# Identification of SNPs in candidate genes potentially involved in bud burst in European beech (*Fagus sylvatica* L.)

By M. Müller[1), 2), *)], S. Seifert[1),2)] and R. Finkeldey[1)]

(Received 16th October 2014)

**Abstract**

European beech (*Fagus sylvatica* L.) is one of the most important deciduous tree species in Central Europe. Higher annual mean temperatures caused by climate change lead to earlier bud burst in spring and/or a delay of leaf senescence in autumn. Since earlier bud burst might increase the late frost risk, adaptive traits like bud burst may gain more importance in the future. Nevertheless, knowledge of the genetic background of leaf unfolding is still scarce for European beech. In the present study, we analyzed parts of ten different candidate genes for bud burst with a total length of 12,290 bp. Comparative sequencing with plant material of 12 different beech populations distributed over Germany revealed 116 SNPs (single nucleotide polymorphisms) and 20 indels. A subset of 46 SNPs was successfully used for genotyping of 100 offspring of a beech population in Germany confirming the suitability of the newly developed SNP set for population genetic studies. The provided data may be useful for further investigations of adaptation in *F. sylvatica*.

*Keywords:* climate change, adaptation, SNPs, candidate genes, European beech, bud burst.

## Introduction

European beech (*Fagus sylvatica* L.) is a widespread and one of the most important deciduous tree species in Central Europe. The species is wind-pollinated, predominantly outcrossing, monoecious and has heavy fruits and therefore a limited seed dispersal. This dominant tree species is of ecological importance but also of great economic interest (e.g., Gömöry et al., 2003; Bolte et al., 2007). Global climate change may affect the vegetation period, growth, health and distribution of trees (European Environment Agency (EEA) 2012). Thereby, phenology is probably one of the most affected adaptive traits (Bertin, 2008). Higher temperatures lead to earlier bud burst and/or a delay of leaf senescence in autumn. As a consequence, the growing season was extended over the last decades as inferred by studies based on different tree and shrub species (Menzel and Fabian, 1999; Menzel, 2000), as well as on more than 650 temperate species including herbaceous plants (Khanduri et al., 2008). Nevertheless, bud burst timing is also influenced by photoperiod and winter chilling. Thereby, the importance of each these different factors for bud burst timing is thought to differ among species. Opportunistic pioneer species react mainly to temperatures whereas late successional species (such as beech) show a more complex response with a large chilling requirement and enhanced photoperiodic sensitivity (Körner and Basler, 2010; Caffarra and Donnelly, 2011; Basler and Körner, 2012). Therefore, beech and other photoperiod sensitive species might not be able to extend the growing season in the same amount as temperature sensitive species (Basler and Körner, 2012).

Paradoxically, warming may also increase the risk of plant frost damage, because warm temperatures might lead to a premature plant development which would result in the exposure of vulnerable plant tissues to late frosts (Gu et al., 2008). European beech is sensitive to late frost events after leaf flushing (Kreyling et al., 2012). Thereby, late frost damage can strongly affect beech vitality and competitiveness. The critical value for frost damage of flushing beech leaves is around –3 °C (Dittmar et al., 2006). Kreyling et al. (2012) emphasize that the frost sensitivity of beech strongly depends on timing, whereby the highest sensi-

[1)] Forest Genetics and Forest Tree Breeding, Büsgen Institute, Faculty for Forest Sciences and Forest Ecology, Georg-August University Göttingen. Büsgenweg 2, 37077 Göttingen, Germany.

[2)] These authors contributed equally to this work.

[*)] Corresponding author: Markus Müller:
Tel.: +49 551 39 14281, Fax: +49 551 39 8367.
E-Mail: mmuellef@gwdg.de



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

tivity can be found directly after leaf flushing and as soon as leaves mature the frost tolerance increases again. Beech provenances differ in bud burst date and also in their risk to late frost (WÜHLISCH et al., 1995; VIŠNJIĆ and DOHRENBUSCH, 2004). In general, provenances from the east and the south of Europe flush earlier than provenances from the western parts of Europe with very early flushing provenances at the eastern edge of the distribution (WÜHLISCH et al., 1995). The frost resistance mainly depends on the annual mean temperature of the populations. Furthermore, there is a positive correlation between flushing date and frost resistance (VIŠNJIĆ and DOHRENBUSCH, 2004).

Since tree species like European beech are main structural and functional elements of forest ecosystems, and the phenological changes described above may have substantial impacts on the vitality of tree populations, it is important to investigate the molecular basis of bud burst. At present, the candidate gene approach is a powerful and efficient method to study the genetic architecture of complex traits (ZHU and ZHAO, 2007). Here, bud burst related candidate genes were analyzed to identify variation within these genes, mainly SNPs (Single Nucleotide Polymorphisms). Parts of ten different candidate genes were analyzed and SNPs in coding, non-coding and untranslated regions were identified. In total, 56 SNPs were selected and their performance tested by genotyping 100 offspring of a natural beech population in Northern Germany. In total, 46 SNPs were successfully genotyped. This newly developed SNP set can be used in further studies to investigate the adaptation potential of European beech to climate change. To our knowledge there is only one other study (LALAGÜE et al., 2014) reporting SNPs in candidate genes potentially involved in bud burst for European beech.

## Material and Methods

### *Plant material*

The sampling design was developed to minimize ascertainment bias. An ascertainment bias occurs most probably if only a small set of individuals from only a part of the species' range is used for the SNP detection and later on, a large set of individuals is genotyped (e.g., HELYAR et al., 2011). Thus, the sampling areas of this investigation are located in five widely separated regions in Germany: in Schorfheide-Chorin in Northeastern Germany, in or near the Lüneburger Heide in Northern Germany, in the Hainich-Dün region in Central Germany, in the Harz Mountains in Central Germany and in the Schwäbische Alb in Southern Germany (*Table 1*). For the sampling sites of the Lüneburger Heide and the Harz Mountains, individuals out of seedling populations were used for comparative sequencing. The seedlings were raised from beechnuts, which were randomly collected in beech stands located in the sampling areas. A microsatellite analysis revealed that the seedling populations are representative for the adult stands of origin (data not shown). For the remaining sampling sites, adult individuals were used. In total, 24 trees from twelve different populations were included in comparative sequencing (two individuals per population). Annual mean temperatures of the investigated regions vary from around 7 °C to around 9 °C (*Table 1*). The performance of the selected SNPs were tested by genotyping 100 individuals of the seedling population Unterlüß sand.

### *DNA extraction*

Total DNA was extracted from leaves using the DNeasy™ PlantKit (Qiagen, Hilden, Germany). The amount and the quality of the DNA were analyzed using 1% agarose gel electrophoresis with 1 x TAE as running buffer (SAMBROOK et al., 1989). DNA was stained with ethidium bromide, visualized by UV illumination and compared to a Lambda DNA size marker (Roche, Mannheim, Germany).

### *Selection of candidate genes*

A literature search was conducted to select candidate genes which might be involved in bud burst behavior. The ten selected genes (*Table 2*) are expected to have an impact on bud burst in oak (DERORY et al., 2006; UENO et al., 2010), a genus of the Fagaceae family related to beech. The Evoltree EST database (http://www.evoltree.eu) and the EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/) were used to find corresponding *F. sylvatica* sequences. These were verified by a BLASTn and BLASTx search (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and used for primer design in order to amplify the corresponding genomic regions.



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table 1.* – Information about the sampling areas.

| Name of the region | Name of the population | Location of the population | Annual mean temperature [° C] | Altitude [m a.s.l.] |
|---|---|---|---|---|
| Schorfheide-Chorin | SEW6 | N 52° 54.447 E 13° 50.502 | 8-8.5 °C [a] | 55 [d] |
| | SEW46 | N 53° 4.327 E 13° 46.647 | 8-8.5 °C [a] | 80 [d] |
| Lüneburger Heide | Göhrde sand | N 53° 08.660 E 10° 52.003 | 8.7 [b] | 85 [b] |
| | Göhrde loam | N 53° 07.379 E 10° 49.224 | 8.7 [b] | 85 [b] |
| | Unterlüß sand | N 52° 49.831 E 10° 18.985 | 8.5 [b] | 117 [b] |
| | Calvörde sand | N 52° 22.819 E 11° 17.406 | 9.2 [b] | 75 [b] |
| | Calvörde loam | N 52° 24.238 E 11° 15.661 | 9.1 [b] | 72 [b] |
| Harz Mountains | Harz | N 51° 49.180 E 10° 15.213 | 7.2 [c] | 458 [c] |
| Hainich-Dün | HEW5 | N 51° 15.830 E 10° 14.457 | 6.5-8 °C [a] | 416 [d] |
| | HEW7 | N 51° 07.863 E 10° 23.126 | 6.5-8 °C [a] | 379 [d] |
| Schwäbische Alb | AEW8 | N 48° 22.953 E 9° 22.943 | 6-7 °C [a] | 766 [d] |
| | AEW40 | N 48° 29.976 E 9° 20.966 | 6-7 °C [a] | 779 [d] |

[a] annual mean temperature for the different regions (Fischer et al., 2010), but not especially for the single plots; [b] Hertel et al., 2013; [c] Seifert (2012); [d] Rajendra and Seifert et al. (2014).

*Amplification, cloning and sequencing of the candidate genes*

The software Primer3 version 0.4.0 (Rozen and Skaletsky, 2000) was used to design primers for amplification and direct sequencing of PCR products. Primers were checked for self-annealing, dimer and hairpin formations with the program OligoCalc version 3.26 (Kibbe, 2007).

PCRs were conducted in a 15 $\mu$l volume containing 2 $\mu$l of genomic DNA (about 10 ng), 1 x reaction buffer (0.8 M Tris-HCl pH 9.0, 0.2 M $(NH_4)_2SO_4$, 0.2% w/v Tween-20; Solis BioDyne, Tartu, Estonia), 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 1 unit of *Taq* DNA polymerase (HOT FIREPol® DNA Polymerase, Solis BioDyne, Tartu, Estonia), 0.3 $\mu$M of each forward and reverse primer (*Table 2*). For amplification of the gene fragment *PP2C* the HotStarTaq® MasterMix (Qiagen, Hilden, Germany) was used. The thermal cycling conditions were the following: an initial denaturation step of 95°C for 15 min followed by 35 cycles of 94°C for 1 min (denaturation), between 50°C and 68°C for 1 min (annealing, see *Table 2* for the different annealing temperatures), 72°C for 1 min (denaturation) and a final extension step of 72°C for 20 min. PCR products were analyzed using 1% agarose gel electrophoresis with 1x TAE as running buffer (Sambrook et al., 1989). DNA was stained with Roti®-Safe GelStain (Roth, Karlsruhe, Germany) and visualized by UV illumination. PCR products were excised from gel and purified using the innuPREP Gel Extraction Kit (Analytik Jena, Jena, Germany). The purified products were cloned into a pCR2.1 vector using the TOPO TA Cloning® Kit (Invitrogen, Carlsbad, USA) with slight modifications (supplementary material 1). Plasmid DNA was extracted using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, Steinheim, Germany). The sequencing reaction was carried out for three different clones per sample for each of the fragments by using the Big Dye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, USA) with both M13 forward and M13 reverse primers. Sequencing reactions were run on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

City, USA), and the sequenced fragments were verified by a BLASTn and BLASTx search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Protein domains within sequences were identified by searching against the NCBI Conserved Domain Database (CDD) (MARCHLER-BAUER et al., 2015)

*Table 2.* – Primer sequences and corresponding annealing temperatures for the selected candidate genes (candidate gene fragments longer than 1,000 basepairs were divided into two parts for sequencing. For the amplification of the product, the primers F part 1 and R part 2 were used. Accession No: EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl/)).

| Abbreviation | Gene | EMBL accession no. | Primer sequence (5'-3') | Annealing temperature |
|---|---|---|---|---|
| *Asi* | *Alpha amylase/ subtilisin inhibitor* | LK022686 | F: GTTGATGAGATCGATTGGAACCCTGAG<br>R: GCCAACGAGGGCAATTACAGAACTA | 68 °C |
| *Arf* | *Auxin response factor* | LK022685 | F part 1: AGTGATAGCAACTCCACAACCGTACC<br>R part 1: GAGTCTTAGGCTCTGAGATGCAAATG<br>F part 2: GTTGACCGGGAGAATGATGTGCTTC<br>R part 2: GTACTCAAGTGACCCCACAGACGTTA | 68 °C |
| *ConsC1* | *Constans like (1)* | LK022687 | F part 1: ACTCTCTTCTGCCGTGCCGACTCAG<br>R part 1: GTCGAGAGACGAAGAAGAAACCTG<br>F part 2: ACTCATCAGTGTCTCAGCCAGAGT<br>R part 2: GGCACGAGAGCTTCGCAGTAGTTAAT | 68 °C |
| *ConsC2* | *Constans like (2)* | LK022688 | F: ACTCTCACTACTCCCACACGTCTAC<br>R: GCTGTCAGTACCCGAACTGTGAAAC | 62 °C |
| *CP10* | *Chloroplast chaperonin like* | LK022689 | F part 1: GAGTAGGGAGTGGTCTGTCTCAGAGG<br>R part 1: TCAAGGGCTTGAGATCCTGT<br>F part 2: CTGGCACCCAAGTTGTGTATT<br>R part 2: ATCCACATGCCTTGAGGCACTTTCACC | 66 °C |
| *CysPro* | *Cystein Proteinase* | LK022694 | F: GACCATGAGTGTGATCCCGAGGAATA<br>R: CTGCATGGCATCAAGCTTCACTTACC | 60 °C |
| *DAG* | *DOF zinc finger protein* | LK022690 | F part 1: CCTTCTCCTTCTCCAACACACT<br>R part 1: TTCAAGTTCTAGACATTCTTTGTCG<br>F part 2: CCAGTCACTCCTCGGCTTAG<br>R part 2: GTACCGTGCGTGCCAAGTAT | 50 °C |
| *FRIGIDA* | *FRIGIDA* | LK022691 | F: GCGCGAGACTTAAAATCGAC<br>R: AAAAACCGTCCAATGCAATC | 50 °C |
| *His3C1* | *Histone 3* | LK022692 | F: GAAGCGAAAAGAGATGGCCCGTACGAA<br>R: GACAGCACAACACCAGTTTGAGATCC | 60 °C |
| *His3C2* | *Histone 3* | LK022693 | F: CTCTCAGAAAGTCCAGAACCCCAAAAGC<br>R: CGCTTAAGCACGTTCGCCACGGATCCTC | 67 °C |
| *NAC* | *NAC transcription factor* | LK022695 | F part 1: TTGTAGCCGGAAATGGGTGT<br>R part 1: GACACGTGGCAAAGTGAAGA<br>F part 2: TTGGGTTTTGTGTCGGATTT<br>R part 2: CCCTTTTGGTGCTAAACTCCAG | 62 °C |
| *PP2C* | *Protein phosphatase 2C* | LK022696 | F part 1: GGGATTTGCTGTGGAGTTGT<br>R part 1:TCTGCAATTGGTGGTTTTGA<br>F part 2: GAAAGAAGAGGTGGAAAGCGTA<br>R part 2: CGTTGTCCGTACTGTGCCTA | 50 °C |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

using the Conserved Domain Search (CD-Search) tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (MARCHLER-BAUER and BRYANT, 2004).

*SNP analyses*

SNPs occurring in only one individual were excluded after comparative sequencing to avoid the analysis of false SNPs because of sequencing errors. To reduce the number of SNPs for genotyping, haplotype tagging SNPs (htSNPs) were identified by using the software HaploBlockFinder version 0.7 (ZHANG and JIN, 2003). Out of these, SNPs were selected for genotyping based on their type (non-synonymous SNPs were preferred), and on their suitability for genotyping (e.g., no indels in vicinity of the target SNP). In total, 56 SNPs (18 non-coding SNPs, 17 synonymous SNPs, 16 non-synonymous SNPs and five SNPs from untranslated regions (UTR)) were chosen for genotyping (supplementary material Table S1). Surrounding sequences of the selected SNPs were sent to KBiosciences UK Ltd for primer design and analysis of the SNPs using the PCR-based KASP™ genotyping assay (Hoddesdon, UK).

*Data analyses*

Analyses of sequence data

For visual examination, editing and alignments of the sequences, CodonCode Aligner version 4.2.4 (CodonCode cooperation, http://www.codoncode.com) and BioEdit version 7.1.3.0 (HALL, 1999) using ClustalW multiple alignment (THOMPSON et al., 1994) were applied. Nucleotide diversity ($\pi$), haplotype diversity and Tajimas' *D* (TAJIMA, 1989) were calculated excluding indels with the software DnaSP version 5.10.01 (LIBRADO and ROZAS, 2009). The same software was used to infer linkage disequilibrium (LD) between SNPs within each candidate gene. $R^2$ values (HILL and ROBERTSON, 1968) were calculated and their significance tested using the Fisher's exact test. Bonferroni correction was used to correct for multiple testing. The LD decay with distance was estimated by regression analysis also implemented in the DnaSP software. Additionally, LD decay for all candidate genes combined was determined by plotting all $r^2$ values against distance (bp) and applying a linear regression analysis using the software STATISTICA version 12 (StatSoft Inc., Tulsa, US).

*Table 3.* – Overview of exons, introns, UTR, indels and SNPs of the analyzed gene fragments.

| Gene name | Abbreviation | Total length (bp) | Length (bp) of exons | Length (bp) of introns | Length (bp) of UTR | No. of indels | Total no. of SNPs | No. of non-coding SNPs | No. of SNPs in UTR | No. of synonymous SNPs | No. of non synonymous SNPs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Auxin response factor* | *Arf* | 1058 | 512 | 546 | 0 | 0 | 13 | 13 | 0 | 0 | 0 |
| *Alpha amylase/subtilisin inhibitor* | *Asi* | 873 | 631 | 0 | 242 | 1 | 10 | 0 | 4 | 2 | 4 |
| *Constans like* | *ConsC1* | 1200 | 935 | 106 | 159 | 0 | 3 | 0 | 0 | 1 | 2 |
| | *ConsC2* | 583 | 551 | 0 | 32 | 1 | 8 | 0 | 1 | 2 | 5 |
| *Chloroplast chaperonin like* | *CP10* | 1594 | 633 | 917 | 44 | 2 | 19 | 11 | 0 | 6 | 2 |
| *Cysteine proteinase* | *CysPro* | 920 | 496 | 212 | 212 | 1 | 12 | 3 | 7 | 2 | 0 |
| *Dof zinc finger protein* | *DAG* | 1210 | 459 | 627 | 124 | 5 | 15 | 11 | 1 | 3 | 0 |
| *FRIGIDA* | *FRIGIDA* | 430 | 430 | 0 | 0 | 1 | 7 | 0 | 0 | 5 | 2 |
| *Histone 3* | *His3C1* | 939 | 394 | 292 | 253 | 2 | 5 | 2 | 1 | 2 | 0 |
| | *His3C2* | 716 | 386 | 292 | 38 | 2 | 11 | 7 | 2 | 2 | 0 |
| *NAC transcription factor* | *NAC* | 1357 | 1030 | 188 | 139 | 5 | 6 | 1 | 2 | 2 | 1 |
| *Protein phosphatase 2C* | *PP2C* | 1410 | 1129 | 281 | 0 | 0 | 7 | 2 | 0 | 2 | 3 |
| Total | | 12,290 | 7,586 | 3,461 | 1,243 | 20 | 116 | 50 | 18 | 29 | 19 |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

Analyses of genotypic data of the population US

Linkage disequilibrium and deviations from Hardy-Weinberg proportions were estimated using the software Genepop version 4.2.1 (ROUSSET, 2008). Thereby, 10,000 demorization steps, 100 batches and 5,000 iterations per batch were used as Markov chain parameters. Neutrality was tested with the Ewens-Watterson test (MANLY, 1985) implemented in the software POPGENE version 1.32 (YEH et al., 1999) using 1000 simulations. The observed heterozygosity ($H_o$) and the expected heterozygosity ($H_e$) were calculated using the software GenAlEx version 6.5 (PEAKALL and SMOUSE, 2006, 2012).

## Results

### *Sequence data*

Fragments of ten different candidate genes for bud burst with a total length of 12,290 bp were analyzed in this study (*Table 3*). From the total length of the analyzed fragments, 7,586 bp accounted for exons, 3,461 bp for introns and 1,243 bp for UTR (for more detailed information about the gene structure compare the entries in the EMBL Nucleotide Sequence Database (*Table 2*)). Protein domains were identified in all candidate gene sequences (supplementary material *Table S1*). The length of the analyzed gene fragments varied from 430 bp for the gene *FRIGIDA* to 1,410 bp for the gene *PP2C*. In total, 20 indels and 116 SNPs were identified (on average one SNP per 106 bp) (supplementary material *Table S2*). In total, 19 SNPs were found to be non-synonymous, whereas all of them led to an amino acid exchange and none caused an early stop codon. The number of haplotypes ranged from three (gene fragment *Asi*) to 15 (gene fragment *CP10*), whereas the mean value was 8.42 (*Table 4*). The mean haplotype diversity was 0.652 ranging from 0.302 (gene fragment *FRIGIDA*) to 0.848 (gene fragment *PP2C*). The nucleotide diversity ranged from 0.57 (gene fragment *ConsC1*) to 4.58 (gene fragment *Arf*). The mean nucleotide diversity over all fragments was higher for non-coding sites than for coding sites (*Table 4*). Tajima's *D* was significantly positive for the gene fragment *His3C1* ($p < 0.05$). The mean linkage disequilibrium within genes was 0.32 (measured by $r^2$). After applying Fisher's exact test 48% of the pairwise comparisons remained significant ($p < 0.05$), and the mean LD for these SNP pairs was 0.64. After the Bonferroni correction 37% of the pairwise comparisons remained significant, and the mean LD was 0.79. The LD decay with distance was different for the analyzed candidate genes (supplementary material *Figure S1*). For four genes (*Arf*, *Asi*, *ConsC2*, *CysPro*) no LD decay over the analyzed distances was detected.

*Table 4.* – Nucleotide Diversity ($\pi$), haplotype diversity and results of Tajima's *D* test of the different gene fragments.

| Name of the Gene fragment | No of haplotypes | Haplotype diversity | Total Nucleotide diversity[+] | Nucleotide diversity[+] | | | | | Tajima's *D* |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Non-coding sites (introns) | Coding sites | UTR | Syn. sites | Non-syn. sites | |
| *Arf* | 12 | 0.805 | 4.58 | 8.87 | 0 | - | 0 | 0 | 1.974 |
| *Asi* | 3 | 0.377 | 3.37 | | 2.60 | 5.58 | 2.64 | 2.59 | 0.804 |
| *ConsC1* | 4 | 0.568 | 0.57 | 0 | 0.73 | 0 | 0.53 | 0.79 | 0.012 |
| *ConsC2* | 8 | 0.807 | 4.23 | - | 3.82 | 11.64 | 3.95 | 3.78 | 1.002 |
| *CP10* | 15 | 0.841 | 3.06 | 3.38 | 2.42 | 0 | 8.51 | 0.40 | 0.110 |
| *CysPro* | 5 | 0.621 | 3.03 | 3.12 | 1.13 | 7.44 | 4.75 | 0 | 0.094 |
| *DAG* | 10 | 0.704 | 3.28 | 4.89 | 1.97 | 2.51 | 8.58 | 0 | 0.492 |
| *FRIGIDA* | 4 | 0.302 | 2.14 | - | 2.14 | - | 6.07 | 0.96 | -1.131 |
| *His3C1* | 5 | 0.566 | 2.25 | 3.51 | 2.57 | 0.32 | 10.32 | 0 | 2.124* |
| *His3C2* | 13 | 0.735 | 2.86 | 5.35 | 0.90 | 9.12 | 3.68 | 0 | -0.517 |
| *NAC* | 9 | 0.645 | 1.44 | 2.37 | 1.10 | 2.85 | 2.82 | 0.60 | 1.119 |
| *PP2C* | 13 | 0.848 | 1.57 | 2.13 | 1.43 | - | 3.41 | 0.82 | 1.067 |
| Mean | 8.42 | 0.652 | 2.70 | 4.20 | 1.89 | 5.64 | 5.02 | 1.24 | 0.596 |

[+] $\pi \times 10^{-3}$; * $p < 0.05$.



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table 5.* – Overview of successfully genotyped SNPs. $H_o$: observed heterozygosity, $H_e$: expected heterozygosity, HWE: Hardy-Weinberg proportions.

| SNP no. | SNP name | Position (bp) | Gene | Characteristic | Substitution | $H_o$ | $H_e$ | P HWE |
|---|---|---|---|---|---|---|---|---|
| 1 | Arf_265 | 239 | *Auxin response factor* | non-coding | A/G | 0.490 | 0.465 | 0.668 |
| 2 | Arf_303 | 277 | | non-coding | A/G | 0.480 | 0.403 | 0.081 |
| 3 | Arf_563 | 537 | | non-coding | A/G | 0.150 | 0.172 | 0.205 |
| 4 | Arf_573 | 547 | | non-coding | T/C | 0.500 | 0.466 | 0.530 |
| 7 | Arf_615 | 589 | | non-coding | A/G | 0.560 | 0.495 | 0.230 |
| 12 | Arf_833 | 807 | | non-coding | A/G | 0.323 | 0.284 | 0.290 |
| 13 | Arf_878 | 852 | | non-coding | A/G | 0.480 | 0.403 | 0.083 |
| 25 | ConsC1_293 | 268 | *Constans like* (1) | non-synonymous | A/T | 0.495 | 0.460 | 0.519 |
| 26 | ConsC1_306 | 281 | | synonymous | A/G | 0.050 | 0.049 | 1.000 |
| 29 | ConsC2_51 | 26 | *Constans like* (2) | UTR | A/G | 0.460 | 0.394 | 0.128 |
| 30 | ConsC2_98 | 73 | | non-synonymous | C/G | 0.020 | 0.020 | 1.000 |
| 31 | ConsC2_147 | 122 | | synonymous | T/G | 0.140 | 0.147 | 0.480 |
| 32 | ConsC2_151 | 126 | | non-synonymous | C/G | 0.480 | 0.461 | 0.828 |
| 33 | ConsC2_211 | 186 | | non-synonymous | T/G | 0.090 | 0.086 | 1.000 |
| 34 | ConsC2_390 | 365 | | synonymous | T/C | 0.090 | 0.086 | 1.000 |
| 36 | ConsC2_488 | 463 | | non-synonymous | T/C | 0.430 | 0.498 | 0.163 |
| 38 | CP10_65 | 39 | *Chloroplast Chaperonin like* | synonymous | T/C | 0.250 | 0.262 | 0.700 |
| 39 | CP10_67 | 41 | | non-synonymous | T/C | 0.080 | 0.077 | 1.000 |
| 45 | CP10_377 | 351 | | non-coding | T/G | 0.429 | 0.408 | 0.805 |
| 47 | CP10_442 | 416 | | non-coding | C/G | 0.247 | 0.232 | 1.000 |
| 48 | CP10_503 | 477 | | synonymous | C/G | 0.130 | 0.156 | 0.138 |
| 50 | CP10_749 | 723 | | synonymous | C/G | 0.260 | 0.255 | 1.000 |
| 55 | CP10_1317 | 1291 | | non-coding | A/G | 0.210 | 0.219 | 0.644 |
| 56 | CP10_1428 | 1402 | | non-synonymous | T/C | 0.253 | 0.236 | 0.689 |
| 58 | CysPro_118 | 118 | *Cystein proteinase* | synonymous | C/G | 0.510 | 0.486 | 0.685 |
| 59 | CysPro_202 | 202 | | synonymous | A/G | 0.040 | 0.040 | 1.000 |
| 63 | CysPro_728 | 728 | | UTR | C/G | 0.080 | 0.077 | 1.000 |
| 65 | CysPro_783 | 783 | | UTR | T/G | 0.515 | 0.487 | 0.683 |
| 71 | DAG_81 | 58 | *Dof zinc finger protein* | UTR | A/G | 0.380 | 0.385 | 1.000 |
| 72 | DAG_289 | 266 | | non-coding | A/T | 0.380 | 0.385 | 1.000 |
| 89 | DAG_1059 | 1036 | | synonymous | T/G | 0.270 | 0.262 | 1.000 |
| 91 | Frigida_54 | 34 | *FRIGIDA* | synonymous | T/C | 0.030 | 0.030 | 1.000 |
| 92 | Frigida_104 | 84 | | non-synonymous | A/G | 0.040 | 0.039 | 1.000 |
| 93 | Frigida_179 | 159 | | non-synonymous | A/G | 0.080 | 0.077 | 1.000 |
| 101 | His3C1_292 | 292 | *Histone 3* (1) | non-coding | T/C | 0.540 | 0.497 | 0.428 |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table 5 (continued).* – Overview of successfully genotyped SNPs. $H_o$: observed heterozygosity, $H_e$: expected heterozygosity, HWE: Hardy-Weinberg proportions.

| SNP no. | SNP name | Position (bp) | Gene | Characteristic | Substitution | $H_o$ | $H_e$ | P HWE |
|---|---|---|---|---|---|---|---|---|
| 108 | His3C2_104 | 104 | *Histone 3* (2) | synonymous | A/C | 0.010 | 0.010 | - |
| 110 | His3C2_186 | 186 | | non-coding | T/C | 0.390 | 0.416 | 0.63 |
| 112 | His3C2_260 | 260 | | synonymous | A/G | 0.350 | 0.326 | 0.757 |
| 123 | NAC_854 | 834 | *NAC transcription factor* | non-synonymous | A/C | 0.280 | 0.332 | 0.13 |
| 124 | NAC_962 | 942 | | synonymous | A/G | 0.140 | 0.147 | 0.479 |
| 129 | NAC_1300 | 1280 | | UTR | A/G | 0.350 | 0.390 | 0.312 |
| 131 | PP2C_315 | 315 | *Protein phosphatase 2C* | non-synonymous | C/G | 0.130 | 0.122 | 1.000 |
| 132 | PP2C_391 | 391 | | synonymous | T/G | 0.460 | 0.495 | 0.548 |
| 134 | PP2C_791 | 791 | | non-synonymous | A/G | 0.020 | 0.020 | 1.000 |
| 135 | PP2C_941 | 941 | | non-coding | T/G | 0.480 | 0.498 | 0.691 |
| 136 | PP2C_1200 | 1200 | | synonymous | A/G | 0.469 | 0.498 | 0.547 |

*Genotypic data of the population US*

In total, 56 SNPs were chosen for genotyping. Eight SNPs were not processed successfully, and two SNPs turned out to be monomorphic. Thus, 46 SNPs were successfully genotyped (15 synonymous SNPs, 14 non-coding SNPs, 12 non-synonymous SNPs and five SNPs from UTR) (*Table 5*). Significant LD was detected for different SNP pairs. In total, 7.8% of all possible SNP pairs were found to be significantly in LD ($p<0.05$) (supplementary material *Table S3*). No significant deviations from Hardy-Weinberg proportions were detected, whereas the locus *His3C2_104* was not polymorphic enough to calculate probabilities for Hardy-Weinberg proportions. No departures from neutral expectations were detected by the Ewens-Watterson test. The observed heterozygosity ($H_o$) ranged from 0.01 to 0.560, whereas the mean value was 0.284. The expected heterozygosity ($H_e$) ranged from 0.01 to 0.498 (mean 0.277) (*Table 5*).

## Discussion

In this study, parts of ten different candidate genes for bud burst of the non-model species *Fagus sylvatica* L. were analyzed. In total, 116 SNPs were identified in the course of comparative sequencing which is, on average, one SNP per 106 bp. This SNP density lies almost exactly in between the values revealed by the two other studies, which analyzed candidate genes in beech. Lalagüe et al. (2014) detected one SNP per 89 bp, whereas Seifert et al. (2012b) detected one SNP per 129 bp. Nevertheless, SNP density strongly depends on the investigated region. More SNPs were identified in non-coding regions than in coding regions, which is in line with the results of other studies (e.g., Hemmer-Hansen et al., 2011; Vollmer and Rosel, 2012; Seifert et al., 2012b).

Three gene fragments (*Arf*, *His3C1*, *PP2C*) analyzed in the present study were partially overlapping with sequences previously reported by Lalagüe et al. (2014). This facilitated a comparison of identified variation within these genes between two independent studies. Largely, identical SNPs were detected in the overlapping gene fragments. Nevertheless, some SNPs were exclusively identified either in one or the other study. Since populations from distinct regions of the distribution area of beech were investigated (Germany and France), some differences in the SNP pattern can be expected. Additionally, several SNPs reported by Lalagüe et al. (2014) have previously been removed as single SNPs in the present study, and thus, turned out to be false negative SNPs. Nevertheless, the partially different SNPs detected in the overlapping gene fragments of the two studies shows again the importance of a carefully



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

chosen ascertainment set to avoid ascertainment bias. Ascertainment bias is the systematic deviation from the expected allele frequency distribution resulting from sampling processes used to find marker loci (HELYAR et al., 2011). Typically SNPs are identified in a small panel of individuals from a part of the species' range. In this case, SNPs with low allele frequencies might not be detected. If a large set of individuals is genotyped with these SNPs, an ascertainment bias can occur affecting any statistical measure that relies on allele frequency (NIELSEN, 2000; NIELSEN et al., 2004; HELYAR et al., 2011). To avoid ascertainment bias, a relatively large sample of individuals for SNP detection should be chosen, which represents all populations included in the final genotyping (MORIN et al., 2004). For that reason, comparative sequencing was carried out with individuals from sampling areas located in five widely separated regions in Germany.

The mean nucleotide diversity ($\pi$) was 0.0027 in the present study. The estimates of $\pi$ might be comparatively conservative, since all SNPs occurring only in one individual were excluded from analysis, and only a limited number of trees and clones per tree (*Escherichia coli* transformants) were used for comparative sequencing. Additionally, nucleotide diversity strongly depends on the investigated genes ranging from 0.00057 to 0.00458 in this study, and thus, a comparison between different studies may be complicated. Nevertheless, both additionally existing studies of nucleotide diversity in European beech reported very similar mean values of $\pi$ (SEIFERT et al., 2012b $\pi$: 0.0026; LALAGÜE et al., 2014 $\pi$: 0.0022) though mainly different gene fragments and/or candidate genes were analyzed. Thus, nucleotide diversity might generally be lower compared to other tree species, since several studies reported higher mean nucleotide diversities based on gene sequences for other Fagaceae (e.g., VORNAM et al., 2007 (*Quercus petraea*) $\pi$: 0.00542; QUANG et al., 2008 (*Quercus crispula*) $\pi$: 0.00693; DERORY et al., 2010 (*Quercus petraea*), and other tree species (e.g., INGVARSSON, 2005 (*Populus tremula*) $\pi$: 0.0111; KRUTOVSKY and NEALE, 2005 (*Pseudotsuga menziesii*) $\pi$: 0.00655). The different levels of nucleotide diversity can be caused by a combination of different factors. For instance, selection, parts of the genome considered, sampling strategies, demographic history, and differences in mutation rates are known factors causing different levels of nucleotide diversity among species (HEUERTZ et al., 2006).

Tajima's *D* test was applied to test the sequence data for selective neutrality. The test was statistically significant for only one gene (*His3C1*). The positive value of Tajima's *D* obtained for that gene indicates balancing selection, but the parameter is known to be highly sensitive to sample size (LARSSON et al., 2013). Since the estimations of Tajima's *D* were based on a low number of individuals in this study, the results should be interpreted cautiously. Further, this parameter is known to be affected by population structure. Studies including populations of the present study detected low but significant differentiation among the populations (RAJENDRA and SEIFERT et al., 2014; CARSJENS et al., 2014). Hence, the influence of population structure on Tajima's *D* cannot completely be ruled out.

Relatively high LD levels were found for the different genes (mean $r^2$ of 0.32; 0.64 after Fisher's exact test, and 0.79 after Bonferroni correction). The LD decayed to lower levels ($r^2 < 0.1$) for most genes within a distance of ca. 1,200 bp, albeit no LD decay was detected for four genes. The relatively high levels of LD combined with a slow LD decay in beech compared to other outcrossing tree species (e.g., INGVARSSON, 2005; HEUERTZ et al., 2006) are in line with the results of LALAGÜE et al. (2014). These authors explained the observed LD patterns with a small effective population size of the investigated population. In general, LD is a result of the interplay of several factors, such as mating system, recombination and mutation rates, selection, population size, structure and population history (KRUTOVSKY and NEALE, 2005). Nevertheless, the estimation of LD decay depends on the sequence length and the level of polymorphism (LALAGÜE et al., 2014). In the present study, the sequence length of three genes was shorter than 800 bp, and the mean distance between SNPs for all pairwise comparisons was 329 bp. Additionally, in six genes less than 10 SNPs were detected. Thus, for a more precise estimation of LD pattern, longer sequences may be investigated.

In total, 46 SNPs were successfully genotyped in a seedling population. In total, 7.8% of all possible SNP pairs were found to be significantly in LD. This percentage is lower compared to other studies. Thus, VIDALIS et al. (2013) found that 39.7% of all possible SNP



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

pairs were in LD investigating different *Quercus* species, and INGVARSSON et al. (2008) found 12.8% of all SNP pairs to be in LD in a study with *Populus tremula*. No significant deviations from Hardy-Weinberg proportions were detected, whereas one locus (*His3C2_104*) was not polymorphic enough to calculate probabilities for Hardy-Weinberg proportions in this population. Additionally, no departures from neutral expectations were detected with the Ewens-Watterson test. The observed ($H_o$) and expected ($H_e$) heterozygosities were similar (mean $H_o$: 0.284; mean $H_e$: 0.277), though strongly different among single SNP markers. The values are slightly lower compared to the results of SEIFERT et al. (2012a), who reported mean observed and expected heterozygosities of 0.326 and 0.324, respectively for the source stand of the seedling population investigated in the present study. Since a microsatellite analysis revealed no significant differences between adult and seedling populations (data not shown), these differences are most likely due to the different SNP markers applied in the two studies.

The selected candidate genes in the present study are putatively involved in flowering, temperature response and stress response. These functions have been associated with bud burst before. Thus, several stress related genes were expressed during bud burst in Norway spruce, suggesting that trees need to protect themselves from unfavorable abiotic factors during bud development (YAKOVLEV et al., 2006). In the same study, genes associated with temperature were expressed which can be expected, since temperature plays an important role in spring phenology. The *CONSTANS* gene has an important role in the regulation of flowering by photoperiod in *Arabidopsis* (GRIFFITHS et al., 2003). This gene is relevant for bud burst analysis, since it is suggested that the pathway regulating bud development is common to vegetative and sexual buds (HORVATH, 2009; ALBERTO et al., 2013). The SNP set developed in the present study can be used in further investigations, especially in genetic association studies. These studies attempt to identify patterns of polymorphisms that vary systematically between individuals with different phenotypes (BALDING, 2006). For instance, SNPs from a *Constans-like* gene were associated with bud burst in oak (ALBERTO et al., 2013), and LIND-RIEHL et al. (2014) found evidence for selection on a *Constans-like* gene between two red oak species. Hence, the SNPs provided in the present study are promising for the use in association studies in European beech.

## Acknowledgements

We thank A. DOLYNSKA, C. RADLER, G. DINKEL and A. CAPELLE for their technical assistance as well as all persons who helped us with the field work.

We thank H. LALAGÜE, K. CSILLÉRY, S. ODDOU-MURATORIO, J. SAFRANA, C. DE QUATTRO, B. FADY, S. C. GONZÁLEZ-MARTÍNEZ and G. G. VENDRAMIN for providing the SNP positions of their study.

The study was supported by the Ministry for Science and Culture of Lower Saxony within the network KLIFF – climate impact and adaptation research in Lower Saxony.

We thank the managers of the three Exploratories, KIRSTEN REICHEL-JUNG, SWEN RENNER, KATRIN HARTWICH, SONJA GOCKEL, KERSTIN WIESNER, and MARTIN GORKE for their work in maintaining the plot and project infrastructure; CHRISTIANE FISCHER and SIMONE PFEIFFER for giving support through the central office, MICHAEL OWONIBI for managing the central data base, and MARKUS FISCHER, EDUARD LINSENMAIR, DOMINIK HESSENMÖLLER, JENS NIESCHULZE, DANIEL PRATI, INGO SCHÖNING, FRANÇOIS BUSCOT, ERNST-DETLEF SCHULZE, WOLFGANG W. WEISSER and the late ELISABETH KALKO for their role in setting up the Biodiversity Exploratories project.

The work has been (partly) funded by the DFG Priority Program 1374 „Infrastructure-Biodiversity-Exploratories“ (DFG Fi 569/12-2). Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG).

We greatly appreciate the useful comments by two anonymous reviewers.

## References

ALBERTO, F. J., J. DERORY, C. BOURY, J.-M. FRIGERIO, N. E. ZIMMERMANN and A. KREMER (2013): Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. Genetics **195**: 495–512.



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

Balding, D. J. (2006): A tutorial on statistical methods for population association studies. Nature Reviews Genetics **7**: 781–791.

Basler, D. and C. Körner (2012): Photoperiod sensitivity of bud burst in 14 temperate forest tree species. Agricultural and Forest Meteorology **165**: 73–81.

Bertin, R. I. (2008): Plant phenology and distribution in relation to recent climate change. The Journal of the Torrey Botanical Society **135**: 126–146.

Bolte, A., T. Czajkowski and T. Kompa (2007): The north-eastern distribution range of European beech – a review. Forestry **80**: 413–429.

Caffarra, A. and A. Donnelly (2011): The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. International Journal of Biometeorology **55**: 711–721.

Carsjens, C., Q. N. Nguyen, J. Guzy, F. Knutzen, I. C. Meier, M. Müller, R. Finkeldey, C. Leuschner and A. Polle (2014): Intra-specific variations of stress-related genes in beech progenies are stronger than drought-induced responses. Tree Physiology **34**: 1348–1361.

Derory, J., P. Léger, V. Garcia, J. Schaeffer, M.-T. Hauser, F. Salin, C. Luschnig, C. Plomion, J. Glössl and A. Kremer (2006): Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). New Phytologist **170**: 723–738.

Dittmar, C., W. Fricke and W. Elling (2006): Impact of late frost events on radial growth of common beech (*Fagus sylvatica* L.) in Southern Germany. European Journal of Forest Research **125**: 249–259.

European Environment Agency (EEA) (2012): Climate change, impact and vulnerability in Europe 2012, an indicator-based report. European Environment Agency, Copenhagen.

Fischer, M., O. Bossdorf, S. Gockel, F. Hänsel, A. Hemp, D. Hessenmöller, G. Korte, J. Nieschulze, S. Pfeiffer, D. Prati, S. Renner, I. Schöning, U. Schumacher, K. Wells, F. Buscot, E. K. V. Kalko, K. E. Linsenmair, E.-D. Schulze and W. W. Weisser (2010): Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic and Applied Ecology **11**: 473–485.

Gömöry, D., L. Paule, I. M. Shvadchak, F. Popescu, M. Sulkowska, V. Hynek and R. Longauer (2003): Spatial patterns of the genetic differentiation in European beech (*Fagus sylvatica* L.) at allozyme loci in the Carpathians and the adjacent regions. Silvae Genetica **52**: 78–83.

Griffiths, S., R. P. Dunford, G. Coupland and D. A. Laurie (2003): The evolution of CONSTANS-Like gene families in barley, rice and *Arabidopsis*. Plant Physiology **131**: 1855–1867.

Gu, L., P. J. Hanson, W. M. Post, D. P. Kaiser, B. Yang, R. Nemani, S. G. Pallardy and T. Meyers (2008): The 2007 eastern US spring freeze: increasing cold damage in a warming world. BioScience **58**: 253–262.

Hall, T. A. (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**: 95–98.

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. I. Taylor, R. Ogden, M. T. Limborg, A. Cariani, G. E. Maes, E. Diopere, G. R. Carvalho and E. E. Nielsen (2011): Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources **11**: 123–136.

Hemmer-Hansen, J., E. E. G. Nielsen, D. Meldrup and C. Mittelholzer (2011): Identification of single nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel organism; the Atlantic cod, *Gadus morhua*. Molecular Ecology Resources **11**: 71–80.

Hertel, D., T. Strecker, H. Müller-Haubold and C. Leuschner (2013): Fine root biomass and dynamics in beech forests across a precipitation gradient – is optimal resource partitioning theory applicable to water-limited mature trees? Journal of Ecology **101**: 1183–1200.

Heuertz, M., E. De Paoli, T. Källman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux and N. Gyllenstrand (2006): Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. Genetics **174**: 2095–2105.

Hill, W. G. and A. Robertson (1968): Linkage disequilibrium in finite populations. Theoretical and Applied Genetics **38**: 226–231.

Horvath, D. (2009): Common mechanisms regulate flowering and dormancy. Plant Science **177**: 523–531.

Ingvarsson, P. K. (2005): Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). Genetics **169**: 945–953.

Ingvarsson, P. K., M. V. Garcia, V. Luquez, D. Hall and S. Jansson (2008): Nucleotide polymorphism and phenotypic associations within and around the *phytochrome B2* locus in European aspen (*Populus tremula*, Salicaceae). Genetics **178**: 2217–2226.

Khanduri, V. P., C. M. Sharma and S. P. Singh (2008): The effects of climate change on plant phenology. Environmentalist **28**: 143–147.

Kibbe, W. A. (2007): OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Research **35**: W43–W46.

Körner, C. and D. Basler (2010): Phenology under global warming. Science **327**: 1461–1462.

Kreyling, J., D. Thiel, L. Nagy, A. Jentsch, G. Huber, M. Konnert and C. Beierkuhnlein (2012): Late frost sensitivity of juvenile *Fagus sylvatica* L. differs between southern Germany and Bulgaria and depends on preceding air temperature. European Journal of Forest Research **131**: 717–725.

Krutovsky, K. V. and D. B. Neale (2005): Nucleotide diversity and linkage disequilibrium in cold-hardi-



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

ness- and wood quality-related candidate genes in Douglas fir. Genetics **171**: 2029–2041.

Lalagüe, H., K. Csilléry, S. Oddou-Muratorio, J. Safrana, C. de Quattro, F. Fady, S. C. González-Martínez and G. G. Vendramin (2014): Nucleotide diversity and linkage disequilibrium at 58 stress response and phenology candidate genes in a European beech (*Fagus sylvatica* L.) population from southeastern France. Tree Genetics and Genomes **10**: 15–26.

Larsson, H., T. Källman, N. Gyllenstrand and M. Lascoux (2013): Distribution of long-range linkage disequilibrium and Tajima's D values in Scandinavian Populations of Norway Spruce (*Picea abies*). G3: Genes, Genomes, Genetics **3**: 795–806.

Librado, P. and J. Rozas (2009): DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**: 1451–1452.

Lind-Riehl, J. F., A. R. Sullivan and O. Gailing (2014): Evidence for selection on a Constans-like gene between two red oak species. Annals of Botany **113**: 967–975.

Manly, B. F. J. (1985): The statistics of natural selection on animal populations. Chapman & Hall, London.

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng and S. H. Bryant (2015): CDD: NCBI's conserved domain database. Nucleic Acids Research **43**: D222–D226.

Marchler-Bauer, A. and S. H. Bryant (2004): CD-Search: protein domain annotations on the fly. Nucleic Acids Research **32**: W327–W331.

Menzel, A. and P. Fabian (1999): Growing season extended in Europe. Nature **397**: 659.

Menzel, A. (2000): Trends in phenological phases in Europe between 1951 and 1996. International Journal of Biometeorology **44**: 76–81.

Morin, P. A., G. Luikart, R. K. Wayne and the SNP workshop group (2004): SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution **19**: 208–216.

Nielsen, R. (2000): Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics **154**: 931–942.

Nielsen, R., M. J. Hubisz and A. G. Clark (2004): Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. Genetics **168**: 2373–2382.

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

Peakall, R. and P. E. Smouse (2012): GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics **28**: 2537–2539.

Rajendra, K. C., S. Seifert, K. Prinz, O. Gailing and R. Finkeldey (2014): Subtle human impacts on neutral genetic variation in European beech (*Fagus sylvatica*). Forest Ecology and Management **319**: 138–149.

Rotzen, S. and H. J. Skaletsky (2000): Primer3 on the WWW for general users and for biologist programmers. *In:* Krawetz, S. and Misener, S. (eds.) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365–386.

Rousset, F. (2008): GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources **8**: 103–106.

Sambrook, J., E. F. Fischer and T. Maniatis (1989): Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor N.Y., Cold Spring Harbor Laboratory.

Seifert, S. (2012): Variation of candidate genes related to climate change in European beech (*Fagus sylvatica* L.). Dissertation. Georg-August-University Göttingen.

Seifert, S., B. Vornam and R. Finkeldey (2012a): A set of 17 single nucleotide polymorphism (SNP) markers for European beech (*Fagus sylvatica* L.). Conservation Genetics Resources **4**: 1045–1047.

Seifert, S., B. Vornam and R. Finkeldey (2012b): DNA sequence variation and development of SNP markers in beech (*Fagus sylvatica* L.). European Journal of Forest Research **131**: 1761–1770.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**: 4673–4680.

Ueno, S., G. Le Provost, V. Léger, C. Klopp, C. Noirot, J.-M. Frigerio, F. Salin, J. Salse, M. Abrouk, F. Murat, O. Brendel, J. Derory, P. Abadie, P. Léger, C. Cabane, A. Barré, A. de Daruvar, A. Couloux, P. Wincker, M.-P. Reviron, A. Kremer and C. Plomion (2010): Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. BMC Genomics **11**: 650.

Vidalis, A., A. L. Curtu and R. Finkeldey (2013): Novel SNP development and analysis at a NADP+-specific IDH enzyme gene in a four species mixed oak forest. Plant Biology **15**: 126–137.

Višnjić, Ć. and A. Dohrenbusch (2004): Frostresistenz und Phänologie europäischer Buchenprovenienzen (*Fagus sylvatica* L.) Allgemeine Forst- und Jagdzeitung **175**: 101–108.

Vollmer, N. L. and P. E. Rosel (2012): Developing genomic resources for the common bottlenose dolphin (*Tursiops truncatus*): isolation and characterization of 153 single nucleotide polymorphisms and 53 genotyping assays. Molecular Ecology Resources **12**: 1124–1132.



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

Vornam, B., O. Gailing, R. Finkeldey, C. Collada, M. Á. Guevera, Á. Soto, N. de María, S. Gonzáles-Martínez, L. Díaz, R. Alia, I. Aranda, J. Climent, M. T. Cervera, P. Goicoechea, V. Léger, E. Eveno, J. Derory, P. Garnier-Géré, A. Kremer and C. Plomion (2007): Naturally occurring nucleotide diversity in candidate genes for forest tree adaptation: magnitude, distribution and association with quantitative trait variation. GABI – The German Plant Genome Research Program, Progress report 2004–2007, pp 116–119.

Wühlisch, Gv., D. Krusche and H.-J. Muhs (1995): Variation in temperature sum requirement for flushing of beech provenances. Silvae Genetica **44**: 343–346.

Yakovlev, I. A., C.-G. Fossdal Ø. Johnsen, O. Junttila and T. Skøppa (2006): Analysis of gene expression during bud burst initiation in Norway spruce via ESTs from subtracted cDNA libraries. Tree Genetics and Genomes **2**: 39–52.

Yeh, F. C., R.-C. Yang and T. Boyle (1999): POPGENE: Microsoft Windows-based freeware for population genetic analysis. University of Alberta, Canada (http://www.ualberta.ca/~fyeh/popgene.html)

Zhang, K. and L. Jin (2003): HaploBlockFinder: haplotype block analyses. Bioinformatics **19**: 1300–1301.

Zhu, M. and S. Zhao (2007): Candidate gene identification approach: progress and challenges. International Journal of Biological Sciences **3**: 420–427.

## Supplementary Material 1

Modifications of the cloning procedure using the TOPO TA Cloning® Kit (Invitrogen, Carlsbad, USA)

*Modifications of the transformation protocol:*

– only half of the volume of the vial of the One Shot® Chemically Competent Cells was used (30 µl),
– heat-shock of the cells during the transformation process for 45 seconds instead of 30 seconds,
– 200 µl of S.O.C. medium was added to,
– 110 µl from each transformation was spread to a selective plate.

*Table S1.* – Protein domains identified in the analyzed candidate genes.

| Gene name | Abbreviation | Domain | Accession no. of the domain in CDD | Position (bp) of the domain in the sequence |
|---|---|---|---|---|
| *Auxin response factor* | *Arf* | AUX/IAA super family | cl03528 | 47-137 |
| *Alpha amylase/subtilisin inhibitor* | *Asi* | Soybean trypsin inhibitor (Kunitz) family of protease inhibitors (STI) | cd00178 | 30-209 |
| *Constans like* | *ConsC1* | BBOX, B-Box-type zinc finger; zinc binding domain | cd00021 | 25-68 |
| | | CCT, CCT motif | pfam06203 | 250-294 |
| | *ConsC2* | BBOX, B-Box-type zinc finger; zinc binding domain | cd00021 | 21-64 and 63-107 |
| *Chloroplast chaperonin like* | *CP10* | cpn10 , Chaperonin 10 Kd subunit | cd00320 | 19-110 and 118-210 |
| *Cysteine proteinase* | *CysPro* | Peptidase_C1A | cd02248 | 3-160 |
| *Dof zinc finger protein* | *DAG* | zf-Dof, Dof domain, zinc finger | pfam02701 | 49-111 |
| *FRIGIDA* | *FRIGIDA* | Frigida super family | cl20350 | 86-142 |
| *Histone 3* | *His3C1* | H2A super family | cl00074 | 9-126 |
| | | Histone H3 (provisional) | PTZ00018 | 1-129 |
| | *His3C2* | H2A super family | cl00074 | 1-127 |
| | | Histone H3 (provisional) | PTZ00018 | 1-127 |
| *NAC transcription factor* | *NAC* | No apical meristem (NAM) protein | pfam02365 | 11-136 |
| *Protein phosphatase 2C* | *PP2C* | PP2Cc; serine/threonine phosphatases, family 2C, catalytic domain | cd00143 | 111-349 |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table S2.* – Characterization of all SNPs and indels identified in the analyzed genes. Bold: SNPs genotyped by KBiosciences UK Ltd (KASP™ Genotyping Assay; Hoddesdon, UK).

| SNP/ indel no. | Gene | Position (bp) | Characteristic | Substitution/Indel sequence | Amino acid change |
|---|---|---|---|---|---|
| **1** | *Auxin response factor (ARF)* | 239 | non-coding | A/G | |
| **2** | | 277 | non-coding | A/G | |
| **3** | | 537 | non-coding | A/G | |
| **4** | | 547 | non-coding | T/C | |
| 5 | | 565 | non-coding | A/G | |
| 6 | | 576 | non-coding | T/C | |
| **7** | | 589 | non-coding | A/G | |
| 8 | | 619 | non-coding | A/G | |
| 9 | | 736 | non-coding | A/T | |
| **10** | | 771 | non-coding | A/C | |
| 11 | | 803 | non-coding | A/G | |
| **12** | | 807 | non-coding | A/G | |
| **13** | | 852 | non-coding | A/G | |
| **14** | *alpha Amylase/subtilisin inhibitor (ASI)* | 84 | non-synonymous | T/C | leucine/ serine |
| 15 | | 313 | synonymous | T/C | |
| **16** | | 325 | synonymous | A/G | |
| 17 | | 363 | non-synonymous | A/G | aspartic acid/ glycine |
| 18 | | 467 | non-synonymous | A/C | arginine/ serine |
| 19 | | 473 | non-synonymous | T/G | alanine/ serine |
| 20 | | 646 | 3'UTR | T/C | |
| 21 | | 656-663 | 3'UTR | deletion AA; insertion: TTGTCAAC | |
| 22 | | 707 | 3'UTR | A/T | |
| 23 | | 787 | 3'UTR | A/G | |
| 24 | | 788 | 3'UTR | A/G | |
| **25** | *Constans like (1)* | 268 | non-synonymous | A/T | phenylalanine/tyrosine |
| **26** | | 281 | synonymous | A/G | |
| **27** | | 390 | non-synonymous | A/G | threonine/ alanine |
| 28 | *Constans like (2)* | 6 | 5'UTR | A | |
| **29** | | 26 | 5'UTR | A/G | |
| **30** | | 73 | non-synonymous | C/G | glycine/ alanine |
| **31** | | 122 | synonymous | T/G | |
| **32** | | 126 | non-synonymous | C/G | valine/ leucine |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table S2.* – Continued.

| | | | | | |
|---|---|---|---|---|---|
| **33** | | 186 | non-synonymous | T/G | alanine/ serine |
| **34** | | 365 | synonymous | T/C | |
| **35** | | 452 | non-synonymous | C/G | glutamic acid/ aspartic acid |
| **36** | | 463 | non-synonymous | T/C | valine/ alanine |
| 37 | | 27 | synonymous | A/C | |
| **38** | | 39 | synonymous | T/C | |
| **39** | | 41 | non-synonymous | T/C | threonine/ isoleucine |
| 40 | | 68 | non-coding | A | |
| **41** | | 78 | non-coding | T/C | |
| 42 | | 116 | non-coding | T/C | |
| 43 | | 275 | non-coding | A/T | |
| 44 | | 319 | non-coding | T/G | |
| **45** | | 351 | non-coding | T/G | |
| 46 | *Chloroplast chaperonin like (CP10 like)* | 354 | non-coding | T | |
| **47** | | 416 | non-coding | C/G | |
| **48** | | 477 | synonymous | C/G | |
| 49 | | 659 | non-coding | C/G | |
| **50** | | 723 | synonymous | C/G | |
| **51** | | 772 | synonymous | T/C | |
| 52 | | 876 | non-coding | T/C | |
| **53** | | 909 | non-coding | T/C | |
| 54 | | 978 | non-coding | A/T | |
| **55** | | 1291 | non-coding | A/G | |
| **56** | | 1402 | non-synonymous | T/C | proline/ leucine |
| 57 | | 1499 | synonymous | T/C | |
| **58** | | 118 | synonymous | C/G | |
| **59** | | 202 | synonymous | A/G | |
| 60 | | 292 | non-coding | T/G | |
| 61 | | 372 | non-coding | T/C | |
| 62 | *Cysteine proteinase* | 408 | non-coding | A/T | |
| **63** | | 728 | 3'UTR | C/G | |
| 64 | | 750 | 3'UTR | G | |
| **65** | | 783 | 3'UTR | T/G | |
| 66 | | 830 | 3'UTR | A/C | |
| 67 | | 833 | 3'UTR | A/G | |
| 68 | | 841 | 3'UTR | A/G | |
| 69 | | 887 | 3'UTR | A/C | |
| 70 | | 890 | 3'UTR | T/G | |
| **71** | | 58 | 5'UTR | A/G | |
| **72** | *Dof zinc finger protein (DAG)* | 266 | non-coding | A/T | |
| 73 | | 272 | non-coding | A | |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table S2.* – Continued.

| | | | | | |
|---|---|---|---|---|---|
| 74 | | 284-291 | non-coding | TTCAA | |
| 75 | | 345 | non-coding | A/G | |
| 76 | | 350 | non-coding | A/C | |
| 77 | | 361 | non-coding | C/G | |
| 78 | | 551 | non-coding | A/G | |
| 79 | | 623 | non-coding | A/T | |
| 80 | | 632 | non-coding | A/G | |
| 81 | | 660-665 | non-coding | microsatellite motif: GTA, three different alleles | |
| 82 | | 700 | non-coding | A | |
| 83 | | 716 | non-coding | A/G | |
| 84 | | 718 | non-coding | A/T | |
| 85 | | 719 | non-coding | A/T | |
| 86 | | 764 | non-coding | T | |
| 87 | | 792 | non-coding | T/G | |
| 88 | | 811 | synonymous | T/C | |
| **89** | | 1036 | synonymous | T/G | |
| 90 | | 1171 | synonymous | T/C | |
| **91** | | 34 | synonymous | T/C | |
| **92** | | 84 | non-synonymous | A/G | serine/asparagine |
| **93** | | 159 | non-synonymous | A/G | serine/asparagine |
| 94 | *Frigida* | 197 | synonymous | T/C | |
| 95 | | 239-244 | coding | microsatellite motif: GAA, three different alleles | glutamate |
| 96 | | 343 | synonymous | C/G | |
| 97 | | 370 | synonymous | A/G | |
| 98 | | 430 | synonymous | T/C | |
| 99 | *Histone 3 (1)* | 128-129 | non-coding | TG | |
| 100 | | 230 | synonymous | T/C | |
| **101** | | 292 | non-coding | T/C | |
| 102 | | 387 | synonymous | T/C | |
| 103 | | 434 | non-coding | A/G | |
| 104 | | 457 | non-coding | G | |
| 105 | | 866 | 3'UTR | A/G | |
| 106 | *Histone 3 (2)* | 20 | 5'UTR | T/G | |
| 107 | | 24 | 5'UTR | T/C | |
| **108** | | 104 | synonymous | A/C | |
| **109** | | 155 | non-coding | T/C | |
| **110** | | 186 | non-coding | T/C | |
| 111 | | 188 | non-coding | A/G | |
| **112** | | 260 | synonymous | A/G | |
| 113 | | 301 | non-coding | A/C | |
| 114 | | 334 | non-coding | G | |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table S2.* – Continued.

| | | | | | |
|---|---|---|---|---|---|
| 115 | | 520 | non-coding | A/C | |
| 116 | | 533 | non-coding | T/C | |
| 117 | | 557 | non-coding | T/C | |
| 118 | | *566* | *non-coding* | *T* | |
| 119 | | 88 | synonymous | A/T | |
| 120 | | 259-260 | non-coding | microsatellite motif: T, three different alleles | |
| 121 | | 553 | non-coding | A/T | |
| 122 | | *619* | *non-coding* | *T* | |
| **123** | | 834 | non-synonymous | A/C | glutamate/aspartic acid |
| **124** | *NAC transcription factor* | 942 | synonymous | A/G | |
| 125 | | 985-987 | coding | microsatellite motif: AAT, two different alleles | asparagine |
| 126 | | 1229 | 3'UTR | T | |
| 127 | | 1241-1245 | 3'UTR | complex indel consisting of A and T | |
| 128 | | 1274 | 3'UTR | A/G | |
| **129** | | 1280 | 3'UTR | A/G | |
| **130*** | | 220 | non-synonymous | T/G | lysine/ asparagine |
| **131** | | 315 | non-synonymous | C/G | alanine/glycine |
| **132** | *Protein phosphatase 2C (PP2C)* | 391 | synonymous | T/G | |
| 133 | | 538 | non-coding | T/G | |
| **134** | | 791 | non-synonymous | A/G | asparagine/ aspartic acid |
| **135** | | 941 | non-coding | T/G | |
| **136** | | 1200 | synonymous | A/G | |

*Table S3.* – SNP pairs significantly in LD in the population US ($p < 0.05$).

| | | | |
|---|---|---|---|
| Arf_265 | Arf_615 | CP10_442 | PP2C_1200 |
| Arf_265 | Arf_833 | CP10_442 | PP2C_941 |
| Arf_265 | Arf_303 | CP10_503 | CP10_1317 |
| Arf_265 | Arf_573 | CP10_503 | CP10_65 |
| Arf_303 | Arf_573 | CP10_503 | CP10_377 |
| Arf_563 | Arf_615 | CP10_503 | CP10_749 |
| Arf_563 | PP2C_391 | CP10_65 | CP10_377 |
| Arf_563 | DAG_289 | CP10_65 | CP10_749 |
| Arf_563 | DAG_81 | CP10_65 | FRIGIDA_179 |
| Arf_563 | CP10_67 | CP10_65 | PP2C_391 |



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

*Table S3.* – Continued.

| | | | |
|---|---|---|---|
| Arf_615 | Arf_833 | CP10_749 | DAG_289 |
| Arf_615 | Arf_303 | CP10_749 | DAG_81 |
| Arf_615 | Arf_573 | CP10_749 | PP2C_791 |
| Arf_615 | ConsC1_306 | CysPro_118 | CysPro_783 |
| Arf_833 | Arf_573 | CysPro_118 | NAC_854 |
| Arf_833 | ConsC1_293 | CysPro_202 | PP2C_941 |
| Arf_878 | Arf_615 | CysPro_202 | PP2C_1200 |
| Arf_878 | Arf_303 | CysPro_202 | His3C2_186 |
| Arf_878 | Arf_265 | CysPro_728 | CysPro_118 |
| Arf_878 | Arf_573 | CysPro_728 | CysPro_783 |
| ConsC1_293 | ConsC1_306 | CysPro_783 | His3C2_186 |
| ConsC1_293 | ConsC2_488 | CysPro_783 | NAC_854 |
| ConsC1_293 | His3C2_260 | DAG_289 | His3C2_186 |
| ConsC2_147 | DAG_81 | DAG_289 | His3C2_260 |
| ConsC2_147 | DAG_289 | DAG_81 | DAG_289 |
| ConsC2_147 | His3C1_292 | DAG_81 | His3C2_186 |
| ConsC2_151 | ConsC2_488 | DAG_81 | His3C2_260 |
| ConsC2_151 | ConsC2_51 | FRIGIDA_104 | FRIGIDA_54 |
| ConsC2_211 | ConsC2_390 | FRIGIDA_54 | NAC_1300 |
| ConsC2_488 | ConsC2_51 | His3C1_292 | NAC_854 |
| ConsC2_488 | ConsC2_147 | His3C2_260 | His3C2_186 |
| ConsC2_488 | NAC_962 | NAC_854 | NAC_962 |
| ConsC2_51 | NAC_854 | NAC_854 | NAC_1300 |
| ConsC2_98 | His3C1_292 | NAC_962 | NAC_1300 |
| CP10_1317 | CP10_65 | PP2C_315 | PP2C_391 |
| CP10_1317 | CP10_377 | PP2C_315 | PP2C_1200 |
| CP10_1317 | CP10_749 | PP2C_315 | PP2C_941 |
| CP10_1428 | CP10_442 | PP2C_391 | PP2C_1200 |
| CP10_1428 | CP10_377 | PP2C_941 | PP2C_391 |
| CP10_377 | CP10_749 | PP2C_941 | PP2C_1200 |
| CP10_442 | CP10_377 | | |

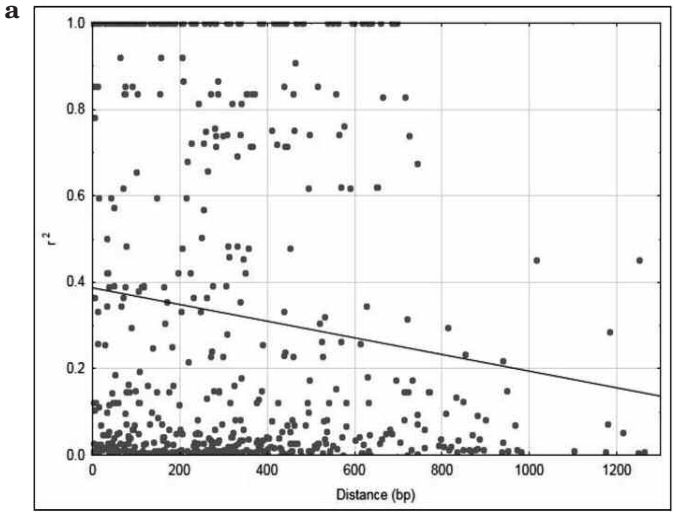


*Figure S1.* – LD plots for the investigated candidate genes. Displayed is a LD plot for all genes combined (a), and LD plots for each of the candidate genes separately (b).



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

**b**

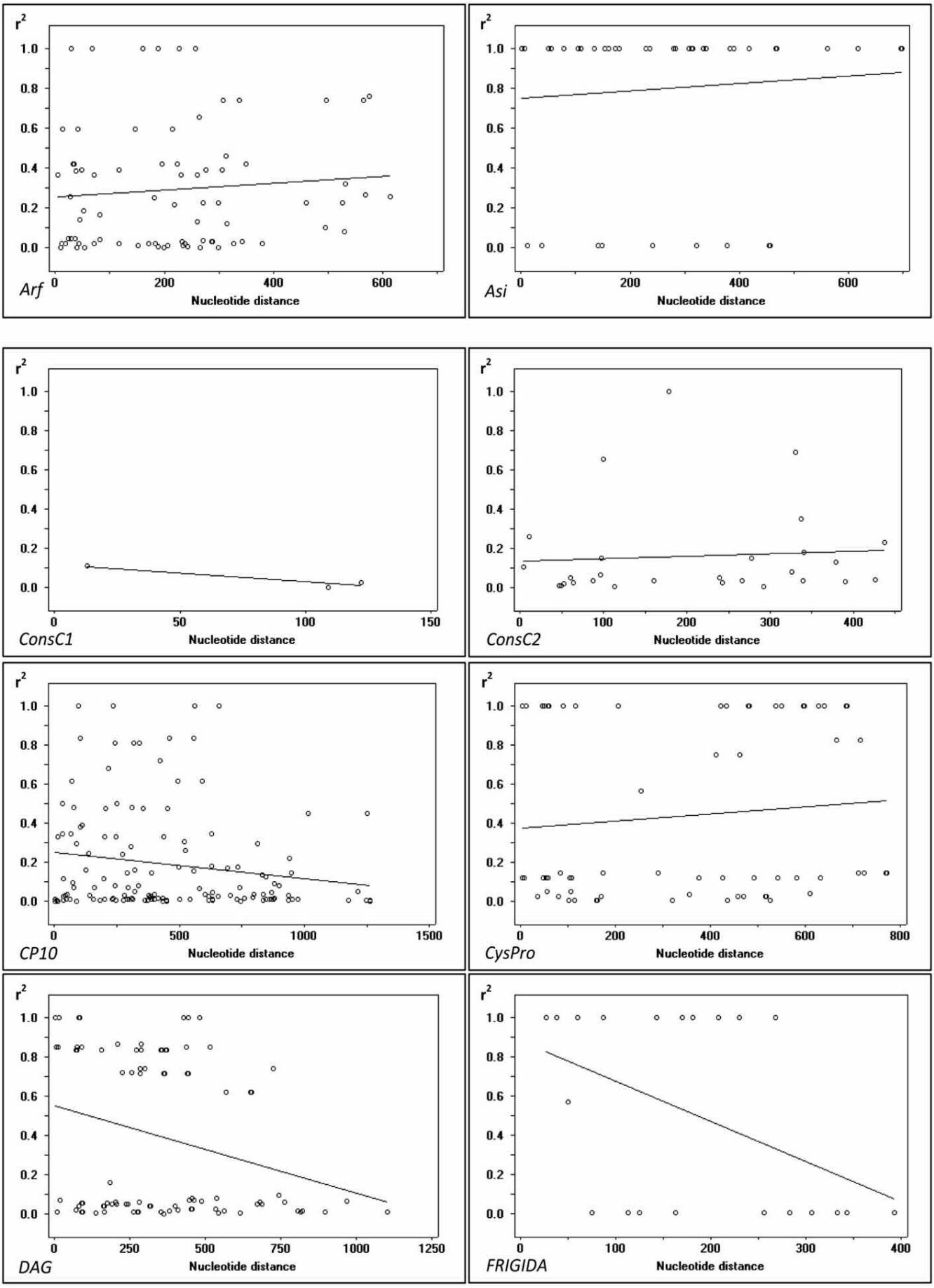


*Figure S1.* – Continued.



DOI:10.1515/sg-2015-0001
edited by Thünen Institute of Forest Genetics

**b**

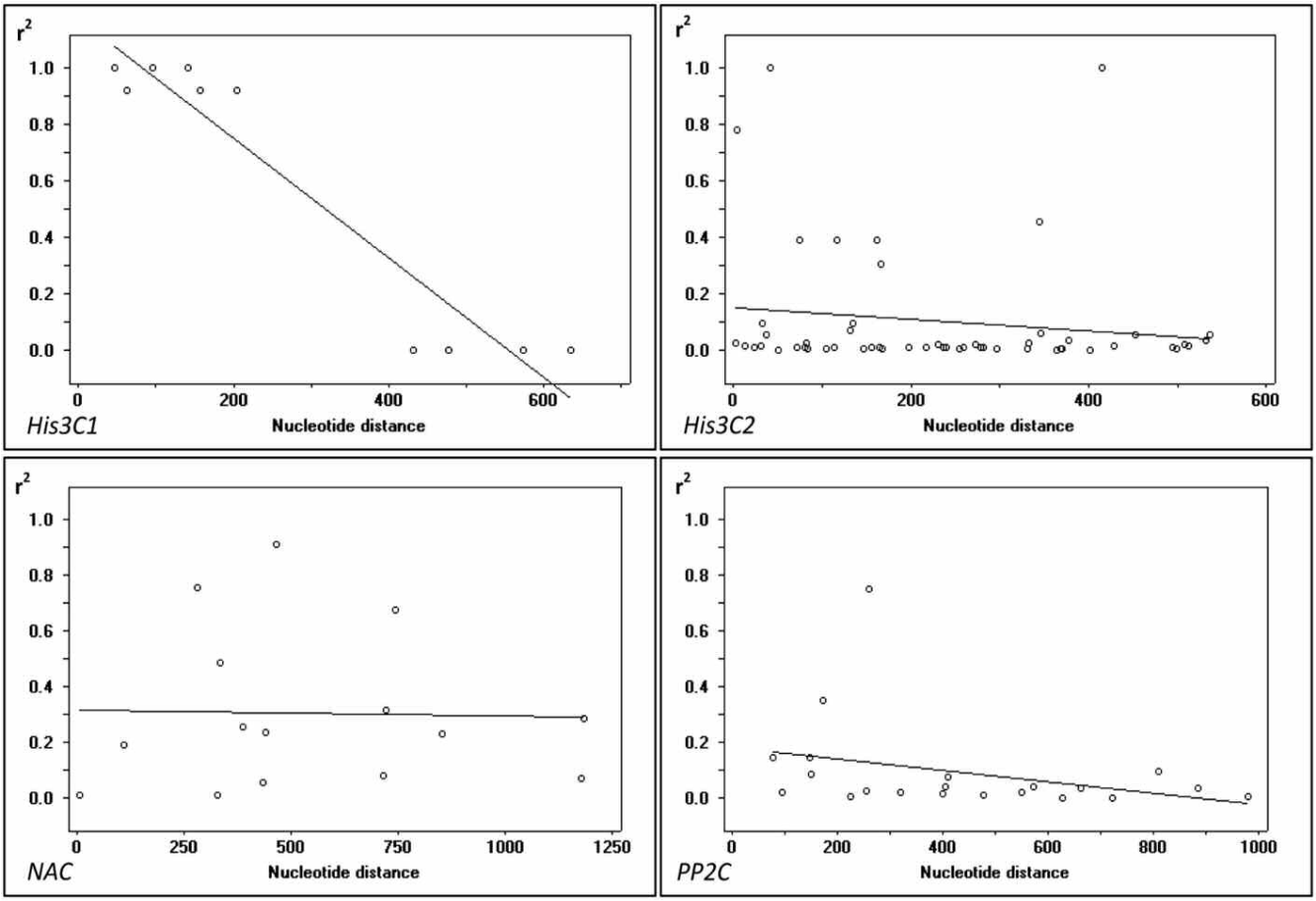


*Figure S1.* – Continued.

# How small and constrained is the genome size of angiosperm woody species

By D. Ohri*)

(Received 8th May 2015)

**Abstract**

Angiosperm hardwood species are generally considered to show an average smaller genome size with a narrow range of variation than their herbaceous counterparts. Various explanations pertaining to limitations of cell size exerted by wood fibers, the requirement of smaller stomata, longer generation time, large population size, etc., have been put forward to account for their small and constrained genome size. Yet studies done in the past several years show that genomically as well as evolutionarily, hardwoods are as diverse and active as their herba-

*) Corresponding author: Deepak Ohri. Amity University Uttar Pradesh (Lucknow Campus), Malhaur (Near Railway Station), Gomti Nagar Extension, Lucknow-226028 (U.P.), India. Telephone: +91-9452734145.
E-Mail: ohri_deepak@rediffmail.com; dohri@lko.amity.edu



DOI:10.1515/sg-2015-0002
edited by Thünen Institute of Forest Genetics