with two different versions of the long allele whilst 6.12C has either two extremely similar versions of the long allele or only one copy of each length allele. All other samples from which a full long allele sequence was produced were DW4, N5.3, DW1, N3.3, DW3, DW5, DW6, N3.1, N3.2, N3.5 and N5.4. All these showed a *B. nana* type banding pattern – i.e. a band for the long allele only. They should have a minimum of two copies of the long allele and a maximum of four. Since the whole sequence could be gathered from a sequence of at least two overlapping copies of the long allele it would indicate that very little difference exists between the copies of the long allele in these samples. Sample 5.34 also showed the same banding pattern as all the *B. nana* samples and also must have at least two copies of the long allele. It is one of the few non *B. nana* samples to produce a complete long allele sequence and may suggest genetic

similarity to *B. nana*. Four of the five central site samples (site 6) produced complete sequences for the short allele (the sample from this site which did not produce a complete sequence was sample 6.12C). From gel data it is likely that sample 6.18F only has one copy of the short ADH allele and hence a complete sequence is to be expected. The other central site samples, however, all are likely to carry two copies of the short allele, and since they produce clear sequences it is likely that these two copies are identical or very similar. This may indicate that there is less genetic variation between central site individuals than individuals from other sites. BPS2, 3.18E, and BIR4 are also likely to carry two copies of the short allele, again suggesting that their copies of the short allele are very similar to identical to one another.

From complete sequence data, phylogenetic trees based on nearest neighbour algorithms resulted in the grouping of all short sequences together (published *B. pubescens* short, 6.18F (short). BPS2 (short), 6.35E (short), 6.29A (short), 3.18E (short), 6.22A (short), BIR4 (short) (Fig. 3.2.6.5)). This suggests strong similarities between all short alleles. Interestingly a second group of long alleles did not form, and the group of short alleles was nested among the long alleles indicating that other than the large area of insertion/deletion there are very few distinguishing differences between long and short alleles. Some groups form between long alleles; DW3 and DW5 group together as do N3.1, N3.2 and N3.5. Samples from the same site did not show any evidence of grouping with the exception of *B. nana* from site three (patch one: windward) where only N3.3 did not group with the others (Fig. 3.2.6.5).

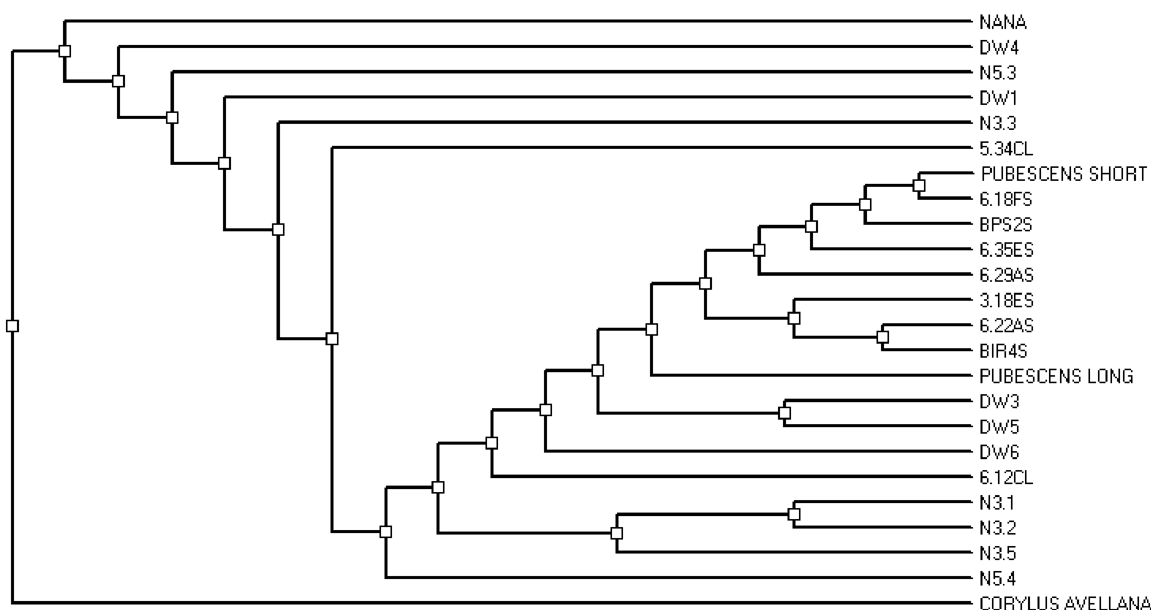

Fig. 3.2.6.5. Dendrogram of relatedness based on partial ADH genes for samples with complete sequences.

The grouping of the short alleles together could suggest less variance in short alleles than in the long alleles perhaps indicating high variability in the region 62-521 base sequence (Fig. 3.2.3.3 – 3.2.3.6) that is present in the long allele but absent in the short allele. Short alleles are also ~459 bases shorter than the long allele, since the long allele is approximately a 1500 base pair sequence a deletion of ~459 bases is a significant section of the sequence. Therefore, purely based on length the short alleles are less likely to have as much variance as the long alleles. There are also fewer complete short allele sequences, and those long sequences that are complete are mainly from samples of morphologically "pure" type *B. nana* which may therefore have less variation between one another. These samples cannot be taken as a good indicator of how other samples may have fallen out in a similarity tree, since these particular samples are unusual in having versions of the same length alleles that are very similar or identical. It is evident there are many variations within these

alleles and those samples with several versions of the alleles are not present in the full sequence data.

There are a variety of problems with any conclusions drawn from the dendrograms of similarity. The majority of these problems are caused by overlapping alleles; misalignments result in a shortage of data due to sequence data only being clear until the point of a deletion or insertion in one version of the allele that is not present in another. In order for the majority of samples to be studied short sections of sequence with few differences and hence fewer misalignments need to be used which give unreliable results as the interesting differences are therefore excluded. The technique also excludes N data (where N is several alleles), N's are counted as 'any' by the analysis. The algorithm works on the premise that it is a sequence from one version of an allele and does not contain overlapping sequence data, hence information about N bases (such as that they indicate two or more versions of the same allele with different bases at that point) is not included in the analysis. This can result in samples which should be similar becoming far removed. The differences between different versions of the same allele are the key to understanding relations. Ideally the individual sequences of each different version of the allele from each individual would need to be sequenced and analysed separately which would require the cloning of each individual to a plasmid in order to separate all the versions of the same length allele.

3.2.7 Conclusions

Using multiple methods of analysis two main groups formed together on several occasions. That these groups formed more than once using different methods would strongly suggest reliable similarities between members of these groups.

One main group which emerged from the short allele sequences is between samples BPS2, BPS3, 6.29A, 6.18F and 6.35E and the published *B. pubescens* sequence. They all have two adenine nucleotides at 575 bp from the start to the sequence (Fig. 3.2.3.6), and they also all share the sequence 5'-TACA-3' at 578 bp from the start, which no other samples share. At 624 bp into these sequences these samples alone read 5'-AACACCATATCAAAGTTCG-3' with the exception of 6.29A which has the sequence 5'-AACACCATATCAAGGTTCG-3'. This difference is a single base substitution, so given the differences seen in other samples (Fig. 3.2.3.7) this is a fairly trivial difference. Although the short sections of sequence data that were used for dendrograms of similarity produced varying results samples 6.18F and 6.35E grouped together for both the forward and reverse section of the short allele suggesting strong relations between these two samples. Given all these similarities it seems likely that these samples are similar to "pure" *B. pubescens.* However only 6.29A and 6.35E showed a "pure" *B. pubescens* banding type pattern, BPS2 has two long alleles one short as in the F1 hybrid, BPS3 has short alleles only and 6.18F has three long one short. It is worth noting that the published *B. pubescens* sequence is not necessarily a "pure" sequence no references are made to attempts to establish the purity of the sample (Järvinen et al., 2004) taking into account that these trees introgressively hybridise it seems likely that this sequence could easily be a hybrid

itself. Regardless of "pureness" these samples evidently are closely related despite different short allele carrying chromosome numbers. Depending on the age of the samples, which is unknown, these individuals could be siblings, parent-offspring relations or more distantly related. The fact that the sequences progress past the region at 624bp where all other samples have a different length section of sequence at this point (of two other possible sequences) or misalign due to carrying two different versions of the allele, means that these samples all share the same version with no differences (excepting the single base substitution in 6.29A). It is therefore likely that these individuals share a closer common ancestry for at least the short allele than with any of the other samples.

Another short allele group forms between samples BPS5, 3.3A, 3.5C, 3.7A, 3.12G, 3.18E, 6.22A, BIR1, BIR3 and BIR4 (Fig. 3.2.3.6). These samples all have three adenine nucleotides at 575 bp into the sequence whereas all other samples have two, or misalign. BIR3 and BPS5 (Fig. 3.2.3.6) share a misalignment at 557-578 bp from the start; at this point the other samples in the group share the sequence 5'-TCA-3' (3.3A, 3.5C, 3.7A. 3.12G, 3.18E, 6.22A, BIR1 and BIR4 (Fig. 3.2.3.6)) which is not seen in any of the other samples. Sequences for 3.12G and 3.3A become misaligned before a section of high variability at 625b in (Fig. 3.2.3.7). 3.18E is the only sample to have the sequence at this point 5'- ACTTATCAGAGTTCA-3' and only samples 3.5C, 3.7A, 6.22A, BIR1 and BIR4 have the sequence 5'-ACTTAAGCTTTTATTAACTT-3'. There are only two other potential sequences seen at this point and these are found in the other group suggesting that the samples which misalign at this point carry at least one version seen in samples 3.5C, 3.7A, 6.22A, BIR1 and BIR4 and one version seen in sample 3.12G 661 bp into the short

allele sequence. Sequence data from samples 3.5C and 3.7A become misaligned (Fig. 3.2.3.7) the remaining samples in the group apart from BIR1 (3.18E, 6.22A, BIR4) continue to form full sequences which in full sequence dendrograms fall out close to one another (Fig. 3.2.6.5). Although sections of forward and reverse sequences for nearest neighbour analysis proved unreliable, samples BIR4, 3.3A and 3.7A grouped both times in these dendrograms of similarity for the short allele. It is evident that members of this group is not as closely related to one another as the first group. These samples evidently have at least one version of the short allele in common, suggesting that they are related but crossing with others may occur. Chromosome numbers support this, samples BPS5, 3.3A, 3.5C, 3.7A, BIR3, either have allele numbers of three short one long or two short one long. All of these samples misalign earlier whereas all other samples (BIR4, 3.12G, 3.18E, 6.22A, BIR1) likely have the allele numbers two short two long or one short one long. All of these samples continue further without misalignments than the others apart from sample 3.12G. This may suggest that samples with a higher ratio of short to long alleles (and potentially higher number of short alleles) have gained a short chromosome from elsewhere i.e. one not associated with the lower ratio short allele site three samples.

The remaining samples that do not fall into either of these groups are all the *B. nana* samples: N3.1, N3.2, N3.3, N3.4 N3.5, N5.1, N5.2, N5.3, N5.4, N5.5, DW1, DW3, DW4, DW5, DW6 and all the *B. pubescens* type samples BPS1, BPS4, BIR2, BIR5, BIR6, 5.28B, 5.34C, 5.32A and 6.12C. Since all the *B. nana* samples, 5.28B and 5.34C do not have short alleles and the two groups that formed are based around the short allele it is evident why these samples do not fall into either of these groups.

The short alleles for samples BPS1, BPS4, BIR2, BIR5, BIR6, 5.32A and 6.12C misalign prior to or at the point at which these two groups initially diverge (Fig. 3.2.3.3 – 3.2.3.6) (all sequence data for the forward section of the short allele of BPS4 was misaligned and hence is not shown in sequence lineups). Hence these samples do not fall into either of these groups and most likely have some versions of the short ADH allele from each group. There are no long allele groups which form on more than one occasion. Very few misalignments on long allele sequences occur which leads to the inference that variation in the long allele is predominantly due to base substitutions rather than insertions or deletions. Full sequence data suggests little similarity between long allele samples and they do not form a distinct group unlike the short allele samples. Some groups for long allele sequences do however occur on rare occasions and those groups which are unlikely to be an artefact due to short sequences or debateable points of misalignment are the groups formed between 6.12C, 6.18F, N5.1 and N5.2 from the ten base pair insertion 210 bp into the long allele, the group DW3 and DW5 from full sequence data and the group formed between all the site three *B. nana* samples (N3.1, N3.2 and N3.5 (Fig. 3.2.6.5)) excluding N3.3.

Some evidence of site individuality can be gained from these groups. The two main groups separate the majority of site six (central) samples into one group and site three (windward) samples into another group. The exception is sample 6.22A which groups with the site three samples. Morphologically *B. pubescens* type samples (BIRX) either group with site three or do not group at all, and never group with the main site six group even though central site samples (site 6) are the most morphologically similar to *B. pubescens* in the patch. Site five (leeward) samples do

not fall into either group suggesting high variation between individuals and similarity to both groups. *B. nana* from site three group with one another (with the exception of sample N3.3) whereas *B. nana* from site five do not group. The morphologically pure *B. nana* samples (DWX) also do not group, with the exception of individuals DW3 and DW5 which group with one another. This suggests some form of genetic selection is occurring in sites three (windward) and six (central) but not five (leeward).

Since site three is closer to site six than it is to site five the differences seen between site three and six are therefore unlikely to be due to short range dispersal of seeds or pollen. *B. pubescens* and *B. nana* seeds are both light and winged (personal observation) and can travel large distances via wind dispersal but are also likely to be secondarily dispersed via snow melt (Pers. comms: R. Baxter. Durham University) as seen in other *Betula* species (Greene & Johnson, 1997; Matlack, 1989). Seeds from a particular tree are therefore unlikely to remain localised. This may therefore suggest there is a particular selective pressure occurring at three that is different from the pressure at six, but is not present or less extreme at site five.

Site five is adjacent to a lake (Fig. 2.1.1.1) and water is therefore likely to be in abundance for these individuals. Site three however is encroaching upon the tundra and is likely to be very dry. Site six is on the top of a hummock (pers. Comms. R. Baxter, R. Holden) and more densely populated therefore competition for water is most likely to be at its highest in site six. Dehydration is known to be a major problem in sapling establishment (Kullman, 1986) and both species are known to

grow well on wet ground (Walters, 1968; Horsfield & Thompson, 1997; Jonasson, 1981; Kirkpatrick & Heal, 2001; Ejankowski & Kunz, 2006).

B. nana is restricted to bogland at lower latitudes (Horsfield & Thompson, 1997; Kirkpatrick & Heal, 2001) and found in bogland, tundra and heath in more northern areas (Jonasson, 1981). *B. pubescens* is also known to tolerate wetter ground than most *Betula* species (Walters, 1968), however it is likely that *B. nana* is more tolerant of wetter ground than *B. pubescens,* since only after drainage of bogland habitats could *B. pubescens* encroach upon *B. nana* habitats (Ejankowski & Kunz, 2006). Since ADH is known to be linked to drought stresses or stresses inducing physiological drought (Freeling & Bennett, 1985; Xie & Wu, 1989; Jarillo et al., 1993; Dolferus et al., 1994; Hoeren et al., 1998), and these two species are more tolerant of wetter locations (Walters, 1968; Jonasson, 1981; Horsfield & Thompson, 1997; Kirkpatrick & Heal, 2001; Ejankowski & Kunz, 2006), it is possible (but not certain since these changes could be silent mutations) that selection of particular ADH alleles may occur at drier sites such as site three and 6, whereas survival at site five which has an abundant water supply, but is not waterlogged, does not require specifically more drought resistant ADH alleles. Although to fully established trees water acquisition may be less troublesome due to large root systems it is during seedling establishment that drought causes the greatest number of deaths (Kullman, 1986). So in dry habitats such as site three and six selection for more drought tolerant saplings is likely to occur and different versions of the ADH allele may be involved in this selection. *B. nana* is capable of surviving in drier habitats in northern latitudes than in more southern latitudes, so it is possible that at more northern latitudes *B. nana* may have developed a different version of the ADH allele

to survive in the drier tundra habitats. It may also have acquired ADH alleles and other genes adapted to drier habitats through introgressive hybridisation with *B. pubescens.* Since site three is much drier than site 5, *B. nana* with ADH alleles adapted to drier habitats are likely to be selected at site three and may explain the tendency of site three *B. nana* samples to group together whereas at site 5, where there is a more copious water supply, this selection pressure does not exist and hence no grouping is seen in site five *B. nana*. Site three samples are likely to be exposed to colder winds at the windward edge of the patch and in winter are likely to be highly exposed to cold and freezing as snow is blown out from the leading edge of the patch and deposited further into the patch (as seen at site 6). Although ADH was not linked to freezing tolerance in *A. thaliana* expression was found to increase in response to cold. A different ADH allele may therefore be selected for in the less dry but colder site three than in the drier less cold site 6. However further experiments would be required to determine if these observed differences in ADH sequence actually have an effect on the ADH protein activity and expression, and are therefore being selected for at the particular sites as discussed above, or if the changes are silent mutations which have no effect on the coding sequence of the protein, and the observed selection is a by-product of selection for something else, likely a nearby gene on the same chromosome. These further experiments would include: comparison of the ADH sequences to other species and using previous knowledge of the relation of protein sequence to ADH function (from other species) to determine if these changes in protein sequence are sufficient to cause a change in protein activity and expressing each individual ADH allele in *Escherichia coli* to link in-vitro ADH activity to the polymorphisms seen.

The potential data acquisition was limited by misalignment of the majority of sequences. Greater resolutions of similarity and relatedness could be deduced by analysing each version of each length allele in each individual, which would require the separation of different versions of the same length allele in each individual. This could be done via cloning into a plasmid, which would give greater resolution and may identify more similarities. However it would still only give data for the similarity and selection of ADH alleles, which may not be representative of the genome as a whole. Hence a method of looking over the entire genome for similarities may provide greater evidence of relations between sites.

Other conclusions, less directly related to the ADH alelle can also be drawn from this data. Firstly that introgressive hybridisation definitely occurs in Abisko where previously most evidence of introgressive hybridisation between *B. pubescens* and *B. nana* came from studies in Iceland (Elkington, 1968; Anamthawat-Jónsson & Tomasson, 1990; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2003; Anamthawat-Jónsson et al., 2010). Due to the discovery of individuals with allele ratios other than two short, two long, long only and two long one short, it is evident that in F1 hybrids, chromosomes with long alleles can pair with chromosomes with short alleles and produce gametes other than haploid long allele only and diploid long and short allele only. It is also evident that morphology cannot be used to indicate purity especially in *B. pubescens* since many of the "pure type" *B. pubescens* showed strong evidence of hybridisation suggesting that a small number of genes, the environment or a combination of both are involved in the different morphologies seen and that the published sequences may not be reliably "pure". Although nearest neighbour dendrograms of sections of the ADH allele were

not particularly reliable for use as an indicator of relations between samples, it was however evident that at least the 450 bp end sequence is highly conserved in both length alleles and species. This may indicate that this is a key region involved in the function of ADH considering the high level of variation that occurs in the earlier sequence.

3.2.8 Summary

It is possible that ADH allele selection is occurring at sites three and 6, most likely linked to drought stress, but not at site five which is well watered. Alternatively the observed selection could be due to selection for a nearby gene on the same chromosome(s) as the ADH alleles so further experimentation, detailed above, would be needed to assess which gene(s) were being selected for. This does however indicate that small scale genetic selection can occur over these patches likely in relation to microhabitats. Analysis of more genes is required to assess if selection is occurring over the entire genome at these sites which could explain the morphologies observed. This study also shows that introgressive hybridisation is taking place between *B. nana* and *B. pubescens* populations in Abisko, which previously had only been shown to occur in Iceland (Elkington, 1968; Anamthawat-Jónsson & Tomasson, 1990; Thórsson, Salmela & Anamthawat-Jónsson, 2001; Anamthawat-Jónsson et al., 2003; Anamthawat-Jónsson et al., 2010) and that morphology alone cannot be used to indicate the purity of an individual, especially in *B. pubescens*.

3.3 RAPD Analysis

3.3.1 RAPD Primers

RAPDs analysis of the samples was undertaken with the aim of comparing banding patterns to identify the similarities of individuals over the entire genome rather than being restricted to one gene, as with the alcohol dehydrogenase analysis. Fourteen primers were tested for suitability by amplification with DNA from a sample chosen at random (BIR4) and running the PCR product on a TAE gel to access the twelve clearest banding patterns for running on the Bioanalyzer (Fig. 3.3.1.1.).

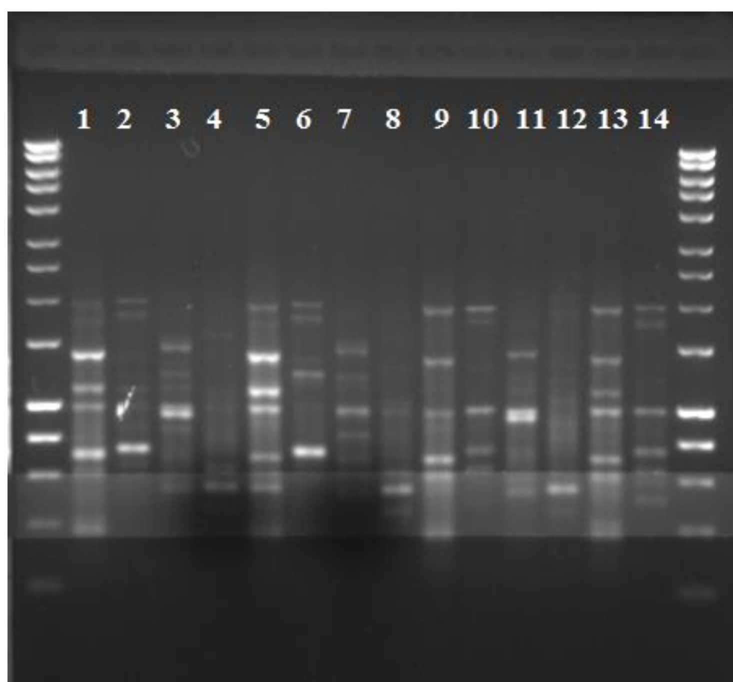

Fig. 3.3.1.1. TAE gel of PCR products (PCR mix 3, Program 4) of Bir4 with RAPD primers (1)ATC33. (2) OPB17. (3)OPB14. (4) OPB11. (5) OPB12. (6) OPB18. (7) ATC71. (8) ATC51. (9) RAPDa. (10) RAPDb. (11) RAPDc. (12) RAPDd. (13) RAPDe. (14) OPB08.

Primers RAPDa, RAPDb, RAPDc, RAPDd, RAPDe, OPB08, OPB11, OPB12, OPB14, OPB17, ATC33, ATC51 produced the best banding patterns (Fig. 3.3.1.1) and were used to amplify fragments of DNA from sample 3.7A. PCR products were then run on the bioanalyzer to choose four primers with clear peak definition and a large range of peaks (Fig. 3.3.1.2.)

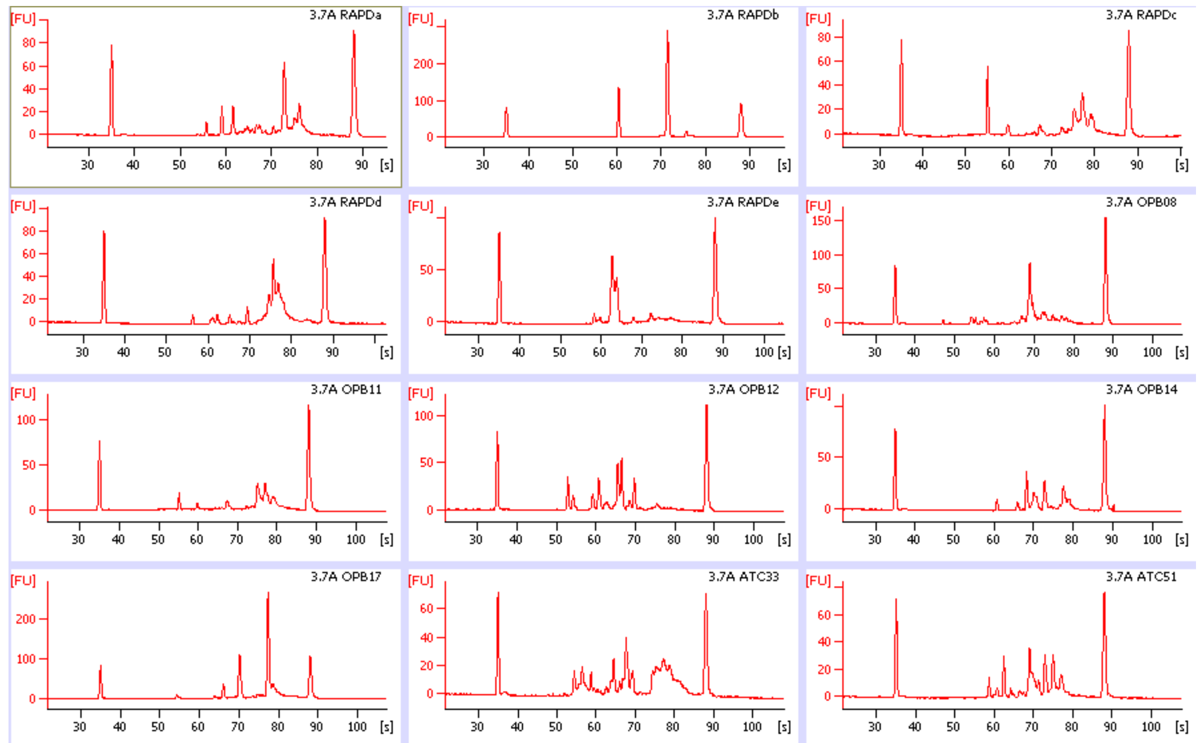

Fig. 3.3.1.2. Electropherogram readouts from the bioanalyser for DNA from sample 3.7A run with various RAPD primers.

RAPDs OPB11, OPB12, OPB14 and ATC51 were selected as the best candidates and used for analysis of further samples. A large range of peaks were produced for each patch and primer (Fig. 3.3.3.2 – 3.3.3.13) Due to a degree of unreliability in peak location (Fig. 3.3.3.1.) these peaks were placed into groups to avoid counting two identical peaks as separate peaks. The presence of peaks in each group was used to give each sample a presence absence code to be analysed with Dice’s coefficient of similarity.

3.3.2 Sample Morphologies

The degree of polycormy of all individuals sampled in patches one and two was recorded at the point of collection, to allow comparison between genetic and morphological similarities between samples. A large degree of polycormy is seen across the trees sampled in patch one and two (Table 3.3.2.1)

Sample ID	**Polycormic?**		**Sample ID**	**Polycormic?**
1NE1.1	Yes		2NE1.1	Yes
1NE1.2	Yes		2NE1.2	Yes
1NE1.3	Yes		2NE1.3	Yes
1NE1.4	Yes		2NE1.4	Yes
1NE1.5	Yes		2NE1.5	Yes
1E1.1	No		2E1.1	10 stems
1E1.2	Yes		2E1.2	16 stems
1E1.3	Yes		2E1.3	Yes
1E1.4	Very		2E1.4	Yes
1E1.5	Very		2E1.5	Yes
1M1	Very		2M1	No
1M2	Hardly		2M2	3 stems
1M3	Yes		2M3	3 stems
1M4	No		2M4	7 stems
1M5	Hardly		2M5	2 stems
1E2.1	No		2E2.1	No
1E2.2	No		2E2.2	3 stems
1E2.3	No		2E2.3	3 stems
1E2.4	Yes		2E2.4	Yes
1E2.5	Yes		2E2.5	No
1NE2.1	Yes		2NE2.1	Yes
1NE2.2	Yes		2NE2.2	Yes
1NE2.3	Yes		2NE2.3	Yes
1NE2.4	Yes		2NE2.4	Yes
1NE2.5	Yes		2NE2.5	Yes

Table 3.3.2.1. Table showing degree of polycormy in patch one and two individuals

3.3.3 Peak Groupings

There is a degree of error in the bioanlayzer readouts with the same peaks varying by approximately two seconds which can cause confusion between peaks that are close together. (Fig. 3.3.3.1)

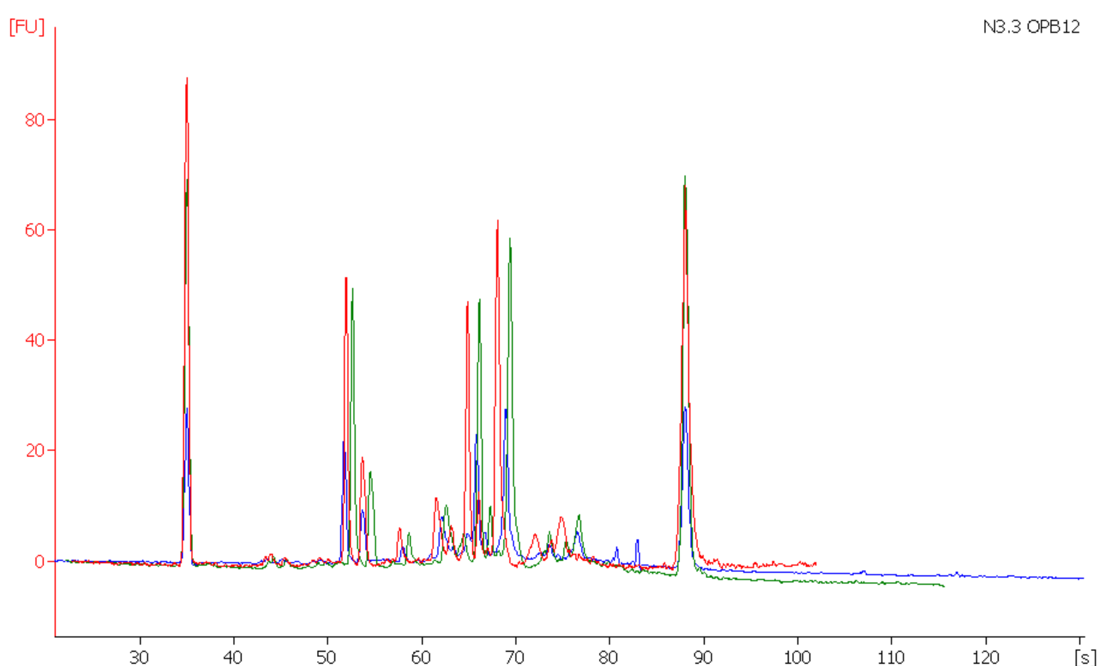

Fig. 3.3.3.1. Example of degree of error for the same sample from the same PCR run and primers on three different bioanalyzer runs.

All samples produced a large variation in peak patterns for each primer (Fig. 3.3.3.2-3.3.3.13). A variety of peak groups were produced for each primer and labelled (Fig. 3.3.3.2-3.3.3.13). Patch three has the least well defined peak groups (Fig. 3.3.3.10-3.3.3.13) and so the basic numbering system of peaks was based on the peak groups from patch three. When peak groups seen in patch three could be split into smaller groups in the other patches they were suffixed with a letter. (Fig 3.3.3.2-3.3.3.9).

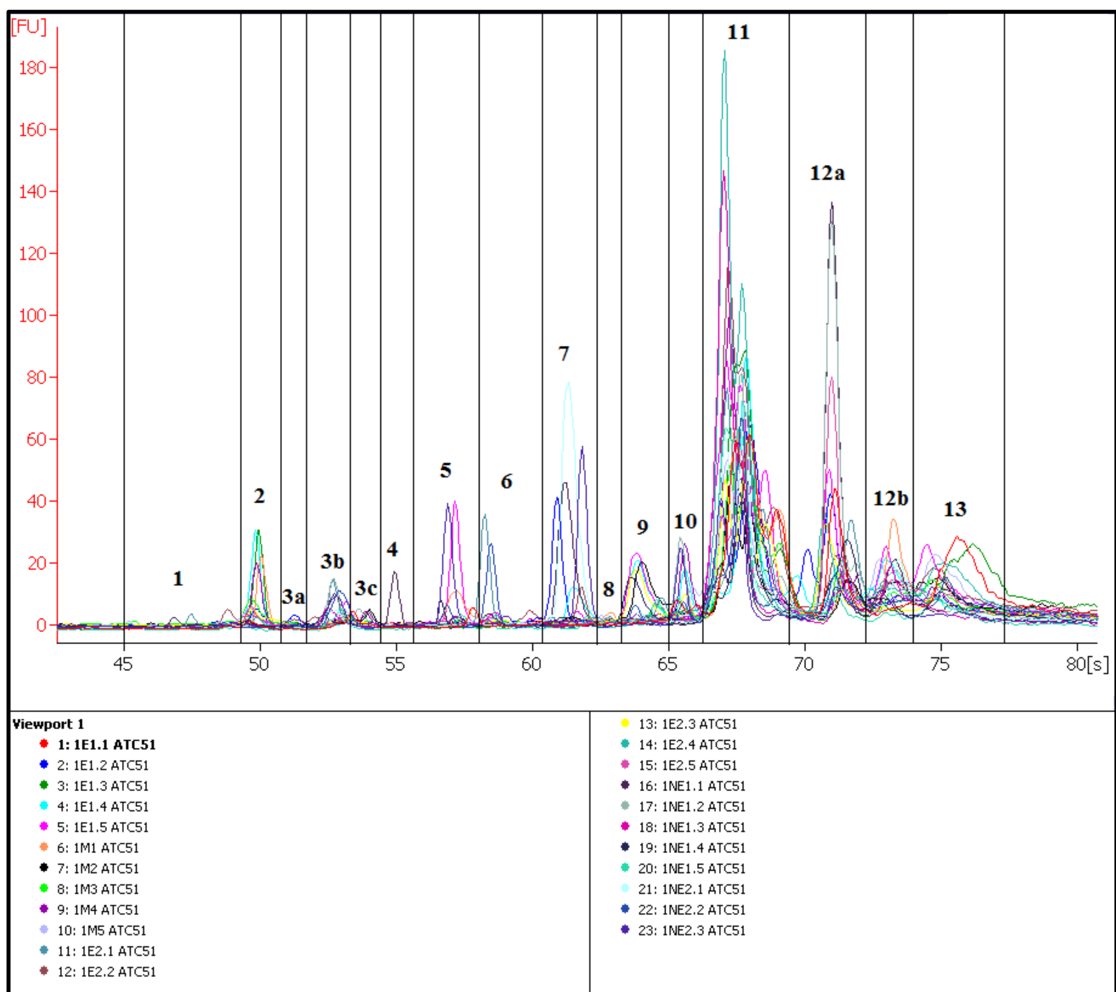

Fig. 3.3.3.2. Electropherogram showing division of peak groups for Patch one samples with primer ATC51

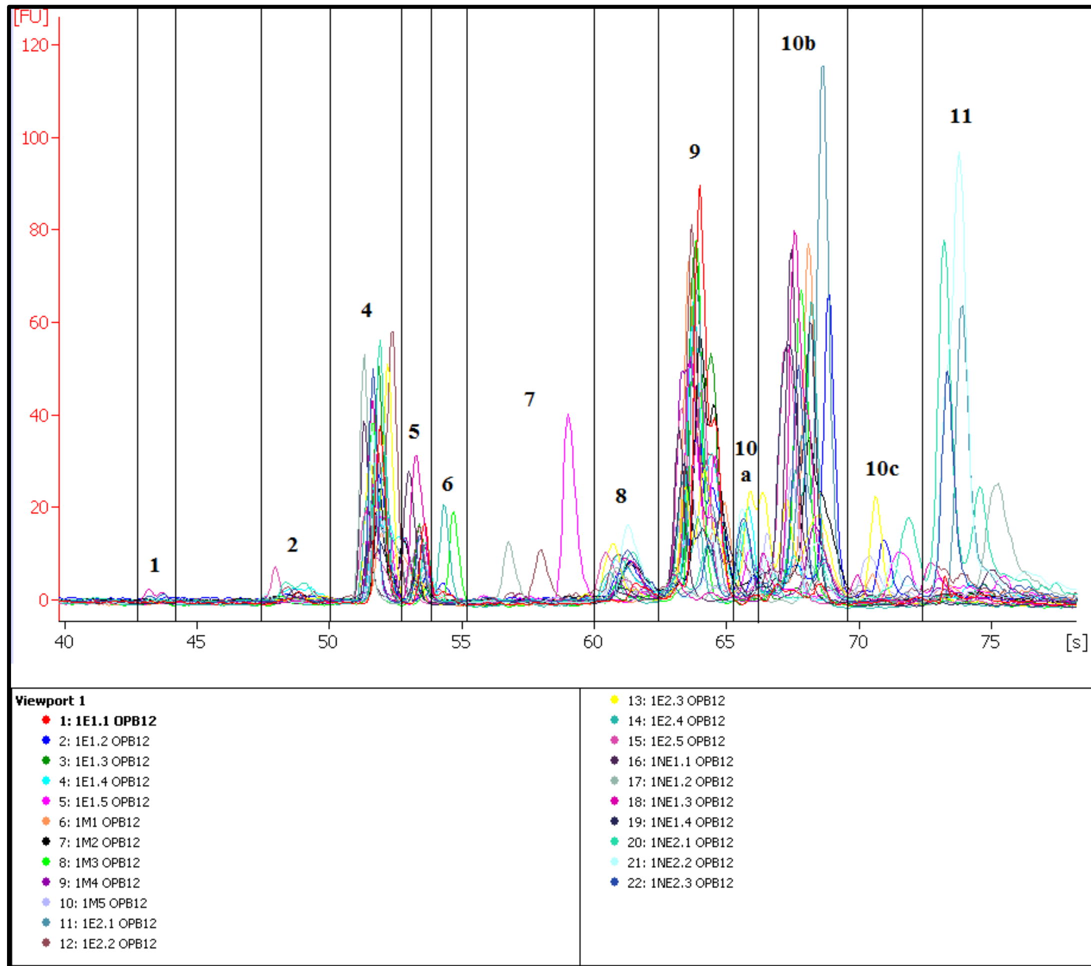

Fig. 3.3.3.3. Electropherogram showing division of peak groups for Patch one samples with primer OPB11

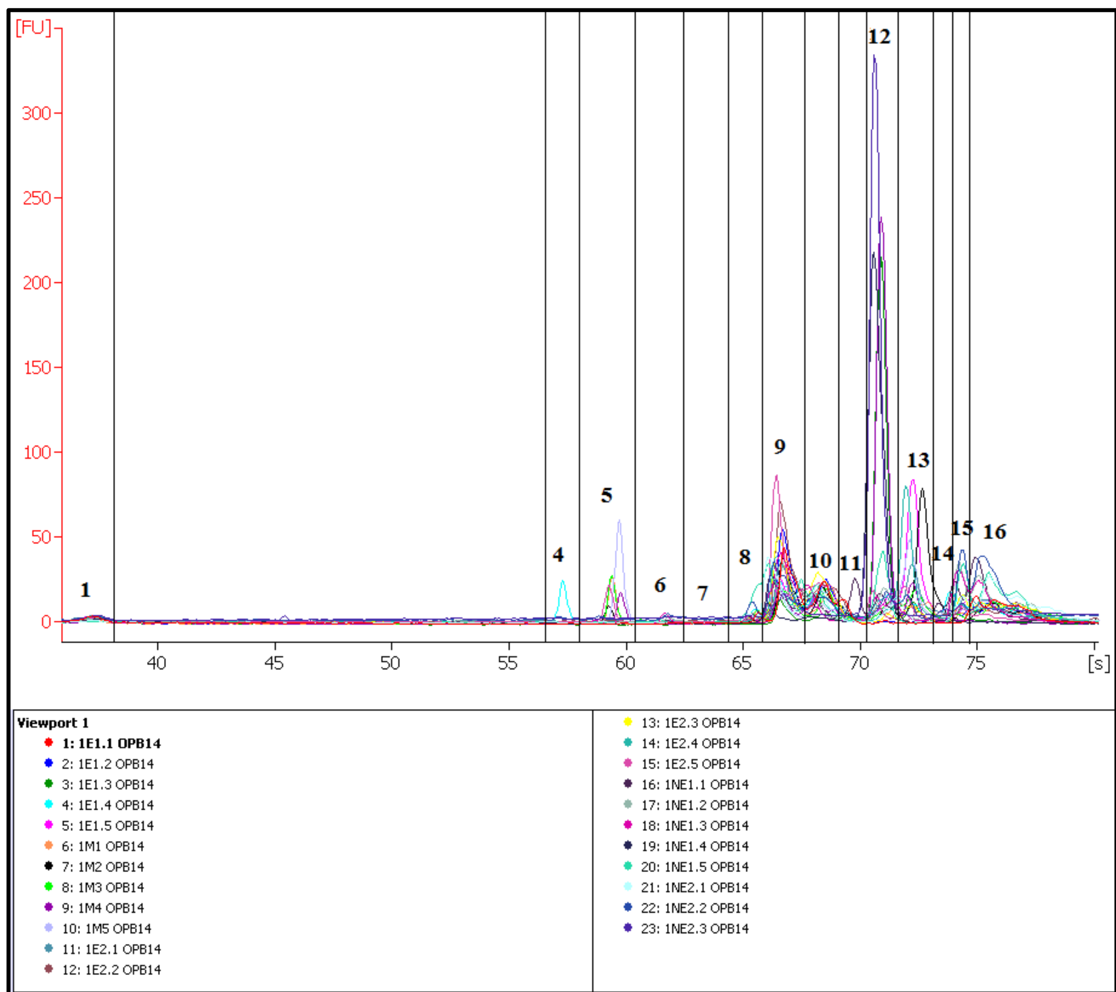

Fig. 3.3.3.4. Electropherogram showing division of peak groups for Patch one samples with primer OPB12

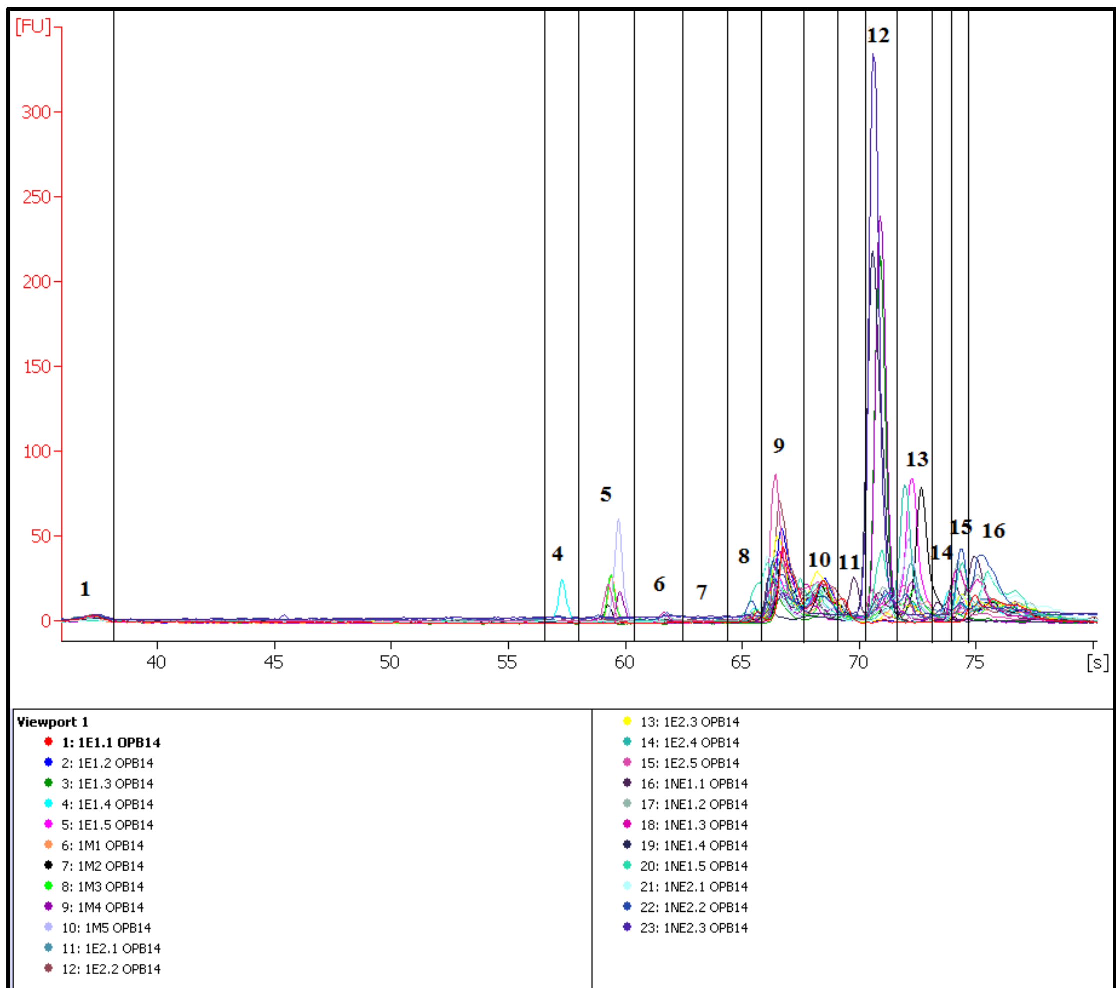

Fig. 3.3.3.5. Electropherogram showing division of peak groups for Patch one samples with primer OPB14

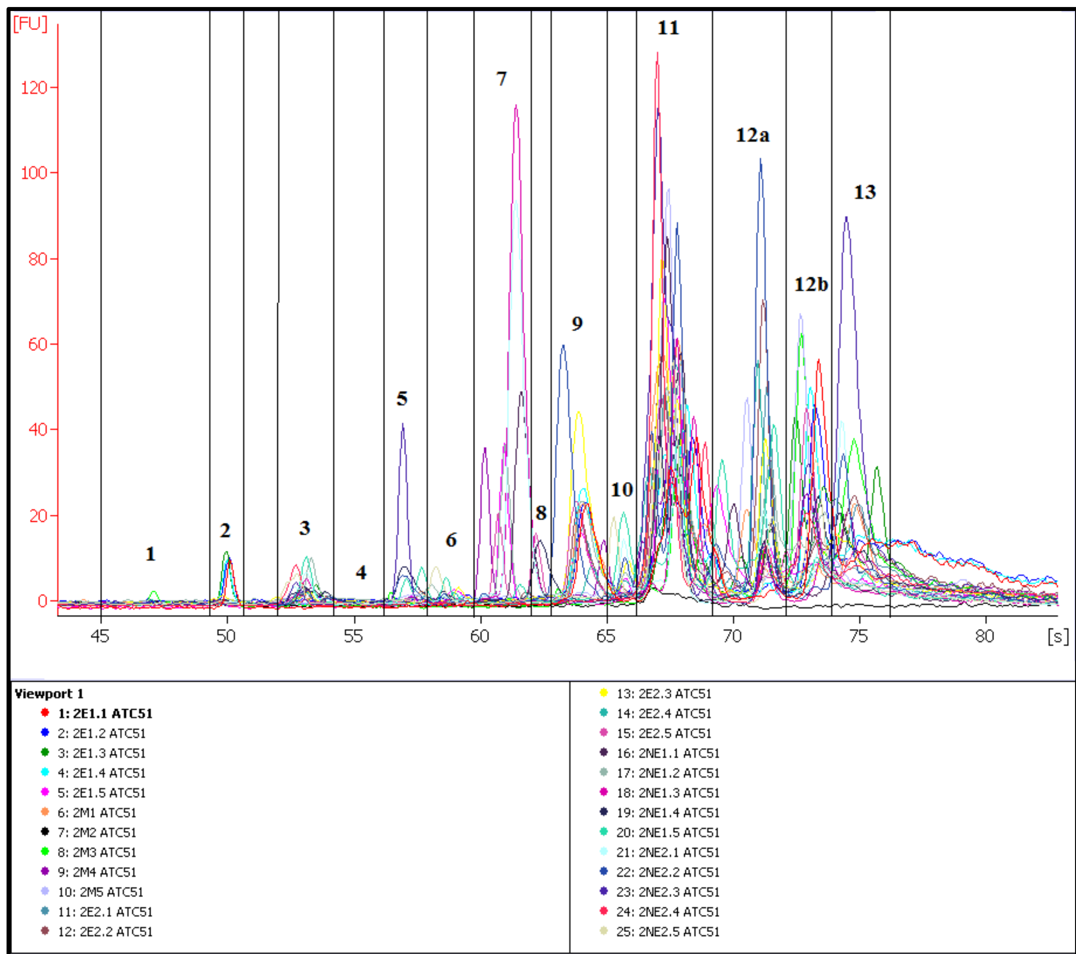

Fig. 3.3.3.6. Electropherogram showing division of peak groups for Patch two samples with primer ATC51

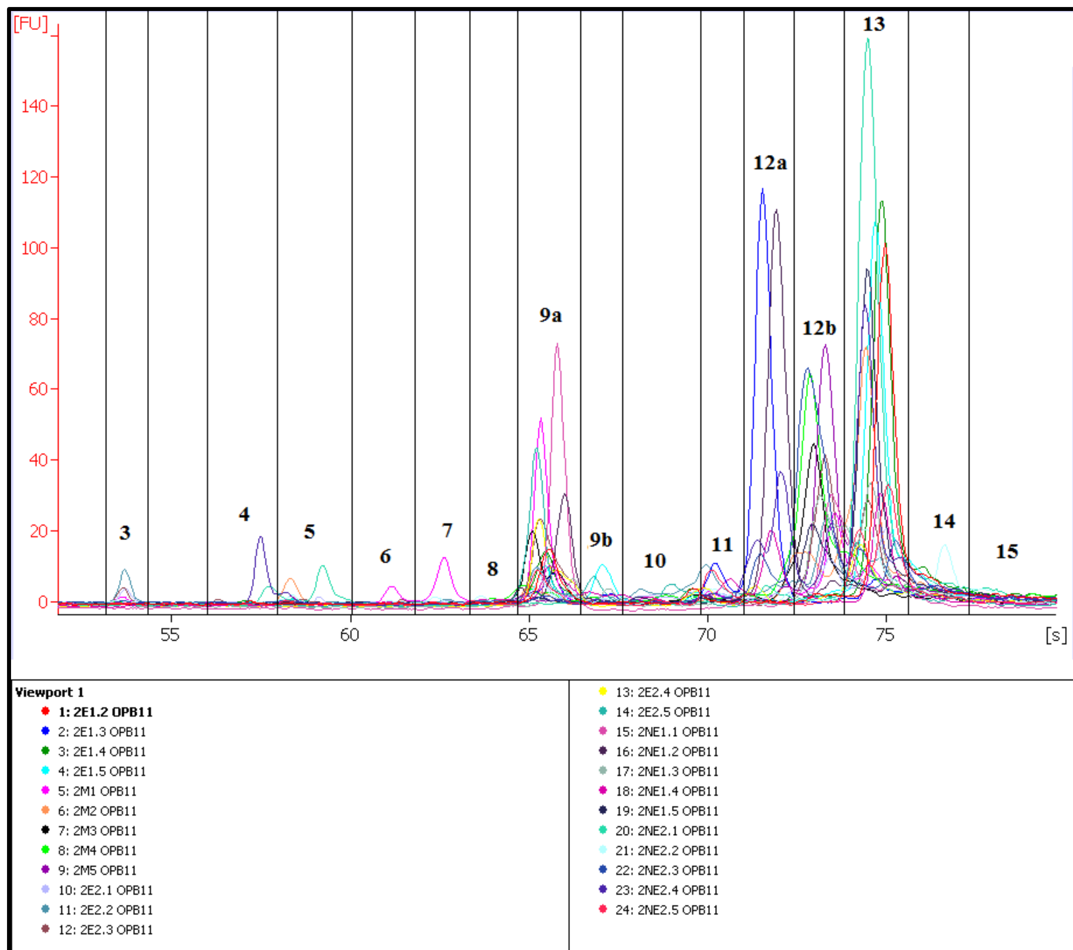

Fig. 3.3.3.7. Electropherogram showing division of peak groups for Patch two samples with primer OPB11

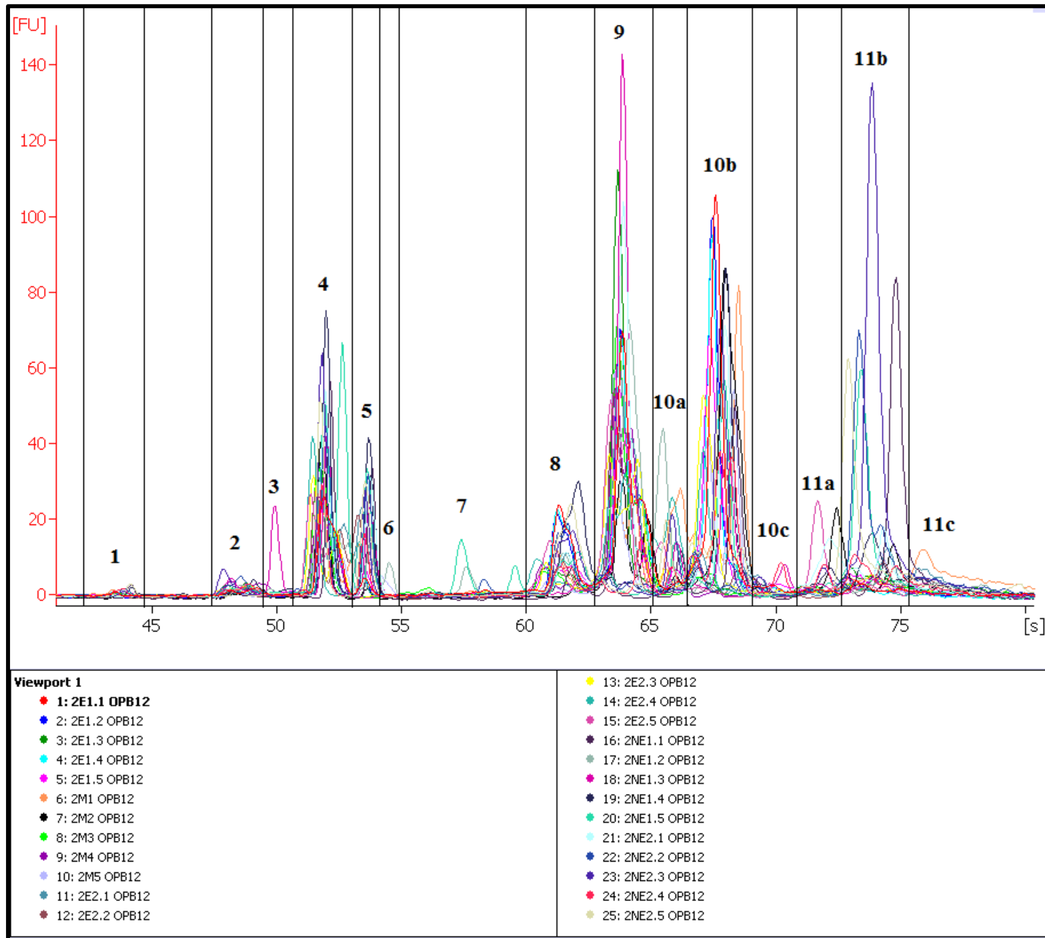

Fig. 3.3.3.8. Electropherogram showing division of peak groups for Patch two samples with primer OPB12

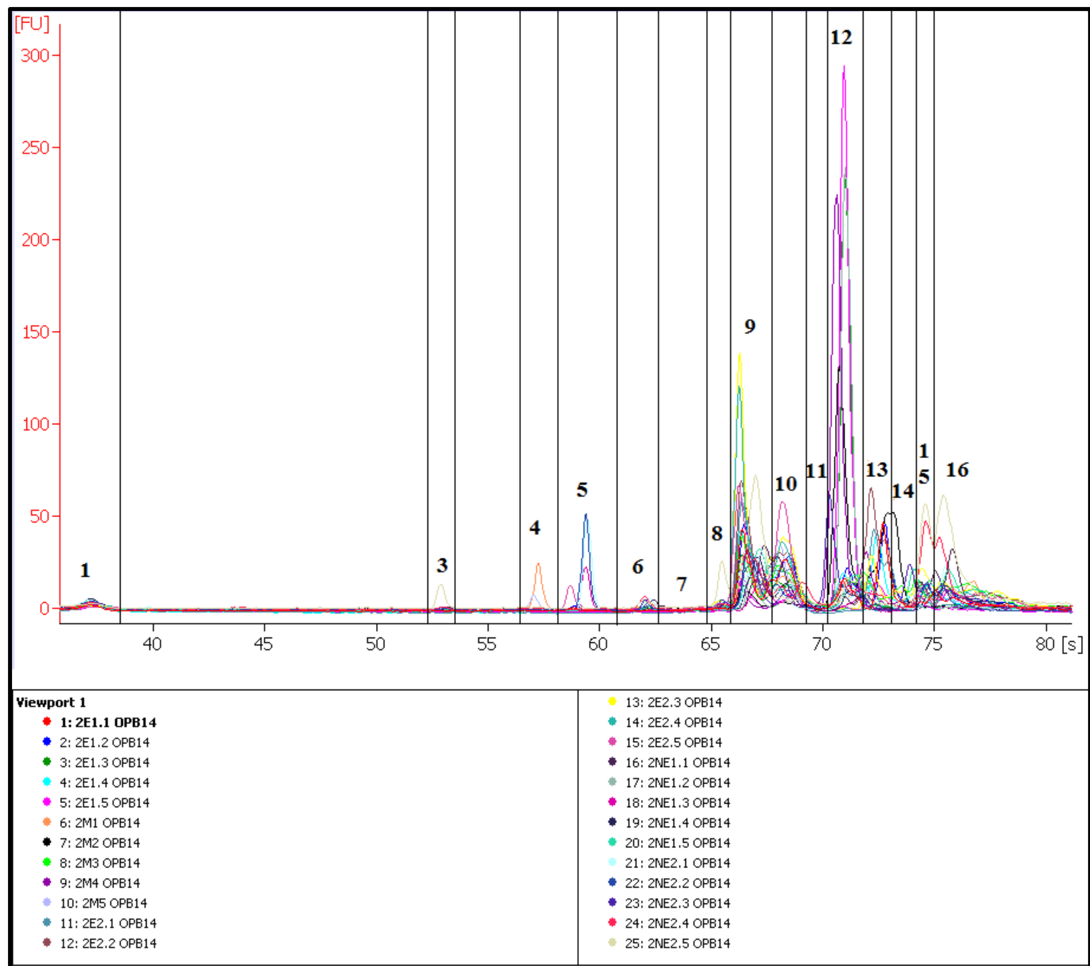

Fig. 3.3.3.9. Electropherogram showing division of peak groups for Patch two samples with primer OPB14

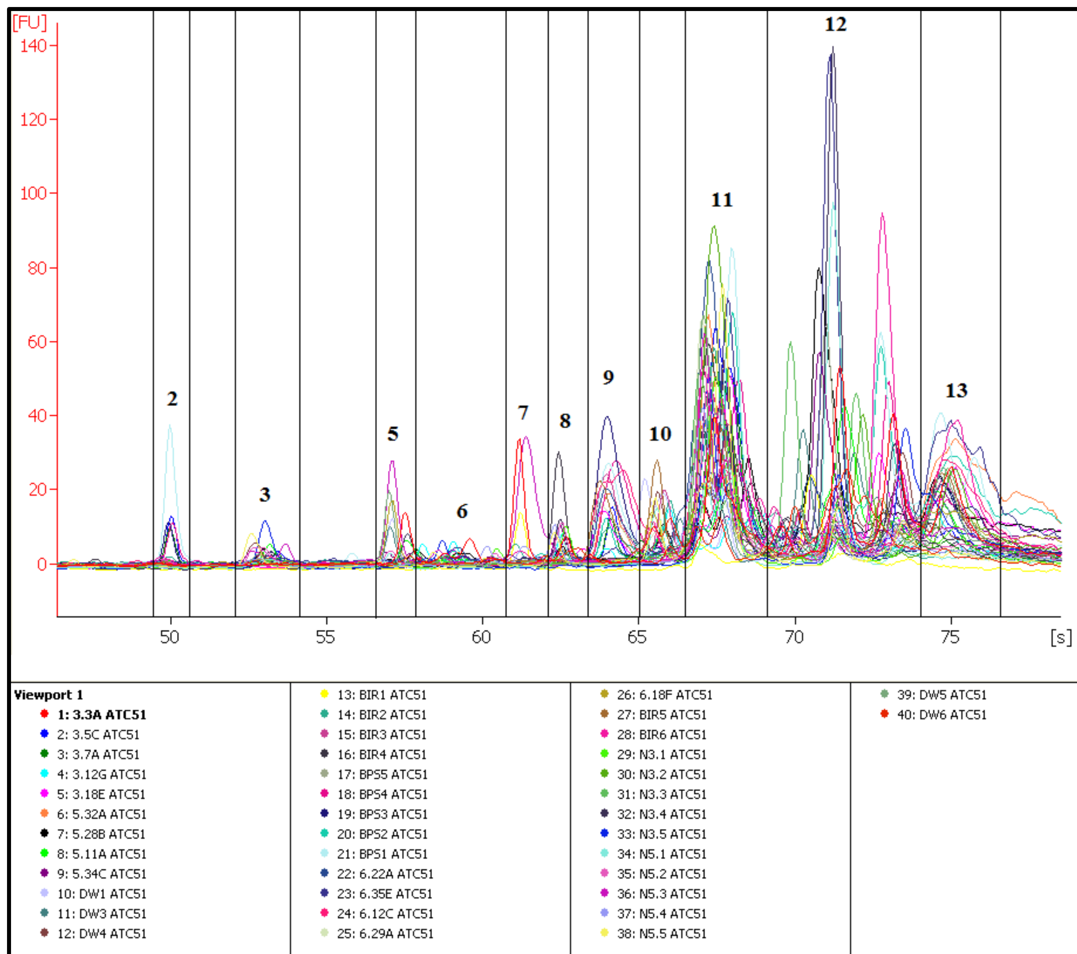

Fig. 3.3.3.10. Electropherogram showing division of peak groups for Patch three samples with primer ATC51

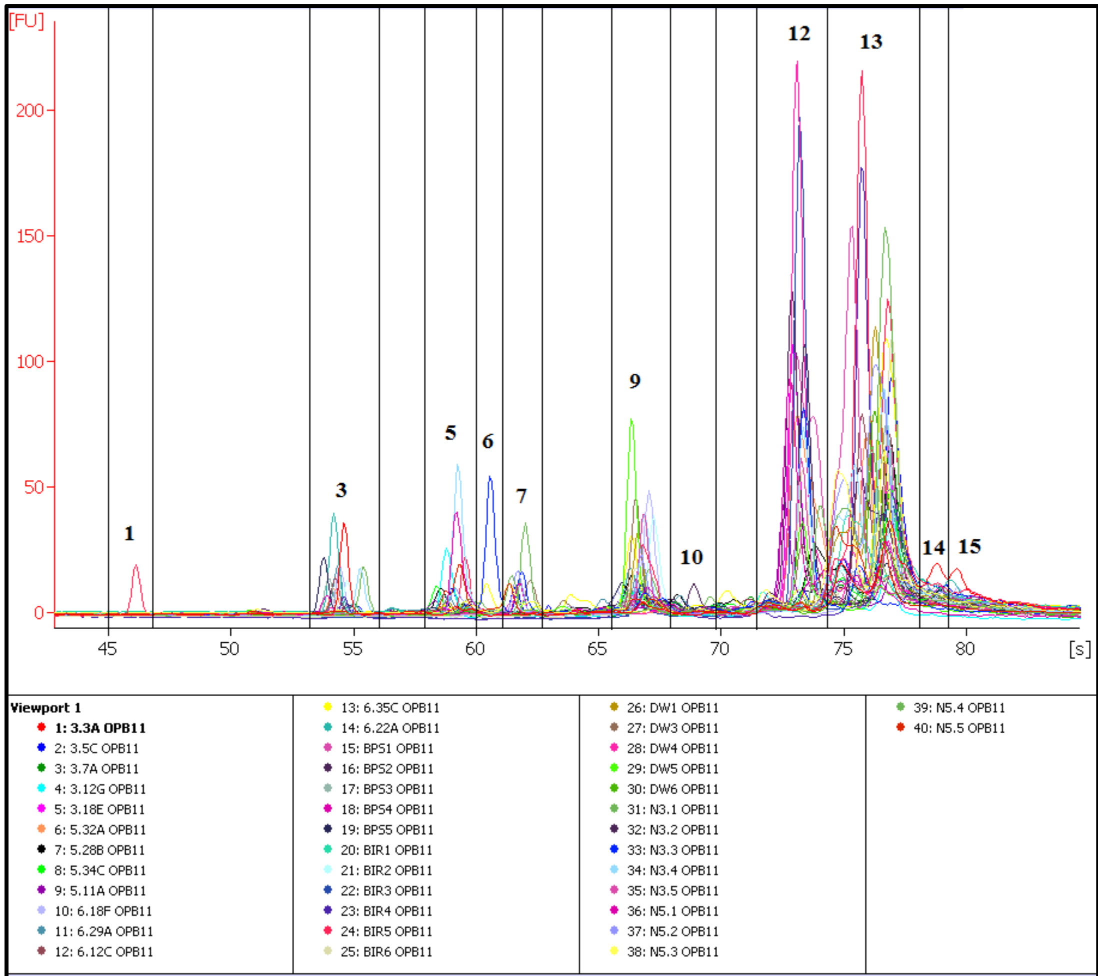

Fig. 3.3.3.11. Electropherogram showing division of peak groups for Patch three samples with primer OPB11

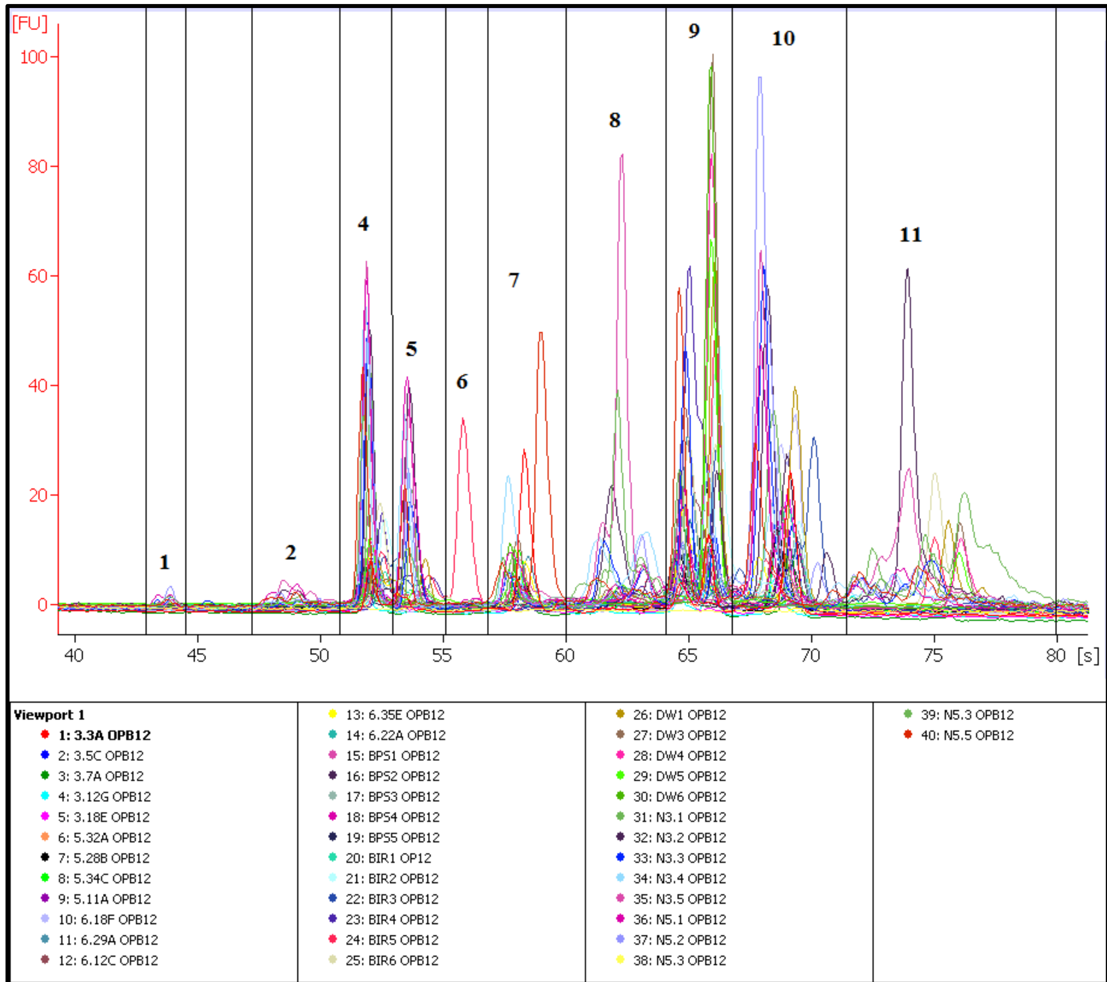

Fig. 3.3.3.12. Electropherogram showing division of peak groups for Patch three samples with primer OPB12

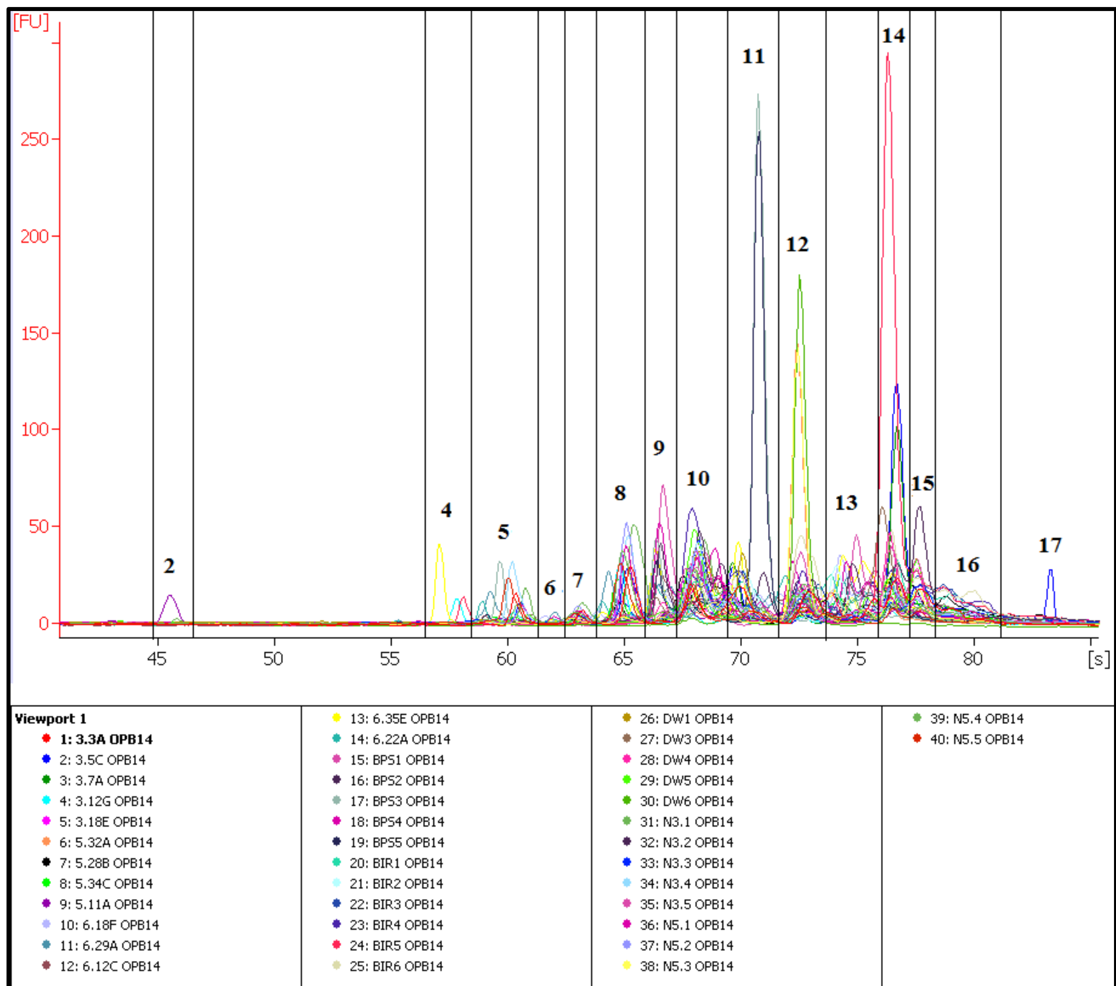

Fig. 3.3.3.13. Electropherogram showing division of peak groups for Patch three samples with primer OPB14

3.3.4 Patch One

On the whole samples in patch one are very similar with the lowest Dice's coefficient of similarity being 0.618 (Table. 3.3.4.1.). Since for Dice's coefficient 1 means samples are identical and 0 is that they are not at all related a reading of 0.618 for the furthest related individuals suggests a close similarity between all samples within this patch.

Table 3.3.4.1. Matrix of results for Dice's similarity coefficient between samples from patch one where 1= identical and 0=no similarity. Lighter shading indicates higher similarity between samples.

	1E1.1	1E1.2	1E1.3	1E1.4	1E1.5	1M1	1M2	1M3	1M4	1M5	1E2.1	1E2.2	1E2.3	1E2.4	1E2.5	1NE1.1	1NE1.2	1NE1.3	1NE1.4	1NE2.1	1NE2.2	1NE2.3
1E1.1	1	0.754	0.769	0.754	0.754	0.754	0.794	0.812	0.75	0.754	0.714	0.821	0.75	0.8	0.746	0.704	0.806	0.676	0.667	0.694	0.747	0.784
1E1.2	0.754	1	0.69	0.677	0.71	0.71	0.689	0.677	0.632	0.742	0.73	0.732	0.737	0.69	0.7	0.719	0.677	0.625	0.643	0.677	0.618	0.746
1E1.3	0.769	0.69	1	0.727	0.697	0.788	0.738	0.697	0.754	0.758	0.746	0.773	0.656	0.71	0.813	0.706	0.754	0.647	0.733	0.696	0.722	0.761
1E1.4	0.754	0.677	0.727	1	0.714	0.714	0.754	0.771	0.708	0.714	0.789	0.759	0.708	0.788	0.765	0.722	0.74	0.639	0.688	0.849	0.711	0.693
1E1.5	0.754	0.71	0.697	0.714	1	0.829	0.754	0.714	0.738	0.743	0.732	0.81	0.708	0.818	0.706	0.722	0.767	0.750	0.656	0.712	0.711	0.720
1M1	0.754	0.71	0.788	0.714	0.829	1	0.841	0.8	0.8	0.829	0.761	0.81	0.738	0.758	0.765	0.722	0.74	0.750	0.750	0.740	0.763	0.747
1M2	0.794	0.689	0.738	0.754	0.754	0.841	1	0.812	0.813	0.783	0.714	0.795	0.75	0.738	0.836	0.732	0.778	0.676	0.762	0.778	0.773	0.730
1M3	0.812	0.677	0.697	0.771	0.714	0.8	0.812	1	0.769	0.743	0.704	0.785	0.769	0.818	0.794	0.694	0.74	0.722	0.719	0.795	0.789	0.693
1M4	0.75	0.632	0.754	0.708	0.738	0.8	0.813	0.769	1	0.769	0.667	0.73	0.7	0.754	0.762	0.627	0.706	0.627	0.712	0.676	0.648	0.629
1M5	0.754	0.742	0.758	0.714	0.743	0.829	0.783	0.743	0.769	1	0.817	0.734	0.677	0.727	0.765	0.778	0.74	0.722	0.656	0.740	0.763	0.720
1E2.1	0.714	0.73	0.746	0.789	0.732	0.761	0.714	0.704	0.667	0.817	1	0.775	0.758	0.746	0.725	0.74	0.676	0.685	0.708	0.757	0.727	0.737
1E2.2	0.821	0.732	0.773	0.759	0.81	0.81	0.795	0.785	0.73	0.734	0.775	1	0.73	0.747	0.779	0.741	0.805	0.741	0.740	0.780	0.753	0.857
1E2.3	0.75	0.737	0.656	0.708	0.708	0.738	0.75	0.769	0.7	0.677	0.758	0.73	1	0.787	0.762	0.716	0.706	0.657	0.610	0.676	0.704	0.657
1E2.4	0.8	0.69	0.71	0.788	0.818	0.758	0.738	0.818	0.754	0.727	0.746	0.747	0.787	1	0.781	0.706	0.754	0.647	0.633	0.725	0.694	0.676
1E2.5	0.746	0.7	0.813	0.765	0.706	0.765	0.836	0.794	0.762	0.765	0.725	0.779	0.762	0.781	1	0.743	0.761	0.657	0.742	0.789	0.757	0.740
1NE1.1	0.704	0.719	0.706	0.722	0.722	0.722	0.732	0.694	0.627	0.778	0.74	0.741	0.716	0.706	0.743	1	0.827	0.730	0.606	0.827	0.769	0.753
1NE1.2	0.806	0.677	0.754	0.74	0.767	0.74	0.778	0.74	0.706	0.74	0.676	0.805	0.706	0.754	0.761	0.827	1	0.773	0.657	0.789	0.835	0.821
1NE1.3	0.676	0.625	0.647	0.639	0.75	0.75	0.676	0.722	0.627	0.722	0.685	0.741	0.657	0.647	0.657	0.73	0.773	1.000	0.606	0.747	0.846	0.753
1NE1.4	0.667	0.643	0.733	0.688	0.656	0.75	0.762	0.719	0.712	0.656	0.708	0.74	0.61	0.633	0.742	0.606	0.657	0.606	1.000	0.687	0.629	0.783
1NE2.1	0.694	0.677	0.696	0.849	0.712	0.74	0.778	0.795	0.676	0.74	0.757	0.78	0.676	0.725	0.789	0.827	0.789	0.747	0.687	1.000	0.861	0.744
1NE2.2	0.747	0.618	0.722	0.711	0.711	0.763	0.773	0.789	0.648	0.763	0.727	0.753	0.704	0.694	0.757	0.769	0.835	0.846	0.629	0.861	1.000	0.790
1NE2.3	0.784	0.746	0.761	0.693	0.72	0.747	0.73	0.693	0.629	0.72	0.737	0.857	0.657	0.676	0.74	0.753	0.821	0.753	0.783	0.744	0.790	1.000

Despite the overall similarity some groups do form, one group is formed between samples 1NE2.1 and 1NE2.2 (Fig. 3.3.4.1) which are very similar with a Dice's coefficient of 0.861 (Table. 3.3.4.1) and more loosely with samples 1E1.4, 1NE1.3, 1NE1.3, 1NE1.1, samples 1E2.2 and 1NE2.3 (which are also very similar to one another with a Dice coefficient of 0.857 (Table. 3.3.4.1)) and sample 1E1.1 (Fig. 3.3.4.1). A second group forms between samples 1M1, 1M2, 1E2.5, 1M5, 1E1.5, 1M3, 1E2.4 and 1E2.1. The first and second group form part of a larger group of samples 1M4 and 1E1.3, whilst samples 1E2.3, 1NE1.4 and 1E1.2 do not group.

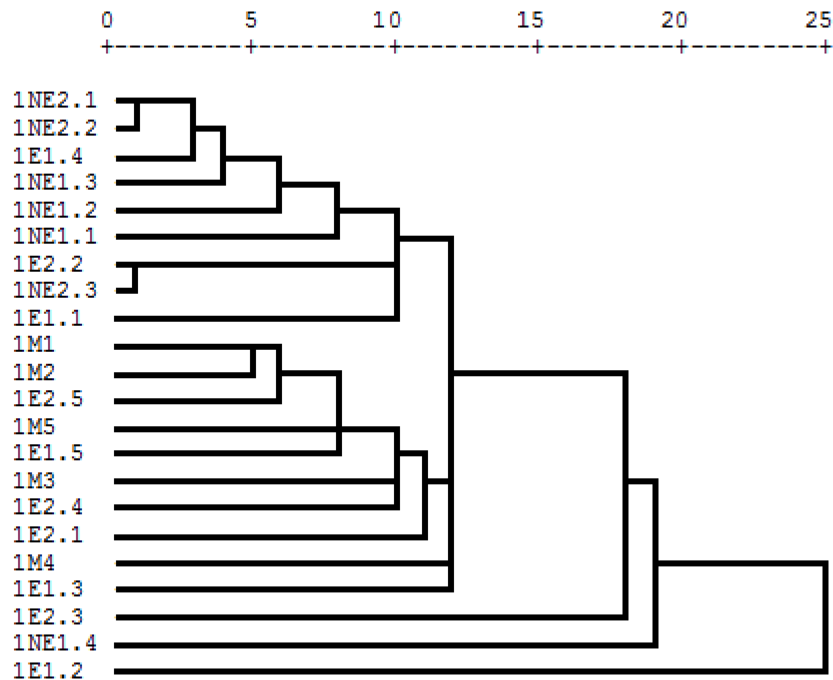

Fig. 3.3.4.1. Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples for patch 1

Despite the overall similarity some groups do form, one group is formed between samples 1NE2.1 and 1NE2.2 (Fig. 3.3.4.1) which are very similar with a Dice's coefficient of 0.861 (Table. 3.3.4.1) and more loosely with samples 1E1.4, 1NE1.3, 1NE1.3, 1NE1.1, samples 1E2.2 and 1NE2.3 (which are also very similar to one

another with a Dice coefficient of 0.857 (Table. 3.3.4.1)) and sample 1E1.1 (Fig. 3.3.4.1). A second group forms between samples 1M1, 1M2, 1E2.5, 1M5, 1E1.5, 1M3, 1E2.4 and 1E2.1. The first and second group form part of a larger group of samples 1M4 and 1E1.3, whilst samples 1E2.3, 1NE1.4 and 1E1.2 do not group.

Genetic distinction between different sites within the patch is minimal. All windward samples (1E1.X) and leeward samples (1E2.X) are found across all groups and amongst those which do not group out (Fig. 3.3.4.1.). This would suggest that windward (1E1.X) and leeward samples (1E2.X) have little similarity to individuals from the same site (1E1 or 1E2) and that very little to no genetic selection is occurring at windward or leeward sites. The majority of *B. nana* (1NE X.X) group together with the exception of 1NE1.4 (a *B. nana* from the windward site) (Fig. 3.3.4.1.). This would indicate that *B. nana* shrubs share a high degree of genetic similarity in this patch. However since the group in which they are found also includes *B. pubescens* samples (1E1.4, 1E2.2 and 1E1.1) (Fig. 3.3.4.1.), and sample 1NE1.4 bears little similarity to any samples in this patch, it would indicate that this similar genetic makeup in these *B. nana* samples is not restricted to *B. nana* type plants. The *B. pubescens* samples that group out with the majority of *B. nana* samples also show a range of morphologies, 1E1.1 and 1E1.4 are polycormic whilst 1E2.2 is not polycormic (Table 3.3.2.1.). This would indicate that variation in observed morphologies of *B. pubescens* is not due to closer (for more polycormic individuals) or more distant (for less polycormic individuals) genetic similarity to *B. nana*.

The majority of the middle site samples (1MX) are found in the second group except for sample 1M4 which is intermediate between groups one and two (Fig. 3.3.4.1.). 1M4 is the only non-polycormic individual from the middle site (1MX), however sample 1E2.1 is also non-polycormic (Table. 3.3.2.1.) and is found in this second group (Fig. 3.3.4.1), thus indicating again that morphology is not linked to any selection occurring at these sites. However there would appear to be genetic selection occurring at the middle site for samples of similar genetic makeup. Since central sites suffer greatly from dehydration stress which is one of the major killers of saplings (Kullman, 1986) individuals with greater tolerance of drought stress, and therefore potentially more closely related individuals with a similar genotype may be selected for at this site.

3.3.5 Patch Two

Samples 2E2.1 and 2E2.2 have an overall Dice similarity of 0.919 (Table 3.3.5.1.), with very few different peaks. Due to the aforementioned degree of error in peak identification this could indicate that these trees may be the same individual from different root suckers mistakenly identified as two separate individuals and are at the very least close relatives. These samples group with sample 2NE1.4 which may indicate an ancestor or descendant of these two samples (Fig. 3.3.5.1). Samples 2E1.2 and 2E.14 also show a strong genetic similarity with a Dice coefficient of 0.881 (Table. 3.3.5.1.) this is also very high which may indicate that these samples are also very closely related, possibly siblings. The lowest similarity score across the entire patch is 0.526, (Table. 3.3.5.1.) this score comes from sample 2M2 which shows very little similarity to any other samples in the patch (Table. 3.3.5.1. and Fig.

3.3.5.1.). If sample 2M2 is excluded a higher level of similarity amongst the patch is seen with the lowest score being 0.618 (Table. 3.3.5.1.). The distinct lack of relatedness of 2M2 to any of the other individuals sampled in the patch may either indicate problems with the PCR or bioanalyser for one of the primers used with 2M2 or, more likely, that sample 2M2 is simply unrelated to the others, perhaps from seed/pollen from further afield than the other trees sampled.

No obvious large groups are formed (Fig. 3.3.5.1.). Those groups that are formed have very high similarity scores indicating immediate family relation between a small number of samples. There is no evidence of differences between the sites, despite a wide range of morphologies (Table. 3.3.2.1.) supporting earlier indications that morphology is not linked to the overall genetic makeup of an individual. Sites from patch two are very close together (Fig. 2.1.1.1.), this may explain the genetic similarity seen across all samples bar 2M2. Unlike the other patches at least one site is not adjacent to a lake (Fig. 2.1.1.1.). It is likely that the conditions across this patch with the exception of wind speed and snow deposition are likely to be fairly similar between the sites sampled which may explain the little to no overall genetic selection occurring between sites in this patch.

Table 3.3.5.1. Matrix of results for Dice's similarity coefficient between samples from patch two where 1= identical and 0=no similarity. Lighter shading indicates higher similarity between samples.

	2E1.2	2E1.3	2E1.4	2E1.5	2M1	2M2	2M3	2M4	2M5	2E2.1	2E2.2	2E2.3	2E2.4	2E2.5	2NE1.1	2NE1.2	2NE1.3	2NE1.4	2NE1.5	2NE2.1	2NE2.2	2NE2.3	2NE2.4	2NE2.5
2E1.2	1.000	0.702	0.881	0.754	0.735	0.667	0.742	0.730	0.727	0.708	0.746	0.769	0.746	0.750	0.762	0.667	0.721	0.708	0.781	0.742	0.667	0.698	0.687	0.714
2E1.3	0.702	1.000	0.759	0.733	0.657	0.596	0.689	0.806	0.708	0.688	0.758	0.719	0.724	0.676	0.645	0.774	0.633	0.719	0.635	0.721	0.649	0.677	0.727	0.667
2E1.4	0.881	0.759	1.000	0.774	0.725	0.612	0.698	0.719	0.716	0.758	0.794	0.727	0.767	0.740	0.781	0.719	0.806	0.727	0.738	0.730	0.711	0.719	0.765	0.676
2E1.5	0.754	0.733	0.774	1.000	0.732	0.588	0.677	0.758	0.696	0.735	0.714	0.706	0.742	0.773	0.667	0.818	0.719	0.765	0.716	0.769	0.692	0.697	0.714	0.712
2M1	0.735	0.657	0.725	0.732	1.000	0.586	0.806	0.685	0.789	0.773	0.805	0.773	0.783	0.780	0.712	0.712	0.704	0.747	0.784	0.639	0.776	0.767	0.779	0.825
2M2	0.667	0.596	0.612	0.588	0.586	1.000	0.577	0.604	0.643	0.655	0.596	0.655	0.571	0.548	0.528	0.604	0.549	0.545	0.593	0.538	0.554	0.528	0.526	0.567
2M3	0.742	0.689	0.698	0.677	0.806	0.577	1.000	0.746	0.800	0.725	0.732	0.754	0.794	0.711	0.687	0.687	0.677	0.812	0.794	0.697	0.734	0.776	0.789	0.784
2M4	0.730	0.806	0.719	0.758	0.685	0.604	0.746	1.000	0.732	0.714	0.750	0.686	0.688	0.753	0.618	0.794	0.727	0.771	0.696	0.716	0.750	0.706	0.694	0.693
2M5	0.727	0.708	0.716	0.696	0.789	0.643	0.800	0.732	1.000	0.822	0.800	0.740	0.776	0.800	0.761	0.789	0.754	0.795	0.806	0.743	0.795	0.817	0.773	0.769
2E2.1	0.708	0.688	0.758	0.735	0.773	0.655	0.725	0.714	0.822	1.000	0.919	0.750	0.758	0.759	0.686	0.743	0.794	0.861	0.732	0.725	0.756	0.743	0.811	0.727
2E2.2	0.746	0.758	0.794	0.714	0.805	0.596	0.732	0.750	0.800	0.919	1.000	0.784	0.765	0.765	0.694	0.722	0.771	0.838	0.712	0.704	0.762	0.778	0.789	0.734
2E2.3	0.769	0.719	0.727	0.706	0.773	0.655	0.754	0.686	0.740	0.750	0.784	1.000	0.818	0.709	0.686	0.686	0.676	0.750	0.732	0.783	0.756	0.743	0.730	0.753
2E2.4	0.746	0.724	0.767	0.742	0.783	0.571	0.794	0.688	0.776	0.758	0.765	0.818	1.000	0.712	0.688	0.688	0.710	0.788	0.769	0.762	0.711	0.781	0.735	0.761
2E2.5	0.750	0.676	0.740	0.773	0.780	0.548	0.711	0.753	0.800	0.759	0.765	0.709	0.712	1.000	0.805	0.753	0.773	0.785	0.744	0.763	0.809	0.753	0.765	0.810
2NE1.1	0.762	0.645	0.781	0.667	0.712	0.528	0.687	0.618	0.761	0.686	0.694	0.686	0.688	0.805	1.000	0.676	0.818	0.714	0.783	0.776	0.725	0.765	0.778	0.720
2NE1.2	0.667	0.774	0.719	0.818	0.712	0.604	0.687	0.794	0.789	0.743	0.722	0.686	0.688	0.753	0.676	1.000	0.727	0.714	0.754	0.746	0.750	0.765	0.722	0.720
2NE1.3	0.721	0.633	0.806	0.719	0.704	0.549	0.677	0.727	0.754	0.794	0.771	0.676	0.710	0.773	0.818	0.727	1.000	0.765	0.776	0.769	0.795	0.788	0.829	0.658
2NE1.4	0.708	0.719	0.727	0.765	0.747	0.545	0.812	0.771	0.795	0.861	0.838	0.750	0.788	0.785	0.714	0.714	0.765	1.000	0.732	0.783	0.805	0.771	0.838	0.805
2NE1.5	0.781	0.635	0.738	0.716	0.784	0.593	0.794	0.696	0.806	0.732	0.712	0.732	0.769	0.744	0.783	0.754	0.776	0.732	1.000	0.735	0.765	0.783	0.740	0.763
2NE2.1	0.742	0.721	0.730	0.769	0.639	0.538	0.697	0.716	0.743	0.725	0.704	0.783	0.762	0.763	0.776	0.746	0.769	0.783	0.735	1.000	0.759	0.776	0.789	0.757
2NE2.2	0.667	0.649	0.711	0.692	0.776	0.554	0.734	0.750	0.795	0.756	0.762	0.756	0.711	0.809	0.725	0.750	0.795	0.805	0.765	0.759	1.000	0.825	0.810	0.828
2NE2.3	0.698	0.677	0.719	0.697	0.767	0.528	0.776	0.706	0.817	0.743	0.778	0.743	0.781	0.753	0.765	0.765	0.788	0.771	0.783	0.776	0.825	1.000	0.833	0.800
2NE2.4	0.687	0.727	0.765	0.714	0.779	0.526	0.789	0.694	0.773	0.811	0.789	0.730	0.735	0.765	0.778	0.722	0.829	0.838	0.740	0.789	0.810	0.833	1.000	0.810
2NE2.5	0.714	0.667	0.676	0.712	0.825	0.567	0.784	0.693	0.769	0.727	0.734	0.753	0.761	0.810	0.720	0.720	0.658	0.805	0.763	0.757	0.828	0.800	0.810	1.000

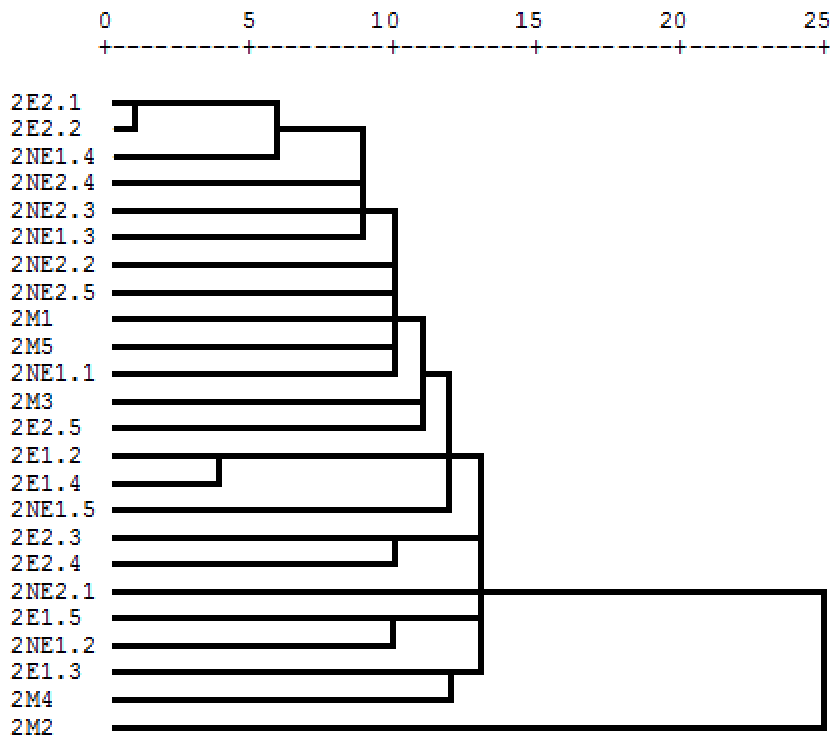

Fig. 3.3.5.1. Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples for patch 2.

3.3.6. Patch Three

Two main groups are formed in this patch. The first group is made up of several smaller groups consisting of samples N3.4, N5.5, N5.4 ,N5.2 which then group again with samples N3.3 and N5.1, which then group with 5.28B, 5.34C (which have high similarity to one another) and 6.18F, and finally with samples N3.1 N3.5, 6.35E (Fig. 3.3.6.1). The second group is not made up of smaller groups and contains samples 6.29A, 6.22A, 3.18E, 6.12C, 3.12G, 3.7A and 3.5C (Fig. 3.3.6.1). These two groups along with samples 5.11A and 3.3A form a larger overall group whilst samples 5.32A and N5.3 do not group out.

In patch three the majority of *B. nana* (NX.X) samples are found in group one bar sample N5.3 which does not group (Fig. 3.3.6.1.), however group one does not consist entirely of *B. nana*, with leeward samples 5.28B, 5.34C and central samples 6.18F and 6.35E also present within the group (Fig. 3.3.6.1.). As found with patch one this would indicate that although the majority of *B. nana* share a similar genetic makeup it is not exclusive to *B. nana* samples or samples with a morphology similar to that of *B. nana*. The leeward samples (5.X) are found everywhere except group two, which could be co-incidental due to the small number of samples relative to the number of groups, but could also indicate that very little selection occurs ar the leeward site with the exception of individuals with genetic make-ups found in the second group being excluded. The second group consists of central site samples (6.X) which are found in both group one and two, and windward samples (3.X) which are only found in group two with the exception of 3.3A which is intermediate between groups one and two. This would suggest that central sites (site six) can support individuals of either group one or group two genetic makeup whereas the windward site (site three) selects for individuals with similar genetic makeup as seen in group two. This selection of specific genotypes for windward sites is not seen in the other patches and would therefore suggest that the selecting factor at this site is something not seen in the other windward sites. Since all windward sites have similar morphologies, and should have similar patterns of wind/ice blast and snow build-up, it would indicate that the driving factor promoting genetic similarity at site three in patch three is not one of these factors and in order for this to be uncovered a range of environmental studies at all sites would need to be undertaken. The presence of central site samples (6.X) in both groups may also indicate that the central site can act as bridge between the windward samples and all other samples.

Since central sites suffer from high levels of drought it could indicate that windward samples from patch three suffer from greater levels of drought than their windward counterparts in patches one and two. This site also supports the indication that the short and polycormic morphology of windward individuals is not due to close relations with *B. nana*.

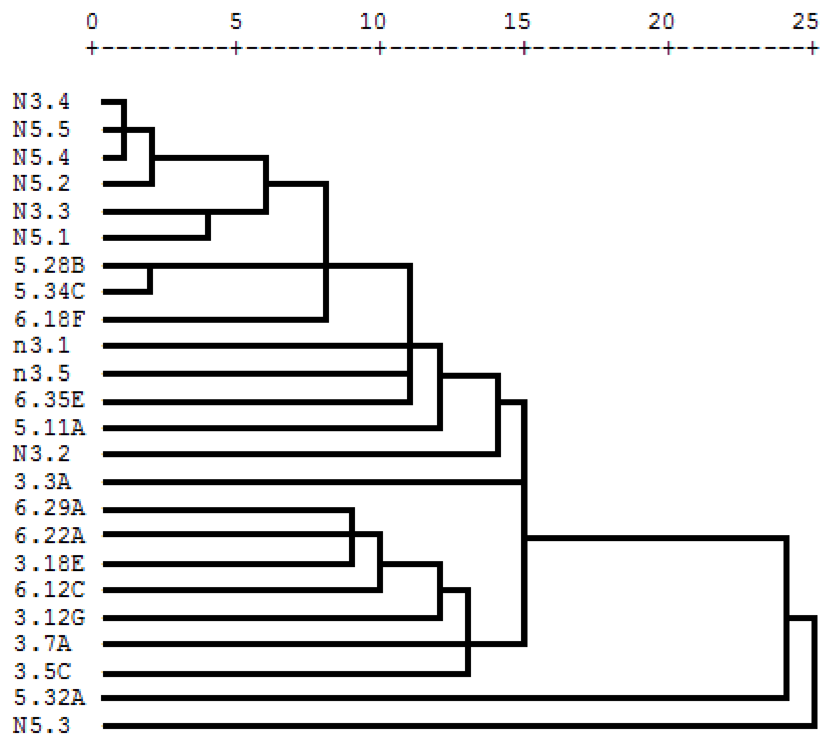

Fig. 3.3.6.1. Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples for patch 3.

Table. 3.3.6.1. Matrix of results for Dice's similarity coefficient between samples from patch three where 1= identical and 0=no similarity. Lighter shading indicates higher similarity between samples.

	3.3A	3.5C	3.7A	3.12G	3.18E	5.32A	5.28B	5.34C	5.11A	6.18F	6.29A	6.12C	6.35E	6.22A	n3.1	n3.2	n3.3	n3.4	n3.5	n5.1	n5.2	n5.3	n5.4	n5.5
3.3A	1.000	.758	.787	.813	.833	.655	.836	.836	.813	.813	.800	.704	.800	.800	.758	.678	.765	.806	.644	.783	.794	.655	.824	.824
3.5C	.758	1.000	.852	.781	.767	.655	.776	.776	.750	.813	.767	.741	.831	.800	.727	.712	.735	.806	.712	.754	.824	.724	.794	.794
3.7A	.787	.852	1.000	.746	.800	.680	.774	.774	.814	.814	.836	.857	.767	.800	.721	.704	.667	.774	.667	.719	.794	.717	.762	.794
3.12G	.813	.781	.746	1.000	.862	.717	.769	.800	.806	.774	.759	.692	.825	.828	.813	.702	.818	.800	.702	.806	.727	.607	.758	.788
3.18E	.833	.767	.800	.862	1.000	.735	.787	.787	.828	.724	.815	.750	.780	.889	.733	.717	.774	.787	.717	.762	.742	.577	.742	.806
5.32A	.655	.655	.680	.717	.735	1.000	.714	.679	.679	.717	.694	.698	.667	.735	.764	.750	.667	.714	.667	.724	.632	.511	.667	.702
5.28B	.836	.776	.774	.769	.787	.714	1.000	.941	.862	.862	.754	.655	.758	.787	.806	.767	.812	.853	.733	.829	.841	.678	.841	.899
5.34C	.836	.776	.774	.800	.787	.679	.941	1.000	.862	.892	.754	.655	.788	.787	.776	.733	.812	.853	.733	.829	.841	.712	.841	.899
5.11A	.813	.750	.814	.806	.828	.679	.862	.862	1.000	.839	.759	.692	.730	.828	.844	.807	.788	.831	.737	.836	.758	.714	.788	.818
6.18F	.813	.813	.814	.774	.724	.717	.862	.892	.839	1.000	.793	.731	.825	.793	.844	.737	.788	.831	.702	.806	.818	.750	.818	.848
6.29A	.800	.767	.836	.759	.815	.694	.754	.754	.759	.793	1.000	.875	.814	.889	.733	.679	.710	.721	.566	.762	.742	.731	.742	.742
6.12C	.704	.741	.857	.692	.750	.698	.655	.655	.692	.731	.875	1.000	.755	.833	.630	.638	.607	.691	.596	.632	.679	.652	.679	.679
6.35E	.800	.831	.767	.825	.780	.667	.758	.788	.730	.825	.814	.755	1.000	.814	.800	.724	.866	.848	.724	.824	.866	.702	.866	.836
6.22A	.800	.800	.800	.828	.889	.735	.787	.787	.828	.793	.889	.833	.814	1.000	.767	.755	.774	.820	.679	.762	.742	.692	.774	.774
n3.1	.758	.727	.721	.813	.733	.764	.806	.776	.844	.844	.733	.630	.800	.767	1.000	.814	.853	.806	.746	.870	.765	.655	.794	.794
n3.2	.678	.712	.704	.702	.717	.750	.767	.733	.807	.737	.679	.638	.724	.755	.814	1.000	.787	.833	.846	.839	.754	.706	.787	.787
n3.3	.765	.735	.667	.818	.774	.667	.812	.812	.788	.788	.710	.607	.866	.774	.853	.787	1.000	.899	.820	.930	.886	.700	.886	.886
n3.4	.806	.806	.774	.800	.787	.714	.853	.853	.831	.831	.721	.691	.848	.820	.806	.833	.899	1.000	.867	.914	.928	.712	.957	.957
n3.5	.644	.712	.667	.702	.717	.667	.733	.733	.737	.702	.566	.596	.724	.679	.746	.846	.820	.867	1.000	.806	.820	.667	.820	.852
n5.1	.783	.754	.719	.806	.762	.724	.829	.829	.836	.806	.762	.632	.824	.762	.870	.839	.930	.914	.806	1.000	.873	.721	.901	.901
n5.2	.794	.824	.794	.727	.742	.632	.841	.841	.758	.818	.742	.679	.866	.742	.765	.754	.886	.928	.820	.873	1.000	.733	.943	.943
n5.3	.655	.724	.717	.607	.577	.511	.678	.712	.714	.750	.731	.652	.702	.692	.655	.706	.700	.712	.667	.721	.733	1.000	.733	.700
n5.4	.824	.794	.762	.758	.742	.667	.841	.841	.788	.818	.742	.679	.866	.774	.794	.787	.886	.957	.820	.901	.943	.733	1.000	.943
n5.5	.824	.794	.794	.788	.806	.702	.899	.899	.818	.848	.742	.679	.836	.774	.794	.787	.886	.957	.852	.901	.943	.700	.943	1.000

Greater differences are seen between samples in this group with the lowest Dice similarity of 0.511. This is despite larger groupings of peaks at this site than at other sites (Fig. 3.3.3.10 – Fig. 3.3.3.13) which should result in greater similarity as there is a higher chance of individuals with different peaks being grouped together under the same peak. This strongly indicates that patch three shows the greatest level of variation between samples. The sites (windward, central and leeward) from which samples were taken in patch three are much further removed from one another than in patches one and two (Fig. 2.1.1.1.). Hence the greater variation between sites in patch three could be due to the greater distance between sample sites. Potentially a wider range of environmental conditions occur over the greater distances between sites selecting for different genotypes.

3.3.7. All Patches

All of the patches were also compared to one another to see if particular patches, species or site locations (windward, leeward or central) grouped together. For this to occur the peak groups of the patches with more well defined peak groups had to be matched with those of the patch with the least well defined peak groups (patch three) (Fig. 3.3.3.10 – Fig. 3.3.3.13.). For example peak groups 3a, 3b and 3c from patch one with primer ATC51 (Fig. 3.3.3.2.) had to be grouped together to match patch peak group three in patch three with primer ATC51 (Fig. 3.3.3.10.).

Several groups were formed when all patches were compared (Fig. 3.3.7.1). Group one consisted of several smaller groups all from patch three. The first smaller group consisted of N3.4, N5.5, N5.4, N5.3 N3.3 N5.1. A second smaller group consisted of

samples 5.28B, 5.34C; these two smaller groups then formed a group together along with sample 6.18F. This slightly larger group formed an even larger group with samples N3.1, N3.5, 6.35E and 5.11A which then grouped with samples N3.2, 3.3A. Group two consisted of a much smaller group of two samples of relatively low similarity 2E2.5 and 2NE1.1. A third group formed between samples 6.29A, 6.22A, 3.18E, 6.12C, 3.12G, 3.7A and 3.5C. Finally the fourth group consisted of two smaller groups. The first smaller group being between samples 1E1.1, 1M3, 1E1.4, 1NE 2.1, 1NE2.2, 1M2, 2NE1.2, 1E2.5, 1M4 and the second smaller group between samples 1M5, 1E2.1, 2E2.1, 2E2.2, 1NE1.4, 2NE2.5, 2NE2.2, 2NE1.3, 2NE2.4, 2M5, 2NE2.3, 2E2.3, 2E2.4, 2M1, 2M3, 1E1.5, 1M1, 1E2.4. These two smaller groups joined with samples 2NE1.5, 2NE2.1, 2E1.5, 1E2.2, 1NE2.3, 1NE1.3, 1NE1.2 and1E1.3 to form the complete fourth group. Groups one, two, three and four joined with samples 2E1.3, 1E2.3, 2M4 to form one final big group. Several samples did not group and these were 2E1.2, 2E1.4 (which are closely related to one another) 1NE1.1, 1NE.14, 1E1.2 and 5.32A, N5.3 AND 2M2 (Fig. 3.3.7.1).

Patch three individuals occurred in two groups distinct from the other patches with the exception of samples 5.32A and N5.3 which did not group. These two main groups formed in patch three are the same ones seen when analysed on its own. Groups two and four consist of a mixture of patch one and two individuals. The only site or species grouping similarity for patches one and two is seen in *B nana* from patch two (2NE2.X) which are mostly in a smaller group within group four along with the majority of patch two leeward *B. pubescens* (2E2.X) (Fig. 3.3.7.1.). There is otherwise very little evidence of genotypic differences between patch one and two, *B. nana* and *B. pubescens* or between samples from the three different types of site (windward, central or leeward).

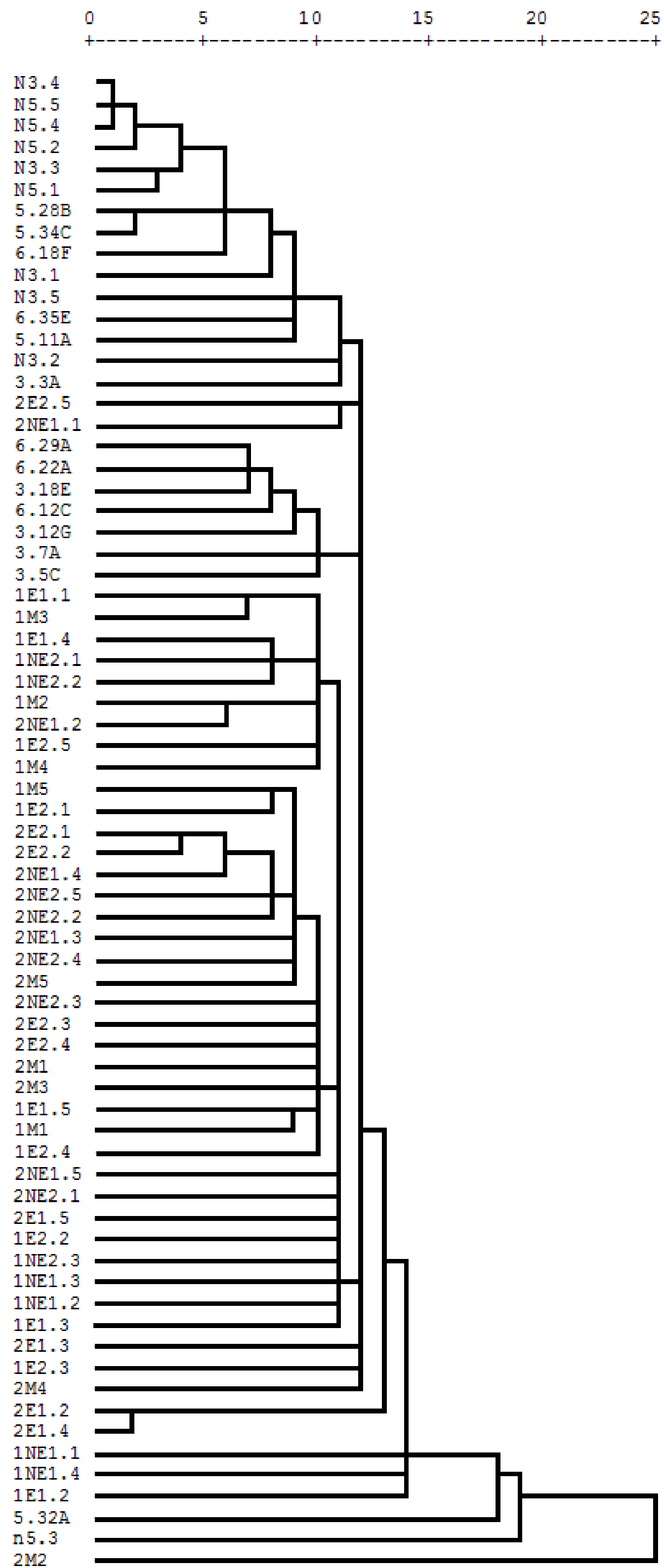

Fig. 3.3.7.1. Dendrogram based on Dice's coefficient of similarity and scaled. Showing similarity between samples across all patches.

There is a greater distinction seen between patch three and the other patches (one and two) than between species or site locations (windward, central or leeward). Patches one and two are more closely related to one another than to patch three, most likely due to patches one and two having greater proximity to one another than to patch three (Fig. 2.1.1.1.). This would indicate that either different selection pressures are occurring at these two patches than at patch three and that particular genotypes have a greater chance of survival in patches one and two than in patch three or that seed and pollen dispersal is more limited than initially thought (Pers. Observation. Pers. comms: R. Baxter, Durham University, Greene & Johnson, 1997 and Matlack, 1989). Since certain individuals from all patches bear little resemblance to any of their main patch members (2E1.2, 2E1.4 ,1NE1.1, 1NE.14, 1E1.2, 5.32A, N5.3 and2M2 (Fig. 3.3.7.1.)) it would suggest that longer range pollen and seed dispersal can occur but this is either infrequent or that local individuals are favoured. A variety of wing loading ratios can be found in *Betula* species which affects seed distribution (Gallagher et al., 2011). Seeds from *B. pubescens* and *B. nana* also show a variety of wing lengths and seed sizes (Personal observation). Seeds with high weight to wingspan ratios are more likely to remain in the same area whilst seeds with lower seed weight to wing-area may travel further (Gallagher et al., 2011). Some seeds are also retained by the seed parent throughout the winter (Pers. com. R. Holden and R. Baxter, Durham University). The fertility of these retained seeds is unknown, however if fertile these seeds may be dropped after winter and remain close to parent plants. Both of these potential methods can ensure that some seeds remain in the same area as their parent where they may have a particular advantage over unrelated saplings, but some seeds also disperse further afield to colonise new similar areas.

Patch three has the strongest evidence of separate groups occurring within the patch (Fig. 3.3.7.1.). Although the sites in patch three are more distant from one another than sites in either patch one and two (Fig. 2.1.1.1.), they are not more distant from one another than patch one is from patch two. Since patch three shows strong differences between samples and sites whereas patches one and two show little difference between one another, it would indicate that the difference between sites in patch three is due to selection for particular genotypes as opposed to only the effects of limited seed dispersal. This also suggests that a wider range of environmental conditions occur across the sites in patch three than across the entirety of patches one and two.

Patch one shows very few site selection differences and patch two has little to none despite both patches having higher peak resolutions than patch three. Since the sites across patches one and two are closer together there may be less variation in environmental conditions across these patches than seen in patch three. The leeward site of patch one (1E2) is adjacent to a lake and the windward and central sites have a lake to the north of them This may indicate that the whole of patch one is relatively wet which could explain the limited degree of selection occurring across the patch. Patch two has no evidence of close proximity to lakes which may indicate all sites are drier in patch two. Although this would explain the overall genetic similarity of individuals in this patch it does not explain their similarity to individuals in patch one. There is however evidence of watercourses to the northeast of patch two which may indicate that patch two is wetter than at first assumed only that adjacent lakes have not formed. Also unknown is what the conditions of the patches were when these trees first became established, which is likely to be when the greatest selection

due to limited water supply occurred, since it is the seedlings that are at greatest risk from drought (Kullman, 1986). As 95 year old saplings have been found above the treeline (Hofgaard, Dalen & Håkan, 2009) the trees in these patches could be of great age and watercourses may well have changed since their initial establishment. The leeward site of patch three (5.X) is also situated next to a lake (Fig. 2.1.1.1.) and shows relatively little evidence of any genetic selection occurring, whereas samples from the presumed drier windward site of patch three (3.X) do show evidence of genetic selection. Both *B. nana* and *B. pubescens* tolerate wetter conditions than most *Betula* species and in more southern latitudes are restricted to wetter areas (Walters, 1968, Jonasson, 1981, Horsfield & Thompson, 1997, Kirkpatrick & Heal, 2001 and Ejankowski & Kunz, 2006). Consequently in wetter conditions most individuals may be able to survive whereas in drier conditions certain genotypes may be selected for. Leeward individuals from patch three (5.X) although showing no strong evidence of undergoing selection within this patch are still distinctly more closely related to members of patch three (including windward site three samples) than to patches one and two with the exception of sample 5.32A which does not fall within any groups (Fig. 3.3.7.1.). This could be a relic from heavier seeds travelling shorter distances, and so the majority of site five samples have genetic similarity to their closer neighbours (i.e. other members of patch three). Of greater probability is that although water availability is key to seedling survival (Kullman, 1986), a wide range of environmental conditions either currently or previously affected selection of particular genotypes amongst all sites and patches. Therefore it seems likely that during patch establishment the environmental conditions at patch three were different from those at patches one and two.

3.3.8 Conclusions and Summary

Both morphology and species are on the whole irrelevant to relationships between samples, with a wide array of morphologies and species grouping together. Although *B. nana* have a tendency to group together within their patch they are often with samples of contrasting morphology and do not group together over all patches, indicating that *B. nana* have more relation to *B. pubescens* from the same/neighbouring patches than to *B. nana* from further afield. The tendency to group within their patch may be due to the lower chromosome number. Pure *B. nana* are diploid whereas pure *B. pubescens* are allopolyploid tetraploids and therefore are more likely to produce a greater number of potential peaks than *B. nana* due to containing two different sets of chromosomes from each ancestral parent. A greater degree of variation is therefore more likely. Observed changes in morphology throughout the patch does not appear to be linked to overall genetic makeup. This would indicate that the morphology seen over all patches is either due to a small number of genes, phenotypic plasticity in response to environmental factors or mechanical damage rather than overall genetic selection or close relation of more polycormic individuals to *B. nana.* This is contrary to a hypothesis forwarded by Vaarama and Valannae, 1973 in Verwijst, 1988, they found that the suckering ability (and hence ability to become polycormic) in *B. nana* and *B. pubescens* was linked to genetic differences between individuals (Vaarama and Valannae, 1973 in Verwijst, 1988) and they theorised the degree of polycormy may be due to hybridization with *B. nana* (Vaarama and Valannae, 1973 in Verwijst, 1988). This current study would refute that the degree of polycormy is due to close genetic relations with *B. nana* although it is possible that if a high degree of introgressive hybridisation is occurring

between these two species, polycormy genes from B. *nana* could be inherited with very few other *B. nana* genes found in the overall genome. All these findings indicate there is a very large degree of introgressive hybridisation occurring between these species in Abisko, northern Sweden.

Small scale genetic selection appears to occur both within and between some patches. This selection is likely to be linked to local conditions and the potentially limited dispersal of seeds with a higher weight to wing ratio may ensure that some of the seeds from parent trees remain in the same area where they are more likely to be adapted to the conditions at that site. Although water availability seems a probable environmental candidate for selection of certain genotypes at particular sites it is likely that several environmental factors are affecting selection. In order to address which environmental factors may be affecting selection of particular genotypes, readings for range of current environmental conditions at the sites would need to be taken. Since established trees are under consideration it is the conditions at the point of establishment which are most likely to be involved in selection of individuals, so paleoecological evidence of conditions at the point of establishment would be more indicative of the actual conditions that drove selection of particular genotypes in saplings. Since the ages of these trees are unknown and they could easily have be offshoots from the original main stem which may no longer remain, meaningful age determination of these trees would be difficult or impossible. Obviously, this would make the collection of paleoecological evidence of environmental conditions in the immediate patch site at the time of establishment equally unfeasible.

A variety of problems with the RAPD procedure present themselves. The variation in peak locations was not expected to occur and hence made results less detailed when potentially several peaks were grouped as one. Future use of the bioanalyser for RAPD analysis should therefore use careful selection of primers that will produce well-spaced peaks. The first batch of samples run on the bioanalyser (patch three) suffered the most from the greatest variance in peak locations (Fig. 3.3.3.1 – 3.3.3.13). Prior to this intensive period of use with the RAPD samples the bioanalyser had only been used infrequently for many years suggesting that frequent use of the bioanalyser may increase readout reliability. Due to *B. pubescens* being tetraploid (and likely an allopolyploid) (Järvinen, et al. 2004., Schenk et al., 2008, Walters, 1968 and Howland, Oliver & Davy, 1995). There are potentially up to four different copies of each gene. Since this data is merely presence/absence data and there is no indication of the level of expression or polymorphisms affecting activity levels, a sample with three copies of one version of a gene and one copy of another would have the same readout as a sample with a different ratio of these genes. The unfortunate consequence would be that samples may be classed as genotypically similar to one another when in reality this different ratio may result in different phenotypes and environmental responses. Dominant and recessive genes in each sample are also given equal weight when using RAPD analysis so apparently similar samples may have very different phenotypes as seen. This RAPD method of analysis does however give a good general overview of sample relationships but not selection that may occur at individual gene levels. This general overview combined with levels of gene expression and closer analysis of specific genes related to morphology and environmental stresses, would give greater detail of the forms of genetic

selection that may be occurring at particular sites within these patches and greater explanation of the observed morphologies

3.4 Physiology and Freezing Shock

3.4.1 Light Response Curves

Light response curves for young and old leaves showing that light saturation occurs at approximately 600µmol m^{-2} s^{-1} for the majority of samples regardless of leaf age (Fig. 3.4.1.1 and Fig. 3.4.1.2). These results were used to select light levels for all saplings.

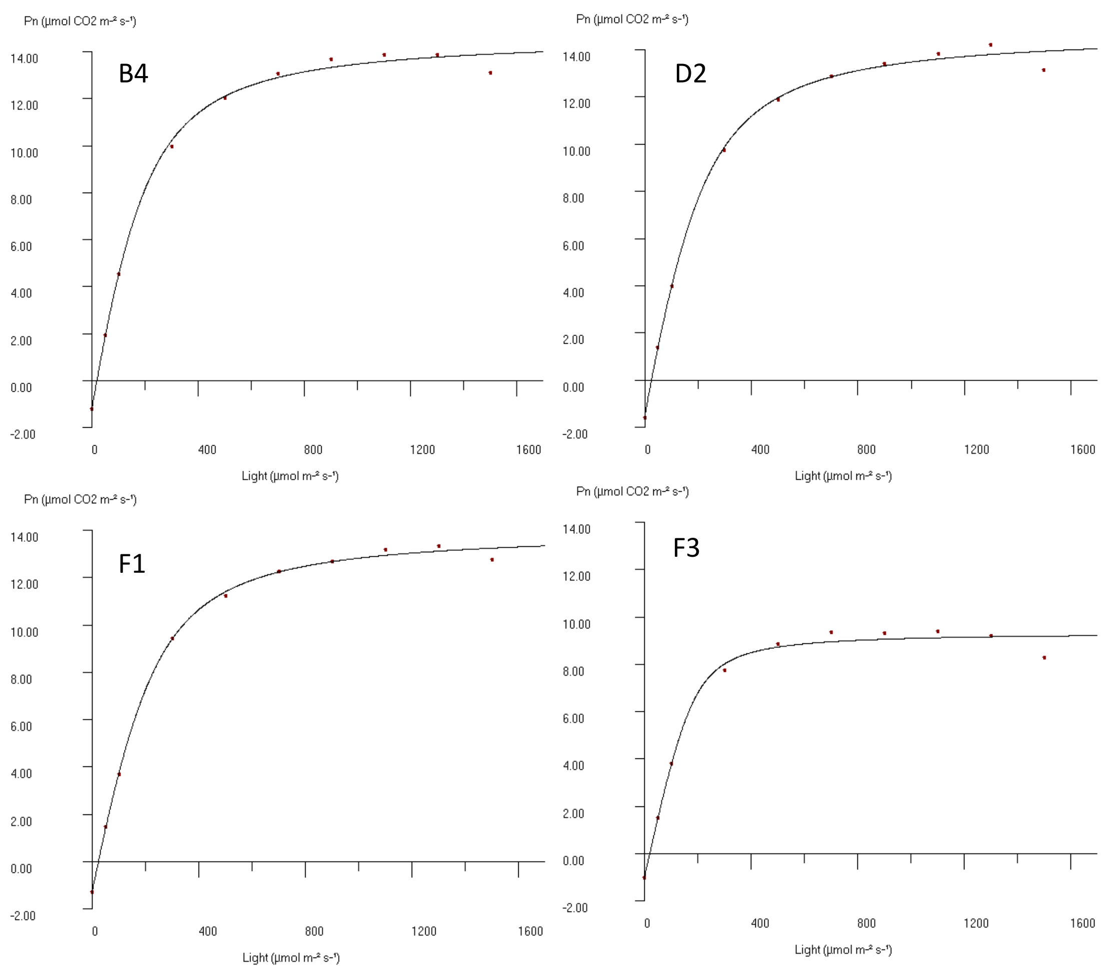

Fig. 3.4.1.1. Light response curves of photosynthesis for older leaves.

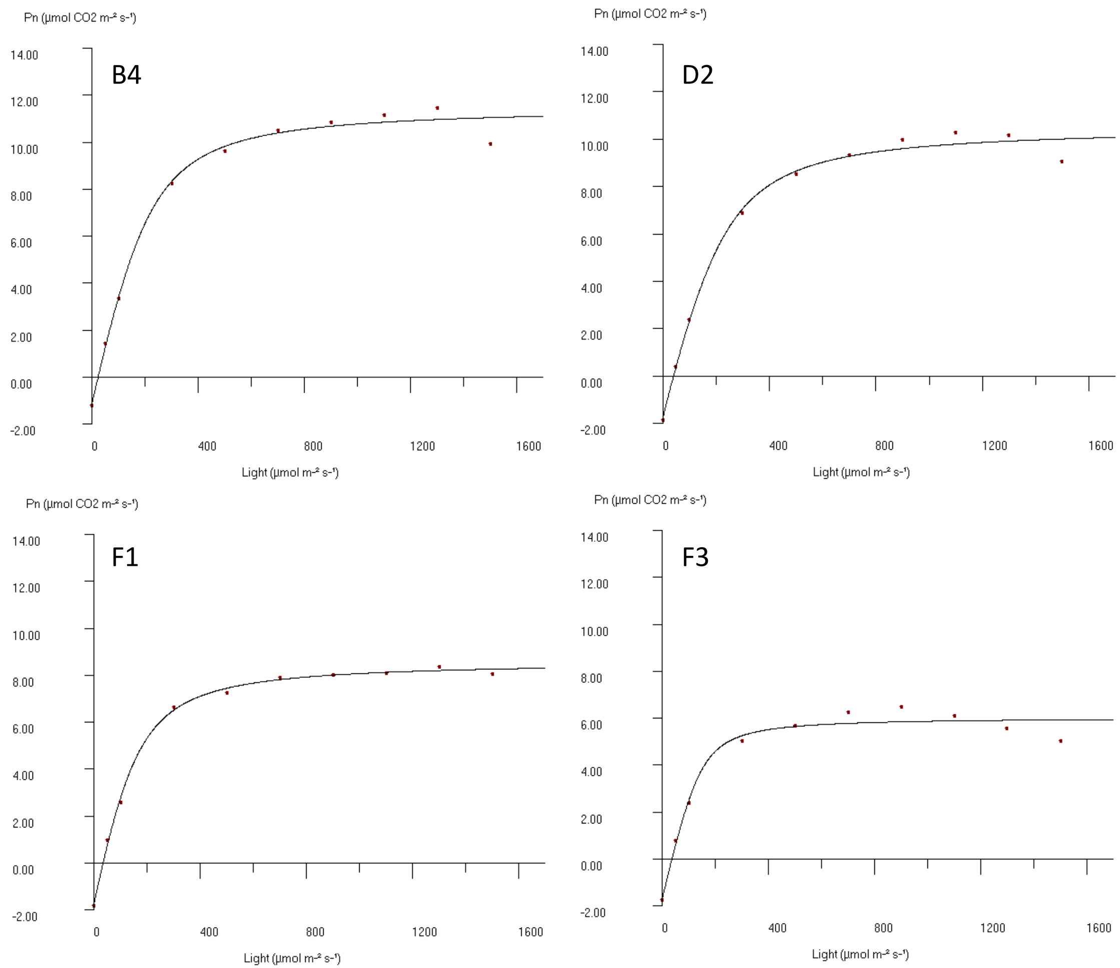

Fig. 3.4.1.2. Light response curves of photosynthesis for young leaves.

3.4.2 Morphology and Survival

All trees in these experiments were grown under the same environmental conditions and yet displayed a wide range of morphologies (Table. 3.4.2.1. & Table. 3.4.2.2.). Heights ranged from 10 mm to 810 mm, stem widths from 1.33 mm to 6.53 mm, 0 to six branches, very downy to slightly downy leaves, very downy to non-downy stems, fully expanded leaf widths from 16.73 mm to 80 mm and fully expanded leaf lengths

from 13.52 mm to 70 mm (Table. 3.4.2.1.).The majority of these characteristics are not linked to one another, for example more branched individuals are not more or less downy than unbranched individuals. The linked characteristics are that longer leaves are wider and conversely shorter leaves narrower, also that taller trees have wider main stems and shorter trees narrower stems. It is somewhat unsurprising that these characteristics are linked as overall leaf shape remains relatively consistent, so leaf lengths and widths would remain relatively within the same ratio, likewise taller trees would require thicker stems to support the extra weight (Table. 3.4.2.1.). This strongly indicates that morphology is not influenced solely by environment and a high degree of morphological difference must be due to the genotype of the individual, and that the characteristics mentioned can be inherited separately.

Acclimation made no difference to survival chances after the -10°C shock. Five of 15 individuals subjected to one week of acclimation at 3°C died after the -10°C shock and the same number died/survived after the shock having been subjected to two weeks of acclimation (Table. 3.4.2.1.). Interestingly, only two of 15 individuals died after a sudden freezing shock without a prior acclimation period.

There are two main potential reasons for ‘sudden-shock only’ individuals having greater survival chances. Firstly, it is possible that this observed difference is due to random chance; those individuals exposed to the sudden shock may have just happened to have a higher percentage of individuals with greater freezing tolerance. A second possibility is that of drought. Individuals undergoing acclimation received less water during that period since there was a lower level of plant water uptake in these colder conditions. Although never completely dry, this reduced water uptake

may have resulted in the death of certain individuals, since saplings are particularity sensitive to drought (Kullman, 1986).

Acclimation periods were at a temperature of 3°C. This is not particularly cold for Arctic species, but was the minimum temperature the growth chambers could maintain for long periods of time. It is plausible that this temperature may be insufficient to induce freezing acclimation in these saplings. Light intensity and period were also set to mimic spring/summer/early autumn, and the sudden drops in temperature to mimic the expected increase in cold/freezing snaps in response to global warming (Waldendorp et al., 2008, Bokhorst et al., 2009). Photoperiod is known to have a large effect on acclimation in these and other *Betula* species (Welling et al., 2004, Li et al., 2003, Rinne & Kaikuranta, 1998) and it therefore seems likely, given the results seen here, that photoperiod is the most important signal to induce acclimation and potentially cold signals alone are insufficient to induce freezing acclimation in *B. pubescens*.

Under the same conditions some individuals died whilst other did not, regardless of acclimation time. Very tiny fluctuations in conditions at the point of freezing (such as slightly damper soil causing greater root damage, degree of vermiculite root packaging that insulate the roots or proximity to chamber edges) could have made these individuals more susceptible to freezing damage (and ultimately death) than others. Alternatively, true genetic differences between individuals resulted in some individuals being more susceptible to freezing stress. Given the wide array of morphologies seen under the same environmental conditions (Table. 3.4.2.1.) and the occurrence of introgressive hybridisation between *B. pubescens* and *B. nana*

(Elkington, 1968; Anamthawat-Jónsson & Tomasson, 1990; Anamthawat-Jónsson et al., 2010; Thórsson, Salmela & Anamthawat-Jónsson, 2001 and Anamthawat-Jónsson et al., 2003) which have differences in freezing tolerance (Stushnoff & Junttila, 1986) it seems highly likely that genetic differences in ability to tolerate freezing conditions would occur and may explain the survival of some individuals over others. This requires further carefully designed experiments to elucidate fully.

Although many individuals survived (Table. 3.4.2.1.), all leaves on all individuals died after the freezing shock. Some leaves were retained, but these were very obviously dead. Since the leaves on all samples died after the freezing shock, although the majority of samples budburst much later, meant that a valuable period of potential photosynthesis was therefore lost. In the arctic the growing season is short so sudden cold/freezing snaps resulting in the loss of leaves, plus the potential for increased respiration of trees during warmer winters, could prove highly detrimental to birch forests, as carbohydrates are depleted and acquisition and storage of photosynthates is also reduced.

Despite the current expansion of the treeline in response to global warming (Truong, Palmé, and Felber, 2007), the data acquired in these experiments indicated that established forest patches may begin to retreat. However it is worth noting that, in the current investigation, these samples were grown in extremely favourable, wind free conditions and most had leaves much larger than their arctic counterparts (Table. 3.4.2.1.). It is possible that the batch of seed collected contained unusually large leaved individuals but this seems unlikely. The growing season is short in Abisko and many stresses such as cold, drought and herbivory are encountered by

the birch plants. It is more likely that a degree of environmental plasticity can occur, to take the maximum advantage of more favourable conditions, such as warmer temperatures, long daylight hours and sufficiency of water that they received in the growth chambers. Larger leaves can therefore be supported due to the favourable conditions to take maximum advantage of the available light, since these individuals have never encountered the risk of drought or loss of leaves to herbivory. These larger leaves and lack of previous exposure to cold may make the leaves and plants more susceptible to sudden freezing shocks, whereas individuals accustomed to cold and other stresses may be more constitutively prepared for freezing and other stresses rather than having to induce all freezing responses at the occurrence of the stress at the expense to leaf area and growth rate. An advantage to the cold/freezing snaps would be the reduction in invertebrate herbivory, which can completely defoliate birch woodlands. With milder winters survival chances of invertebrate herbivores will increase and could potentially lead to a greater number of complete defoliations which are evidentially highly detrimental to the trees. Sudden cold snaps would kill off the majority of these herbivores and given the indication that *Betula pubescens* uses temperature cues to bud burst (Welling et al., 2004) they may refoliate within the same season with a greatly reduced herbivore load and therefore be able to recoup some of the lost photosynthesis time resulting from herbivory leaf loss and the cold snap.

Given the extremely large degree of genetic variation and developmental plasticity possible in *B. pubescens* it would seem likely that although initial retreat of forest patches may occur when the net loss of energy of some individuals becomes greater than their gain, other individuals more tolerant of the new conditions may already

exist in the gene pool and be able to quickly recolonise and potentially forest patches will once again expand.

Why certain individuals do not die in response to freezing shock and others do is not conclusively answered in this study. In order to further investigate the cause of survival/death, further molecular and physiological studies are required. Further studies should investigate levels of gene expression to ascertain which genes are induced in response to cold and freezing, and if certain individuals constitutively have some genes active whilst others do not. Also, investigations as to whether there are particular expression patterns seen in different individuals should be undertaken. Sequencing of genes linked to cold/freezing acclimation/survival, such as CBF or genes found from expression studies, could indicate if specific polymorphisms in the gene correlate to higher levels of tolerance. These findings could then be correlated with genetic studies on established trees in the field. Seedlings could also be grown under an array of different conditions to see how morphology may be affected by different environmental factors, and these results compared to their genotype and gene expression levels to clarify the relationship between environmental plasticity and genetic variation.

Table. 3.4.2.1.a Table of treatment types, morphologies and survival of individual saplings. Individuals that were dead after the experiment are highlighted in red.

Test type	ID	Date tested	Height (mm)	Stem width (mm)	Number of Branches	Fully Expanded Leaf length (mm)	Fully Expanded Leaf Width (mm)	Stem type	Leaf type
1 week	Q5	17.8.12	20	1.83	0	17	19	Downy	Downy
1 week	L6	17.8.12	115	2.48	4	40	40	Downy	Downy
1 week	O12	17.8.12	130	2.5	2	30	35	Downy stripes	Downy
1 week	Q12	17.8.12	100	2.53	3	24	29	Downy stripes	Downy
1 week	E9	17.8.12	120	2.58	2	42	50	Downy stripes	Downy
1 week	A5	17.8.12	810	5.86	5	70	80	Downy stripes	Downy
1 week	L10	17.8.12	55	1.52	3	23.5	27	Downy	Slightly downy
1 week	A18	17.8.12	20	3.22	3	30	30	Downy	Slightly downy
1 week	K7	17.8.12	60	1.64	0	27	25	Downy stripes	Slightly downy
1 week	P15	17.8.12	65	2.01	1	30	35	Downy stripes	Slightly downy
1 week	H2	17.8.12	65	1.48	2	25	26	Slightly downy stripes	Slightly downy
1 week	N7	17.8.12	135	2.44	6	32.9	33	Slightly downy stripes	Slightly downy
1 week	A8b	17.8.12	535	4.55	3	70	75	Slightly downy stripes	Slightly downy
1 week	D3	17.8.12	110	4	2	31.5	39.58	Downy stripes	Very slightly downy
1 week	Q4	17.8.12	60	1.97	0	29	25.4	Downy stripes	Very slightly downy
2 weeks	O5	15.8.12	29	1.5	0	16.95	24.44	Downy	Downy
2 weeks	K20	15.8.12	80	2	3	34	30	Downy	Downy
2 weeks	P23	15.8.12	58	2.23	0	28.9	35.26	Downy	Downy
2 weeks	L5	15.8.12	165	2.98	4	55	62	Downy	Downy

Table. 3.4.2.1.b Table of treatment types, morphologies and survival of individual saplings. Individuals that were dead after the experiment are highlighted in red.

Test type	ID	Date tested	Height (mm)	Stem width (mm)	Number of Branches	Fully Expanded Leaf length (mm)	Fully Expanded Leaf Width (mm)	Stem type	Leaf type
2 weeks	L1	15.8.12	10	2.39	2	35	50	Slightly downy	Downy
2 weeks	B8	15.8.12	115	2.75	2	39	40	Slightly downy	Downy
2 weeks	K29	15.8.12	195	2.6	4	55	63	Downy stripes	Downy
2 weeks	K9	15.8.12	50	1.33	0	22.34	25.6	Slightly downy stripes	Downy
2 weeks	D17	15.8.12	110	2.59	3	45	42	Slightly downy stripes	Downy
2 weeks	P26	15.8.12	37	2	0	34	40	Downy	Slightly Downy
2 weeks	B4	15.8.12	775	6.53	0	70	68	Downy	Slightly Downy
2 weeks	K4	15.8.12	21	3.42	0	21.44	26.29	Downy stripes	Slightly Downy
2 weeks	O10	15.8.12	170	5.74	3	45	55	Downy stripes	Slightly Downy
2 weeks	P13	15.8.12	16	1.5	0	18.49	18.4	Very downy stripes	Slightly Downy
2 weeks	P3	15.8.12	34	1.37	1	19	25	Very slightly downy	Very slightly downy
Shock	Q28	22.8.12	35	1.66	0	25.06	28.7	Downy	Downy
Shock	Q10	22.8.12	55	1.97	3	22.21	27.27	Downy	Downy
Shock	K8	22.8.12	35	1.92	0	22.06	27.4	Downy stripes	Downy
Shock	N10	22.8.12	65	1.92	4	27	32	Downy stripes	Downy
Shock	N9	22.8.12	120	3.46	2	42.62	46	Downy stripes	Downy
Shock	R7	22.8.12	40	1.91	3	24.11	24.11	Slightly downy stripes	Downy
Shock	R6	22.8.12	44	2.39	1	41.54	48.8	Slightly downy stripes	Downy
Shock	O2	22.8.12	16	3.53	0	22.54	23.77	Downy	Slightly Downy

Table. 3.4.2.1.c Table of treatment types, morphologies and survival of individual saplings. Individuals that were dead after the experiment are highlighted in red.

Test type	ID	Date tested	Height (mm)	Stem width (mm)	Number of Branches	Fully Expanded Leaf length (mm)	Fully Expanded Leaf Width (mm)	Stem type	Leaf type
Shock	N15	22.8.12	65	1.83	2	23.13	26.42	Downy stripes	Slightly Downy
Shock	P22	22.8.12	45	2.26	0	25	25.37	Downy stripes	Slightly Downy
Shock	Q33	22.8.12	20	3.05	0	13.52	16.73	Not downy	Slightly Downy
Shock	Q14	22.8.12	40	1.72	3	25.69	31.38	Downy	Very Downy
Shock	L14	22.8.12	65	2.04	0	24.67	26.74	Downy stripes	Very Downy
Shock	D10	22.8.12	110	2.43	5	40.62	40.62	Very downy stripes	Very Downy
Shock	A15	22.8.12	330	4.7	2	66.47	69.37	Very downy stripes	Very Downy

Table. 3.4.2.2.a Table showing details of branches of individuals characterised in (Table. 3.4.2.1.). Dead individuals are highlighted in red.

Individual	Branch Number	Height up main stem (mm)	Length (mm)	Branches
A15	1	90	45	0
A15	2	115	39	0
A18	1	1	15	0
A18	2	1	20	0
A18	3	20	10	0
A5	1	10	670	0
A5	2	10	100	0
A5	3	15	90	0
A5	4	20	710	0
A5	5	20	300	0
A8b	1	190	135	0
A8b	2	220	160	0
A8b	3	250	130	0
B8	1	0	28	0
B8	2	7	4	0
D10	1	0	11	0
D10	2	4	10	0
D10	3	8	13	0
D10	4	8	8	0
D10	5	8	6	0
D17	1	7	20	0
D17	2	12	26	0
D17	3	12	27	0
D3	1	0	32	0
D3	2	15	7	0
E9	1	30	40	0
E9	2	20	35	0
H2	1	0	2	0
H2	2	0	7	0
K20	1	19	26	0
K20	2	19	4	0
K20	3	25	12	0
K29	1	0	8	0
K29	2	0	34	0
K29	3	0	24	0
K29	4	0	18	0
L1	1	6	21	0
L1	2	7	21	0

Table. 3.4.2.2.b Table showing details of branches of individuals characterised in (Table. 3.4.2.1.). Dead individuals are highlighted in red.

Individual	Branch Number	Height up main stem (mm)	Length (mm)	Branches
L10	1	11	6	0
L10	2	16	6	0
L10	3	16	8	0
L5	1	0	37	0
L5	2	0	30	0
L5	3	0	13	0
L5	4	0	45	0
L6	1	0	5	0
L6	2	5	7	0
L6	3	5	4	0
L6	4	5	18	0
N10	1	12	6	0
N10	2	12	7	0
N10	3	14	9	0
N10	4	14	13	0
N15	1	24	6	0
N15	2	22	3	0
N7	1	0	25	0
N7	2	10	35	0
N7	3	20	15	0
N7	4	20	30	0
N7	5	20	20	0
N7	6	20	25	0
N9	1	7	10	0
N9	2	15	8	0
O10	1	0	20	0
O10	2	0	23	0
O10	3	0	48	0
O12	1	1	17	0
O12	2	1	70	0
P15	1	13	10	0
P3	1	10	5	0
Q10	1	12	9	0
Q10	2	12	6	0
Q10	3	12	6	0
Q12	1	10	26	0
Q12	2	10	4	0
Q12	3	0	4	0

Table. 3.4.2.2.c Table showing details of branches of individuals characterised in (Table. 3.4.2.1.). Dead individuals are highlighted in red.

Individual	Branch Number	Height up main stem (mm)	Length (mm)	Branches
Q14	1	0	26	0
Q14	2	0	10	0
Q14	3	0	9	0
R6	1	0	7	0
R7	1	0	12	0
R7	2	0	7	0
R7	3	0	22	0

Chapter Four: Conclusion

4.1 Conclusions

4.1.1 Origins of forest patch structure

Selection within patches for particular genotypes at windward, central and leeward sites (which display different morphologies) was a rare occurrence. Any genotype selection observed proved inconsistent between sites from different patches. The majority of difference in genotype occurs between the different patches; as distance between patches increases the genotypes show greater differences between one another. There is evidence, however, that selection for particular versions of a gene can occur at different sites within a patch, as seen with the ADH allele. This would indicate that genetic selection can occur at various sites across a patch but it acts upon individual genes rather than the genotype as a whole. Since selection does not occur over the entire genotype and more polycormic individuals did not show a stronger relationship to *B. nana*, (which is highly polycormic), than to any other individual sampled, this may indicate that the patch morphologies seen are due to environmental factors such as mechanical damage or developmental plasticity rather than relatedness of individuals to *B. nana*. Chamber studies on seedlings grown under the same conditions did not support this hypothesis, as they showed a large array of morphologies, since these differences cannot be due to environmental variations they must have a genetic origin, indicating that difference in observed

morphologies across the patch may be due to selection for particular genes that influence the degree of polycormy.

There is some evidence from the growth chamber studies of *B. pubescens* that polycormic individuals have a slightly higher likelihood of greater freezing tolerance than monocormic individuals, which may be due to several factors. Genes other than polycormy gene(s) could influence the degree of polycormy as well as freezing tolerance such as CBF, overexpression of which in *A. thaliana* causes stunted growth (Welling & Palva, 2008) (Chapter. 4.4). However being polycormic did not ensure freezing survival in the growth chamber experiments nor did monocormy ensure freezing susceptibility. Alternatively the observed slightly greater probability of survival for polycormic individuals could be due to more recent hybridisations with *B. nana* therefore being more likely to inherit both the *B. nana* polycormy gene(s) and genes linked to greater freezing tolerance from *B. nana*. However as seen from genetic studies across the patch more polycormic individuals do not show greater relationship to *B. nana* than to non polycormic individuals, Indicating that genes affecting polycormy and genes inferring greater freezing tolerance do not need to be inherited together or be active concurrently. If cold tolerance genes do play a part in the observed morphologies, evidence from the growth chamber studies show that they are not the sole influencing factor. Therefore it seems likely that either polycormy may have several potential genetic origins other than just the developmental polycormy gene(s) or that the polycormic form and hence polycormy genes themselves are being selected for at these sites. Potential reasons for active selection of the various degrees of poly/monocormy include selection for increased polycormy at the windward site to reduce the amount of damage from iceblast, to

take advantage of the slightly increased temperature below the boundary layer and get maximum coverage of snow during winter despite the majority of snow being blown out from the windward edge further into the patch. Central sites which would have higher depths of snow and increased protection from iceblast damage due to protection by the other trees, and wind speeds dropping as the wind penetrates the patch, would not need to select for such a high degree of polycormy as at the windward sites. Leeward samples have the greatest protection from ice damage being protected by the other trees and experiencing the lowest wind speeds and hence would not need to select for polycormy. However the leeward sites are susceptible to higher levels of freezing with less snow build up indicating that the advantage of being taller and monocormic and outcompeting the neighbouring trees is greater than the disadvantage of reduced snow cover at the leeward site. Hence it seems likely that the main driving factor for the degree of polycormy is not linked to freezing tolerance but linked to the degree of potential iceblast damage. Hence the range of morphologies shown across the patch from shorter polycormic to taller monocormic individuals could be due to selection for a small number of genes influencing polycormy which would not been seen in the overall genotype studies.

Also evident from growth chamber experiments is the high plasticity of *B. pubescens,* with leaves growing to a much larger size in the chambers than seen in the wild. This is likely to be a plastic response to favourable conditions in the growth chambers. Although it is evident that there is a high genetic component to polycormy, it seems likely that if *B. pubescens* can plastically alter leaf size in response to environmental conditions it may also be capable of plastically altering the degree of polycormy in response to environmental conditions.

It seems likely that the overall forest patch structure is a combination of mechanical, genetic and plastic responses. Further experiments would be required to elucidate the degree of effect on morphology these potential influences have. Developmental plasticity tests would have to be undertaken in growth chambers to see how differing conditions affect development of seedlings. Mechanical damage tests could also be undertaken to see to what degree of frequent mechanical damage increased polycormy and if naturally more polycormic individuals had greater survival chances than monocormic individuals under frequent damage. Gene expression studies could be undertaken to identify which genes are involved in cold and freezing responses in these species and these genes then sequenced from samples across the patches to look at variance across the patch. Seedlings of different morphologies could also be compared to identify potential genes that may affect morphology and these genes then compared across the patches in Abisko.

4.1.2 Possible Consequences of Global Warming on Forest Patches

Physiological tests have shown that some saplings can survive sudden cold shocks in summer whilst others cannot; it is therefore possible that some individuals in established forests may be lost to increasing occurrences of sudden cold shocks in response to global warming (Waldendorp et al, 2008 and Bokhorst et al, 2009). The genetic differences between patches and sites within patches may result in separate patches and sites responding differently to the environmental changes. Although particular genotypes tend to remain in the same area there is evidence of more long distance dispersal of unrelated individuals into these patches. This combined with the

wide array of different genotypes, genes and potentially high range of plastic responses would indicate that there may already be individuals present in the gene pool capable of surviving the new environmental conditions. Established forest may initially retreat due to some established individuals being unable to tolerate sudden freezing shocks in summer, but different patches may respond differently given the differing range of genotypes seen across different patches. Overall, however, *B. pubescens* should be capable of responding relatively quickly to any negative effects arising from global warming, and quickly recolonize any lost areas and are likely to expand further with increasing temperature as already seen above the treeline (Truong, Palmé & Felber, 2007).

4.1.3 Other Findings of Particular Interest

An interesting discovery is the closer relationship of individuals of *B. nana* to individuals from neighbouring patches of *B. pubescens* than to other *B. nana* surrounding other patches. There is no evidence that *B. nana* has a genotype distinct from *B. pubescens*. This would indicate that introgressive hybridisation between these species is extremely high in Abisko to the extent that referring to *B. pubescens* and *B. nana* as separate species is not justified in this location and potentially other locations of co-occurrence. The separate naming is purely convenient for morphological descriptions and bears no relation to overall genotype with different forest patches having greater genetic difference between one another than *B. nana* does from *B. pubescens*.

Bibliography

ACIA. 2005. *Arctic Climate Impact Assessment*. Cambridge: Cambridge University Press.

Anamthawat-Jónsson, K., Thórsson, Æ. (2003). Natural Hybridisation in Birch: Triploid Hybrids Between *Betula nana* and *B. Pubescens*. *Plant Cell, Tissue and Organ Culture*. 85: 99-107.

Anamthawat-Jónsson, K., Thórsson, Æ., Temsch, E., Greilhuber, J. (2010). Icelandic Birch Polyploids – The Case of a Perfect Fit in Genome Size. *Journal of Botany*. doi:10.1155/2010/347254.

Anamthawat-Jónsson, K., Tomasson, T. (1990). Cytogenetics of Hybrid Introgression in Icelandic Birch. *Hereditas.* 112: 65-70.

Anamthawat-Jónsson, K., Tomasson, T. (1999). Brief Report: High Frequency of Triploid Birch Hybrid by *Betula nana* seed parent. *Hereditas*. 130: 191-193.

Anschlag, K., Broll, G., Holtmeier, F. (2008). Mountain Birch Seedlings in the Treeline Ecotone, Subarctic Finland: Variation in Above- and Below-Ground Growth Depending on Microtopography. *Arctic, Antarctic and Alpine Research.* 40: 609-616.

Arif, L., Bakir, M., Khan, H., Al Farhan, A., Al Homaidan, A., Bahkali, A., Al Sadoon, M., Shobrak, M. (2010). Application of RAPD for Molecular Characterization of Plant Species of Medicinal Value from an Arid Environment. *Genetics and Molecular Research.* 9: 2191-9198.

Bokhorst, S., Bjerke, J., Tommervik, H., Callaghan, T., Phoenix, G. (2009). Winter Warming Events Damage Sub-Arctic Vegetation: Consistent Evidence From an Experimental Manipulation and a Natural Event. *Journal of Ecology*. 97: 1408-1415.

Brown, I., Al-dawoody, S. (1979) Observations on Meiosis in Three Cytotypes of *Betula alba* L. *New Phytologist.* 83: 801-811.

Couch, J., Fritz, P. (1990) Isolation of DNA from Plant High in Polyphenolics. *Plant Molecular Biology Reporter.* 8: 8-12.

Dellaporta, S., Wood, J., Hicks, J. (1983) Isolation of DNA from Higher Plants. *Plant Molecular Biology Reporter*. 4: 19–21.

Dharmar, K., John De Britto, A. (2011) RAPD Analysis of Genetic Variability in Wild Populations of *Withania somnifera* (L.) Dunal. *International Journal of Biological Technology.* 2:21-25.

Dolferus, R., Jacobs, M., Peacock, W., Dennis, E. (1994). Differential Interations of Promoter Elements in Stress Responses of the Arabidopsis ADH Gene. *Plant Physiology.* 105:1075-1087.

Doylle, J., Doyle, J. (1987). A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochemical Bulletin*. 19:11-15.

Dye, D. (2002). Variability and Trends in the Annual Snow-cover Cycle in Northern Hemisphere Land Areas, 1972–2000. *Hydrological processes.* 16: 3065-3077.

Edwards K., Johnstone C., Thompson C. (1991). A Simple and Rapid Method for the Preparation of Plant Genomic DNA for PCR Analysis. *Nucleic Acids Research*. 19: 1349.

Ejankowski, W., Kunz, M. (2006). Reconstruction of Vegetation Dynamics in "Linje" Peat-Bog (N Poland) Using Remote Sensing Method. *Biodiversity Research and Conservation.* 1: 111-113.

Elkington, T. (1968). Introgressive Hybridization Between *Betula nana* L. and *B. Pubescens* Ehrh. In North-West Iceland. *New Phytologist*. 67: 109-118.

Freeling, A., Bennett, D. (1985). Maize Adh1. *Annual Review of Genetics.* 19:297-323.

Gallagher, F., Pechmann, I., Isaacson, B., Grabosky, J. (2011). Morphological Variation in the Seed of Gray Birch (*Betula populifolia*): The Effects of Soil-Metal Contamination. *Urban Habitats* 6.

Greene, A., Johnson, E. (1997). Secondary Dispersal of Tree Seeds on Snow. *Journal of Ecology*. 85:329-340.

Guillemaut, P., Mardchal-Drouard, L. (1992). Isolation of Plant DNA: A Fast, Inexpensive, and Reliable Method. *Plant Molecular Biology Reporter.* 10: 60-65.

Hall, T.A. (1999). BioEdit: a User-Friendly Biological Aequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41:95-98.

Hartley, S., Firn, R. (1989). Phenolic Biosynthesis, Leaf Damage, And Insect Herbivory in Birch (*Betula pendula*). *Journal of Chemical Ecology.* 15: 275-283.

Haukioja, E., Niemelä, P., Sirén, S. (1985). Foliage Phenols and Nitrogen in Relation to Growth, Insect Damage, and Ability to Recover After Defoliation, in the Mountain Birch *Betula pubescens* ssp *tortuosa*. *Oecologia (Berlin).* 65:214-222.

Hoeren, F., Dolferus, R., Wu, Y., Peacock, W., Dennis, E. (1998). Evidence for a Role for AtMYB2 in the Induction of the Arabidopsis Alcohol Dehydrogenase Gene (ADH1) by Low Oxygen. *Genetics*. 149: 479-490.

Hofgaard, A., Dalen, L., Håkan, H. (20090. Tree Recruitment Above the Treeline and Potential for Climate-driven Treeline Change. *Journal of Vegetation Science.* 20:1133-1144.

Horsfield, G., Thompson, D., (Eds) (1997). The Ecology and Restoration of Montane and Subalpine Scrub Habitats in Scotland. *Scottish Natural Heritage Review*. 83.

Howland, D., Oliver, R., Davy, A. (1995). Morphological and Molecular Variation in Natural Populations of Betula. *New phytologist.* 130:117-124.

Jarillo, J., Leyva, A., Salinas, J., Martinex-zapter, J. (1993). Low Temperature Induces the Accumulation of Alcohol Dehydrogenase mRNA in *Arabidopsis thaliana*, a Chilling-Tolerant Plant. *Plant Physiology.* 101: 833-837.

Järvinen, P., Palmé, A., Morales, L., Lännenpää, M., Keinänen, M., Sopanen, T., Lascoux, M. (2004). Phylogenetic Relationships of *Betula* Species (Betulaceae) Based on Nuclear ADH and Chloroplast MatK Sequences. *American Journal of Botany.* 91: 1834-1845.

Jonasson, S. (1981). Plant Communities and Species Distribution of Low Alpine *Betula nana* Heaths in Northernmost Sweden. *Vegetatio.* 44:51-64.

KAPA3G Plant PCR Kit, Technical Data Sheet, version 1.11

Karlsson, P., Weih, M. (1996). Relationships between Nitrogen Economy and Performance in the Mountain Birch *Betula pubescens* ssp. *tortuosa. Ecological Bulletins.* 45: 71-78.

Karlsson, P., Weih, M. (2001). Soil Temperatures Near the Distribution Limit of the Mountain Birch (*Betula pubescens* ssp. *czerepanovii*): Implications for Seedling Nitrogen Economy and Survival. *Arctic, Antarctic and Alpine Research*. 33: 88-92.

Kirkpatrick, A., Heal, K. (2001). Dwarf Birch (Betula nana) Performance in Relation to Soil Water and Soil Nutrients. *Scottish Natural Heritage Research Survey and Monitoring Report.* 170.

Kjällgren, L. and Kullman, L. (1998). Spatial Patterns and Structure of the Mountain Birch Tree-Limit in the Southern Swedish Scandes – A Regional Perspective. *Geografiska Annaler.* 80: 1–16.

Koehler, K., Center, A., Cavender-Bares, J. (2011). Evidence to a Freezing Tolerance- Growth Rate Trade-Off in the Live Oaks (*Quercus* series *Virentes*) Across the Tropical-Temperate Divide. *New Phytologist*. 193: 730-744

Kohler, J., Brandt, O., Johansson, M., Callaghan, T. (2006). A Longterm Arctic Snow Depth Record from Abisko, Northern Sweden 1913–2004. *Polar Research.* 25: 91–113.

Kullman, L. (1984). Transplantation Experiments with Saplings of *Betula pubescens* ssp. *tortuosa* near the Tree-Limit in Central Sweden. *Holarctic Ecology*. 7:289-293.

Kullman, L. (1986). Demography of *Betula pubescens* ssp. *tortuosa* Sown in Contrasting Habitats Close to the Birch Tree-Limit in Central Sweden. *Vegetatio.* 65: 13-20.

Kullman, L. (1989). Recent Retrogression of the Forest-Alpine Tundra Ecotone (*Betula pubescens Ehrh.* ssp.*tortuosa (Ledeb.) Nyman*) in the Scandes Mountains, Sweden. *Journal of Biogeography*. 16:83-90.

Kullman, L. (2001). 20th Century Climate Warming and Tree-Limit Rise in the Southern Scandes of Sweden. *Ambio.* 30: 77-80.

Kullman, L. (2002). Rapid Recent Range-Margin Rise of Tree and Shrub Species in the Swedish Scandes. *Journal of Ecology*. 90:68-77.

Lefort, F., Douglas, G. (1999). An Efficient Micro-method of DNA Isolation From Mature Leaves of Four Hardwood Tree Species *Acer*, *Fraxinus*, *Prunus* and *Quercus*. *Annals of Forest Science.* 56:259-263.

Li, C., Junttila, O., Ernstsen, A., Heino, P., Palva, E. (2003). Photoperiodic Control of Growth, Cold Acclimation and Dormancy in Development in Silver Birch (*Betula pendula*) Ecotypes. *Physiologia Plantarum.* 117:206-212.

Maltas, E., Vural, H., Yildiz, S. (2011). Extraction of Genomic DNA from Polysaccharide- and Phenolics-rich *Ginkgo biloba. Journal of Medicinal Plants Research.* 5: 332-339.

Milbau, A., Graae, B., Shevtsova, A., Nijs, I. (2009). Effects of a Warmer Climate on Seed Germination in the Subarctic. *Annals of Botany*. 104:287-296.

Maliouchenko, O., Palmé, A., Buonamici, A., Vendramin, G., Lascoux, M. (2007) Comparative Phylogeography and Population Structure of European *Betula* Species, With Particular Focus on *B. pendula* and *B. pubescens*. *Journal of Biogeography.* 34: 1601-1610.

Matlack, G. (1989) Secondary Dispersal of Seed Across Snow in *Betula* lenta A Gap-Colonizing Tree Species. *Journal of Ecology.* 77: 853-869.

MoBio PowerPlant® Pro DNA Isolation Kit Description (2010) available from http://www.mobio.com/plant-dna-isolation/powerplant-pro-dna-isolation-kit.html accessed 03/10/12.

MoBio PowerPlant® Pro DNA Isolation kit instruction Manual. version 01132012.

Nurmi, K., Ossipov, V, Haukioja, E., Pihlaja, K. (1996). Variation of Total Phenolic Content and Individual Low-Molecular-Weight Phenolics in Foliage of Mountain Birch Trees (*Betula pubescens* ssp. *Tortuosa*). *Journal of Chemical Ecology*. 22: 2023-2040.

Nuttal, M., Callaghan, T. eds. (2000). *The Arctic: Environment, People, Policy*. Amsterdam: Harwood Academic.

Oakenfull, R., Baxter, R., Knight, M. (2013). A C-Repeat Binding Factor Transcriptional Activator (CBF/DREB1) from European Bilberry (Vaccinium myrtillus) Induces Freezing Tolerance When Expressed in *Arabidopsis thaliana*. *PLoS ONE* 8(1): e54119. doi: 10.1371/journal.pone.0054119

Palme, A., Palsson, Q., Lascoux, M. (2004). Extensive Sharing of Chloroplast Haplotypes Among European Birches Indicates Hybridization Among *Betula pendula*, *B. pubescens* and *B. nana. Molecular Ecology.* 13:167-178.

Parmesan, C., Yohe, G. (2003). A Globally Coherent Fingerprint of Climate Change Impacts Across Natural Systems. *Nature*. 421: 37-42.

Paterson, A., Brubaker, C., Wendel, J. (1993). A Rapid Method for Extraction of Cotton (Gossypium spp.) Genomic DNA Suitable for RFLP or PCR Analysis. *Plant Molecular Biology Reporter.* 11(2):122-127.

Rinne, P., Kaikuranta, P. (1998). Onset of Freezing Tolerance in Birch (*Betula pubescens* Ehrh.) Invovles LEA Proteins and Osmoregulation and is Impaired in and ABA-deficient Genotype. Plant. *Cell and Environment*. 21: 601-611.

Rundqvist, S., Hedenås, H., Sandström. (2011). Tree and Shrub Expansion Over the Past 34 Years at the Tree-Line Near Abisko, Sweden. *Ambio*. 40: 683-692.

Sambrook, J., Russel, D. (2001). *Molecular cloning: A laboratory manual*. 3rd edition. New York: Cold Spring Harbor Laboratory Press.

Schenk, M., Thienpont, C., Koopman, W., Gilissen, L., Smulders, M. (2008). Phylogenetic Realtionships in *Betula* (Betulaceae) based on AFLP markers. *Tree genetics and Genomes.* 4:911-924.

Strange, C., Prehn, D., Arce-Johnson, P. (1998). Isolation of *Pinus* Radiate Genomic DNA Suitable for RAPD analysis. *Plant Molecular Biology Reporter*. 16:1-8.

Stevens, G., Fox, J. (1991). The Causes of Treeline. *Annual Review of Ecology and Systematics.* 22:177-191.

Stonehouse, B., 1989. *Polar Ecology*. Glasgow: Blackie.

Stushnoff, C., Junttila, O. (1986). Seasonal Development of Cold Stress Resistance in Several Plant Species at a Coastal and a Continental Location in North Norway. *Polar Biology*. 5:129-133.

Sveinbjörnsson, B., Kauhanen, H., Nordell, O. (1996). Treeline Ecology of Mountain Birch in the Torneträsk Area. *Ecological Bulletins.* 45: 65-70.

Sybenga, J. (1996). Chromosome Pairing Affinity and Quadrivalent Formation in Polyploids: do Segmental Allopolyploids Exist? *Genome.* 19: 1176-84.

Taylor, A., Clydesdale, F. (1987). Potential of Oxidised Phenolics as Food Colourants. *Food Chemistry.* 24:301-313

Thórsson, Æ. Salmela, E., Anamthawat-Jónsson, K. (2001). Morphological, Cytogenetic, and Molecular Evidence for Introgressive Hybridization in Birch. *The American Genetic Association.* 92:404-408.

Truong, C., Palmé, A., Felber, F. (2007). Recent Invasion of the Mountain Birch *Betula pubescens* ssp. *tortuosa* Above the Treeline Due to Climate Change: Genetic and Ecological Study in Northern Sweden. *Journal of Evolutionary Biology.* 20: 369-380.

Verwijst, T. (1988). Environmental Correlates of Multiple-Stem Formation in *Betula pubescens* ssp. *tortuosa*. *Vegetatio*. 76: 29-36.

Walters, S. (1968). *Betula* L. in Britain. *Proceedings of the Botanical Society of the British Isles.* 7: 179-180.

Weih, M., Karlsson, P. (1999). The Nitrogn Economy of Mountain Birch Seedlings: Implications for Winter Survival. *Journal of Ecology.* 87:211-219.

Welling, A., Palva, T. (2008). Involvement of CBF Transcription Factors in Winter Hardiness in Birch. *Plant Physiology.* 147: 1199-1211.

Welling, A., Rinne, P., Viherä-Aarnio, A., Kontunen-Soppela, S., Heino, P., Palva, E. (2004). Photoperiod and Temperature Differentially Regulate the Expression of Two Dehydrin Genes During Overwintering of Birch (*Betula pubescens* Ehrh.). *Journal of Experimental Botany.* 55: 507-516.

Woldendorp, G., Hill, M., Doran, R., Ball, M. (2008). Frost in a future Climate: Modelling Interactive Effects of Warmer Temperatures and Rising Atmospheric CO_2 on the Incidence and Severity of Frost Damage in a Temperate Evergreen (*Eucalyptus pauciflora*). *Global Change Biology.* 14: 294-308.

Xie, Y., Wu, R. (1989). Rice Alcohol Dehydrogenase Genes: Anaerobic Induction, Organ Specific Expression and Characterization of cDNA Clones. *Plant Molecular Biology*. 13:55-68.