Genetic analyses to determine the genetic variation and region of origin of introduced northern red oak (*Quercus rubra* L.) populations

Dissertation

To attain the doctoral degree (Dr. forest.)

of the Faculty of Forest Sciences and Forest Ecology

Georg-August-Universität Göttingen

Submitted by

Tim Pettenkofer

born on the 10.10.1987 in Berlin, Germany

Göttingen, March 2020

1st Referee	PD Dr. Ludger Leinemann
2nd Referee	Prof Dr. Christian Ammer
Date of oral examination	27.09.2019

"If you only read the books that everyone else is reading,
you can only think what everyone else is thinking."
— Haruki Murakami, Norwegian Wood

ACKNOWLEDGEMENTS

First and foremost, I want to thank my supervisor Ludger Leinemann for his guidance and valuable input over the past three years. His kindness in all matters, scientific experience and anecdotes during coffee break will stay memorable for me and made stressful times pass much faster. Speaking of anecdotes during coffee breaks - I also want to thank Barbara Vornam, who was always there when I had a question and who regrettably really did retire after threatening us to do so for quite a while. I think I speak for everyone when I say that this meant a great loss not only on the personal- but also on the professional level for the department.

I am also very grateful for having had Oliver Gailing as head of the department. Being a renowned red oak specialist and a determined researcher, he contributed not only his unique expertise, but, through his corrections, let my manuscripts shine in a new light, too. The same is true for Konstantin 'Kostya' Krutovsky, who I want to thank for his valuable comments and suggestions that let me see the bigger picture.

This study was made possible by the funding of the Federal Ministry of Food and Agriculture (BMEL) through the Fachagentur Nachwachsende Rohstoffe (FNR). As it is a joint project together with the silviculture department, I want to thank my colleague Katharina Burkardt especially for the enjoyable time during data collection in the forests. I also thank her, Christian Ammer, Torsten Vor and Andreas Parth in general for the fruitful collaboration over the past years and hope that our paths will cross again someday.

For inspiring and motivating me to work in the field of forest genetics during my studies and to start my doctorate, I want to thank Martin Ziehe, Reiner Finkeldey and the late Sarah Seifert.

I would not be able to finish this thesis without Alexandra Dolynska, who was an excellent mentor in any lab-related tasks and who I admire for her organisational skills and straight-forwardness. Not any less important for the realisation of this thesis are Christine Radler, Markus Müller, Natalie Breidenbach, Oleksandra 'Lesya' Kuchma, Devrim Semizer Cuming and Oliver Caré. They provided me with hints and ideas when I was stuck, and I enjoyed every

minute of essential procrastination. For genuine English proofreading, I want to thank Edward Jones for an incredible job when I asked him to look over some of my manuscripts.

Along the way, I met a lot of friendly people and colleagues who supported me or the study in one way or the other. For all who I forgot to mention here by name – I have not forgotten your contribution!

I am incredibly grateful for my parents Joachim and Domenika and my siblings Moritz and Maxi Pettenkofer for supporting me in every way during all the years of my studies. I would undoubtably be somewhere and someone else if it were not for you.

Lastly, I want to use these lines to thank my love Monika Mozurevičiūtė. Tu visada sugebi nuraminti ir palaikyti, kai man to labiausiai reikia. Tu tapai dalimi manęs ir aš negaliu įsivaizduoti savo gyvenimo be tavęs.

PREFACE

The aspiration of this work is to study the genetic variation of northern red oak populations in Germany and in the natural distribution range. To make inferences about the putative origin of introduced red oak stands, chloroplast markers were used. In addition, a new set of chloroplast CAPS markers was developed to enhance the resolution of spatial chloroplast variation patterns. The use of neutral and potentially adaptive nuclear markers allowed inferences on the occurrence of putative genetic bottlenecks. Also, recommendations regarding the origin of the material for future establishments of red oak stands could be made.

This thesis is comprised of four scientific papers which either are already, or are in the process, to be published in relevant peer-reviewed journals. Also, a new approach was chosen in the structure of the thesis: Being in place of a separate introduction, a review paper was submitted to a renowned German scientific journal to make the information available for the German readership. It includes a general introduction to the topic, a description of all relevant studies and their outcome in this field, as well as a synopsis in both German and English. For this reason, this manuscript, with the English title ‘Studies of the genetic variation and origin of northern red oak (*Quercus rubra* L.) in natural and introduced populations‘, was places first.

The second paper (Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison with reference native North American populations) discusses the spatial distribution of chloroplast microsatellite haplotypes of both German and North American populations. The third paper (Development of novel *Quercus rubra* chloroplast genome CAPS markers) provides a new set of CAPS markers which proved to enhance the spatial resolution of chloroplast haplotypes in combination with cpSSR markers used in Paper 2. For this reason, the chloroplast genomes of eight individuals from different regions within the red oak’s natural distribution range were sequenced.

The thesis ends with the last paper (Genetic variation of introduced red oak (*Quercus rubra*) stands in Germany in comparison to North American populations), which focusses on the genetic variation of German and North American populations on potentially adaptive as well as putatively neutral nuclear microsatellite markers.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i
PREFACE iii
Eine Übersicht zu Untersuchungen der Herkunft und genetischen Variation der Roteiche (*Quercus rubra* L.) in natürlichen und eingeführten Populationen 1
1 Einleitung 4
2 Studien zur genetischen Variation der Roteiche 6
3 Schlussfolgerung und Ausblick 14
4 Zusammenfassung 15
5 Summary 16
6 Literaturverzeichnis 18
Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison with reference native North American populations 25
1 Introduction 28
2 Materials and methods 31
3 Results 35
4 Discussion 39
5 References 42
6 Supplementary Files 47
Development of novel *Quercus rubra* chloroplast genome CAPS markers 55
1 Introduction 59
2 Material and methods 60
3 Results 68
4 Discussion 71
5 References 71
Genetic variation of introduced red oak (*Quercus rubra*) stands in Germany in comparison to North American populations 77
1 Introduction 80
2 Material and Methods 83
3 Results 88
4 Discussion 93
5 References 96
6 Supplementary Files 100
CONTRIBUTIONS TO THE SCIENTIFIC PAPERS 107
SYNOPSIS 109
ZUSAMMENFASSUNG 111
CURRICULUM VITAE ..... **Fehler! Textmarke nicht definiert.**

LIST OF ABBREVIATIONS

AMOVA	Analysis of Molecular Variance
CAPS	Cleaved Amplified Polymorphic Sequence
COL	CONSTANS-like
EST	Expressed Sequence Tag
Indel	Insertion and deletion
IAM	Infinite Allele Model
IR	Inverted Repeat
LSC	Large Single Copy
MCMC	Markov Chain Monte Carlo
NJT	Neighbor-Joining Tree
PCoA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
PVP	Polyvinylpyrrolidone
RADseq	Restriction site Associated DNA sequencing
RFLP	Restriction Fragment Length Polymorphism
SMM	Stepwise Mutation Model
SNP	Single Nucleotide Polymorphism
SSC	Small Single Copy
(cp/n) SSR	(chloroplast/nuclear) Short Sequence Repeat
TPM	Two-Phase Model

Paper 1

Eine Übersicht zu Untersuchungen der Herkunft und genetischen Variation der Roteiche (*Quercus rubra* L.) in natürlichen und eingeführten Populationen

Published

Allgemeine Forst- und Jagdzeitung

14 February 2020

doi:10.23765/afjz0002036

Eine Übersicht zu Untersuchungen der Herkunft und genetischen Variation der Roteiche (*Quercus rubra* L.) in natürlichen und eingeführten Populationen

Tim Pettenkofer[1] • Ludger Leinemann[1] • Oliver Gailing[1,2,3]

Kontakt: Tim Pettenkofer[1], E-Mail: t.pettenkofer@gmail.com; Oliver Gailing[1,2,3], Ludger Leinemann[1], E-Mail: lleinem@gwdg.de; E-Mail: ogailin@gwdg.de

[1] Georg-August-Universität Göttingen, Fakultät für Forstwissenschaften und Waldökologie, Abteilung Forstgenetik und Forstpflanzenzüchtung, Büsgenweg 2, 37077 Göttingen, Deutschland

[2] Michigan Technological University, School of Forest Resources and Environmental Science, Houghton, 49931 Michigan, USA

[3] Georg-August-Universität Göttingen, Center for Integrated Breeding Research (CiBreed), Göttingen, Deutschland

Keywords: Roteiche, *Quercus rubra*, Herkunft, Genetische Variation, Nichtheimische Baumart

1 Einleitung

Die Roteiche (*Quercus rubra* L.) hat ihr natürliches Verbreitungsgebiet im Osten und Nordosten der USA und im Südosten von Kanada und wurde bereits Ende des 17. Jahrhunderts zunächst in Frankreich eingeführt (Goeze 1916; Hickel 1932). Erst später wurde die Roteiche in Parks und botanischen Gärten in ganz Europa angepflanzt (Bauer 1951). In Deutschland erlebte die Roteiche zwei intensivere Anbauphasen: Eine in der zweiten Hälfte des 19. Jahrhunderts und eine weitere zur Jahrhundertwende vom 19. zum 20. Jahrhundert (Bauer 1951; Nagel 2015). Gegenüber heimischen Weißeichenarten kann die Roteiche bereits nach lediglich 80-120 Jahren geerntet werden und weist einen vergleichsweise niedrigen Wasser- und Nährstoffbedarf auf (Nagel 2015). Relativ anspruchslos zeigt sich die Roteiche generell in Hinblick auf ihren Standort: Sie benötigt mindestens 600 mm jährlichen Niederschlag und erträgt harte Winter und heiße Sommer mit mittleren Temperaturen zwischen −14 bis 26 °C (Dreßel und Jäger 2002). Aktuell ist die Roteiche mit 44.000 ha die flächenmäßig bedeutendste nichtheimische Laubbaumart in Deutschland (Bundesministerium für Ernährung und Landwirtschaft (BMEL) 2014). Ihr Holz findet in zahlreichen Produkten Verwendung, ist jedoch aufgrund seiner unverthyllten Gefäße nicht für den Bau von Weinfässern geeignet (Dengler 1944). Die Roteiche gilt in Deutschland als anbauwürdig und kann die Anpassung von Wäldern an prognostizierte Klimaänderungen verbessern (Klemmt et al. 2013; Nagel 2015).

1.1 Zum Thema Invasivität und Einführung nichtheimischer Baumarten

In vielen Ländern Europas (z.B. Belgien, Polen, Tschechische Republik, Litauen) gilt die Roteiche unter anderem aufgrund ihrer hohen Verbreitungsfähigkeit auf weniger fruchtbaren Böden und negativen Auswirkungen auf die Anzahl und Vielfalt von Mikroorganismen und Gräsern im Unterstand als invasive Art (Möllerová 2005; Branquart et al. 2007; Riepšas und Straigytė 2008; Chmura 2013). In Deutschland wird die Roteiche aufgrund von ineffektiven Vektoren für die Samenverbreitung, starkem Verbiss der Verjüngung und der Abwesenheit von vegetativer Vermehrung nicht als invasiv eingestuft (Nagel 2015). Zudem gilt die Roteiche als weniger schattentolerant als die Hauptbaumart Rotbuche (Niinemets und Valladares 2006; Nagel 2015). Dennoch kann die Roteiche heimische Eichenarten durch ihre höhere Schattentoleranz und Konkurrenzfähigkeit in allen Entwicklungsstadien überwachsen (Stratmann und Warth 1987; Vor und Lüpke 2004; Kuehne et al. 2014). In frühen Stadien ist

die Roteiche sogar in der Lage dominante heimische Baumarten wie die Rotbuche zu überwachsen, wenn die Lichtverhältnisse es zulassen (Vor und Lüpke 2004).

Die natürliche Verjüngung gelingt der Roteiche in ihrem heimischen Verbreitungsgebiet immer weniger. Als Gründe dafür werden neben dem Verbiss der Sämlinge unter anderem auch die steigende Verhütung von Waldbränden und die Hinwendung zur Naturverjüngung im Unterstand genannt (Crow 1988; Dey und Parker 1996; Nicolescu et al. 2018). Auch in Deutschland ist die Naturverjüngung nicht das übliche Vorgehen bei der Bestandesbegründung. Die meisten deutschen Roteichenbestände wurden durch Pflanzung oder Aussaat als Reinbestand oder in Trupps im Mischbestand begründet (Nagel 2015).

1.2 Genetische Aspekte der Einführung von nichtheimischen Baumarten

Die Einführung einer neuen Art bzw. die „Gründung" einer neuen Population in einem fremden Ökosystem geht in den meisten -wenn nicht allen- Fällen auch mit einem mehr oder weniger ausgeprägten genetischen Flaschenhals einher (Nei et al. 1975; Dlugosch und Parker 2008). Diese Gründerpopulationen sind oft klein und weisen bei der Abwesenheit von Genfluss nur einen Bruchteil der genetischen Variation ihres natürlichen Verbreitungsgebiets auf, wodurch ihr Anpassungspotential deutlich beeinträchtigt werden kann (Nei et al. 1975; Barrett und Husband 1990; Magni Diaz 2004; Graw 2005). Verschiedene Untersuchungen haben jedoch gezeigt, dass mehrfache Einführungen die genetische Variation einer Gründerpopulation nachhaltig erhöhen und so sowohl ihr Anpassungs-, als auch ihr Verbreitungspotential verbessern können (Dlugosch und Parker 2008; Suarez und Tsutsui 2008).

Grundsätzlich wird bei der Einführung von Arten das Ziel verfolgt, möglichst viel genetische Variation und damit zumindest theoretisch auch ein entsprechend hohes Anpassungspotential in der neuen Umwelt zu etablieren. Bei Baumarten mit sehr großen Verbreitungsgebieten kann allerdings die geografisch/genetische Differenzierung der Art nicht unberücksichtigt bleiben. Aufgrund örtlich differenzierter adaptiver Prozesse sind häufig nicht alle Provenienzen einer Art gleichermaßen für die im Zielland herrschenden Klimabedingungen und Standortverhältnisse geeignet. Die Wahl der richtigen Herkunft beeinflusst Merkmale wie die Frost- und Trockenstressresistenz sowie den Zeitpunkt des Austriebs und kann zur Steigerung der Holzproduktion beitragen (Kriebel et al. 1976; Liesebach und Schneck 2011). Liesebach und Schneck (2011) führen zum Beispiel aus, dass Herkünfte aus dem nördlichen Verbreitungsgebiet „wüchsiger" im Vergleich zu Herkünften

aus dem südlichen Verbreitungsgebiet sind und somit Ertragssteigerungen erwarten lassen. Offenbar ist jedoch der genetische Ursprung aller europäischen Roteichenpopulationen gänzlich unbekannt (Nicolescu et al. 2018). Bauer (1954) vermutet, dass das Saatgut hauptsächlich aus dem nördlichen Bereich der Allegheny Mountains im Nordosten des natürlichen Verbreitungsgebietes stammt.

Das Ziel dieser Arbeit ist eine zusammenfassende Betrachtung der vorhandenen Arbeiten zur Herkunft und genetischen Vielfalt deutscher bzw. europäischer Roteichenpopulationen im Vergleich zu Populationen aus dem natürlichen Verbreitungsgebiet. Sie soll relevante Erkenntnisse zusammenfassen und eine Grundlage für zukünftige Untersuchungen zu dieser Baumart bieten. Dabei wird das Ziel verfolgt, das innerartliche genetische Potential der Roteiche zu charakterisieren und für die Erhöhung der Anpassungsfähigkeit des Waldes in Deutschland zu nutzen.

2 Studien zur genetischen Variation der Roteiche

2.1 Untersuchungen der genetischen Variation der Roteiche in Nordamerika

Phänotypische Untersuchungen in Herkunftsversuchen gehörten zu den ersten Arbeiten zur Erfassung der genetischen Variation der Roteiche in ihrem natürlichen Verbreitungsgebiet (siehe auch Tabelle 1). Neben der Wuchsleistung und Sterblichkeitsrate waren vor allem auch der Austriebszeitpunkt im Frühjahr und die herbstliche Laubverfärbung von Interesse (z.B. Kriebel et al. 1976; Schlarbaum et al. 1982; Kriebel et al. 1988). Schlarbaum et al. (1982) zogen aus ihren Beobachtungen sogar Rückschlüsse auf nacheiszeitliche Rückwanderungsbewegungen der Roteiche. Daubree und Kremer (1993) untersuchten erstmals den Zusammenhang der Variation an Allozym-Genorten mit phänotypischen Merkmalen für die Roteiche. Neu war dabei auch der Vergleich der genetischen Variation von 23 in Europa eingeführten Roteichenbeständen unbekannter Herkunft mit der von neun nordamerikanischen Populationen eines Herkunftsversuches in Frankreich. Die Ergebnisse von Daubree und Kremer (1993) zeigen eine klinale Veränderung der Merkmale Austriebszeitpunkt und herbstliche Laubverfärbung in Nord-Süd Richtung. Europäische Bestände liegen dabei im intermediären Bereich. Die Autoren schließen aus ihren Ergebnissen, dass die eingeführten Bestände ihren Ursprung in verschiedenen Regionen innerhalb des natürlichen Verbreitungsgebiets haben (Daubree und Kremer 1993). Die Ergebnisse dieser Herkunftsversuche weisen auf genetische Unterschiede in den beleuchteten adaptiven Merkmalen hin.

Romero-Severson et al. (2003) haben als erste Genorte im Chloroplastengenom der Roteiche genutzt, um regionale Verteilungsmuster im US Bundesstaat Indiana zu untersuchen. Sie nutzten dafür drei PCR-RFLP (engl. Restriction Fragment Length Polymorphism) Marker und zeigten vor allem die hohe genetische Differenzierung zwischen Populationen auf regionaler Ebene (G_{ST} = 0,73) als Vorteil von Chloroplasten-Markern auf. Magni et al. (2005) nahmen diese Studie zum Anlass für die Untersuchung von 290 Bäumen in 66 Populationen des gesamten natürlichen Verbreitungsgebiets an fünf neuen PCR-RFLP Markern. Im Gegensatz zu Romero-Severson et al. (2003) fanden Magni et al. (2005) nur eine vergleichsweise geringe Differenzierung zwischen Populationen (G_{ST} = 0.46), welche in Richtung Norden zunimmt. Die Autoren vermuten nacheiszeitliche Rückwanderungsbewegungen als eine mögliche Ursache. Auch Birchenko et al. (2009) untersuchten weite Teile des natürlichen Verbreitungsgebiets (35 Populationen) an Chloroplasten-Markern und fanden eine abnehmende genetische Vielfalt von Süd nach Nord. Allerdings konnte auch hier keine genaue Aussage über nacheiszeitliche Rückwanderungsrouten gemacht werden. Sie nutzten dieselben PCR-RFLP Marker wie Romero-Severson et al. (2003).

Gerwein und Kesseli (2006) untersuchten erstmals die genetische Variation und regionale Verteilungsmuster an vier nuklearen Mikrosatelliten-Genorten und drei Chloroplasten-Mikrosatelliten (cpSSR) im US-Bundesstaat Massachusetts im Nordosten des natürlichen Verbreitungsgebiets. Sie verglichen dabei fünf Primärwald- mit fünf Sekundärwald-Populationen. Die Ergebnisse zeigen eine eher geringe genetische Differenzierung für beide Markersysteme. Interessant ist jedoch der cpSSR Locus D2T2 (Dumolin-Lapegue et al. 1999), welcher durch das häufige Vorkommen eines privaten Allels in einer Sekundärwald-Population eine außerordentlich hohe genetische Differenzierung (F_{ST} = 0.92) aufweist (Gerwein und Kesseli 2006). Leider gehen Gerwein und Kesseli (2006) in ihrer Studie nicht auf mögliche Ursachen ein und geben auch keine weiteren Informationen zu der betroffenen Population und den vorkommenden Allelen. Zukünftige Untersuchungen an cpSSR Loci der Roteiche sollten diesen Marker miteinbeziehen.

Untersuchungen im Norden der USA konnten erfolgreich die genetische und morphologische Unterscheidung von den eng verwandten und hybridisierenden Roteichen-Arten *Q. rubra* und *Q. ellipsoidalis* nachweisen (Gailing et al. 2012; Lind und Gailing 2013). Dazu wurden sowohl neutrale Mikrosatelliten- (SSR) und potentiell adaptive (EST-SSR)

Marker, als auch morphologische Blattmerkmale genutzt. Darüber hinaus konnten zwei potentiell selektive Marker, sogenannte ‚Outlier' (3A05 und GOT021), identifiziert werden. Letzterer ist möglicherweise an der Ausprägung der Blattmerkmale und der Trockenstressresistenz beteiligt (Lind und Gailing 2013). Lind-Riehl et al. (2014) identifizierten in einer weiteren Studie den EST-SSR FIR013 als dritten Outlier-Locus zur möglichen Unterscheidung zwischen *Q. rubra* und *Q. ellipsoidalis*. FIR013 liegt im kodierenden Bereich des Gens *CONSTANS-like*, welches u.a. mutmaßlich an der Regulierung des Blühzeitpunkts beteiligt ist (Lind-Riehl et al. 2014). Eine weitere Arbeit, welche sich mit der Unterscheidung zwischen *Q. rubra*, *Q. ellipsoidalis* und *Q. velutina* beschäftigt, ist die von Zhang et al. (2015). Sie analysierten 37 Reinbestände beider Arten (davon 20 *Q. rubra*) im Bereich der Großen Seen und konnten an den untersuchten drei Chloroplasten-Mikrosatelliten Markern zwar keine Unterscheidung zwischen den Arten (G_{ST} = 0,023), wohl aber den Genfluss zwischen benachbarten Populationen verschiedener eng verwandter Arten nachweisen.

Eine weitere Studie zum gesamten natürlichen Verbreitungsgebiet der Roteiche stellt die von Borkowski et al. (2017) dar. Die Autoren untersuchten hierbei 23 Populationen an 10 neutralen Kern-Mikrosatelliten, um die genetische Differenzierung zwischen Populationen auf lokaler und regionaler Ebene zu charakterisieren. Sie fanden, ähnlich wie in vorangegangenen Studien, eine klinale Veränderung des populationsspezifischen F_{ST} -Werts, welche sie auf nacheiszeitliche Ausbreitungsbewegungen zurückführen (Borkowski et al. 2017).

Zuletzt untersuchten Merceron et al. (2017) die genetischen Strukturen der Roteiche sowohl in ihrem natürlichen Verbreitungsgebiet als auch in ausgewählten europäischen Populationen an 69 SNP Markern. Ihre Ergebnisse zeigen in Nordamerika einen Nord-Süd-Gradient mit abnehmender genetischer Variation in Richtung Norden. Auf der Grundlage dieses Gradienten vermuten sie die Herkunft europäischer Populationen im Norden des natürlichen Verbreitungsgebietes. In Europa fehlt neben einem dritten genetischen Cluster auch jegliche geografisch Verteilungsstruktur.

Tabelle 1 Zusammenfassung der Ergebnisse der wichtigsten Studien zur Roteiche

Autor	Gebiet	h_S	h_T	H_O	H_e	N_a	F_{ST}	G_{ST}
Daubree und Kremer 1993[1]	Europa	-	-	-	0,33	13,00	-	-
	Nordamerika	-	-	-	0,27	15,10	-	-
Sork et al. 1993[1]	Nordamerika	-	-	-	-	-	0,09	-
Aldrich et al. 2002[2]	Nordamerika	-	-	0,68	0,74	7,50	-	-
Romero-Severson et al. 2003[3]	Nordamerika	0,21	0,78	-	-	-	-	0,73
Magni Diaz 2004[3]	Europa	0,33	0,40	-	-	-	-	0,18
Magni et al. 2005[3]	Nordamerika	0,22	0,41	-	-	-	-	0,47
Gerwein und Kesseli 2006[4,5]	Nordamerika	-	-	-	-	6,70	0,31	-
		-	-	-	-	24,10	0,03	-
Birchenko et al. 2009[3]	Nordamerika	-	-	-	-	-	-	0,58
Sullivan et al. 2013[6]	Nordamerika	-	-	0,63	0,76	7,89	0,08	-
Lind und Gailing 2013[7,2]	Nordamerika	-	-	0,61	0,71	11,17	0,01	-
		-	-	0,73	0,84	14,50	0,02	-
Lind-Riehl et al. 2014[6]	Nordamerika	-	-	-	0,72	15,00	-	-
Zhang et al. 2015[4]	Nordamerika	0,35	0,44	-	-	2,30	-	0,21
Borkowski et al. 2017[2]	Nordamerika	-	-	-	-	10,27	0,04	-
Merceron et al. 2017[8]	Nordamerika	-	-	-	0,33	-	-	-
	Europa	-	-	-	0,33	-	-	-
Pettenkofer et al. 2019[4]	Deutschland	0,29	0,34	-	-	8,00	-	0,14
	Nordamerika	0,18	0,65	-	-	8,00	-	0,73
Pettenkofer et al. (eingereicht, b)[6]	Deutschland	-	-	0,61	0,63	9,40	0,01	-
	Nordamerika	-	-	0,59	0,65	9,44	0,03	-

h_S = Genetische Diversität innerhalb von Populationen; h_T = Gesamte genetische Diversität; H_o = Beobachtete Heterozygotie; H_e = Erwartete Heterozygotie; N_a = Anzahl der Allele; F_{ST}, G_{ST} = Genetische Differenzierung; Markertyp: [1] Alloenzyme; [2] SSR; [3] cpDNA PCR-RFLP; [4] cpSSR; [5] nSSR; [6] SSR und EST-SSR; [7] EST-SSR; [8] SNP

2.2 Herkunft und Variation europäischer Populationen

Schon Bauer (1951) regte in seiner Dissertation zur Roteiche die Untersuchung des Ursprungs deutscher Bestände an. Dennoch war Magni Diaz (2004) auf europäischer Ebene der Erste, der in einem bislang unveröffentlichten Manuskript in seiner Dissertation die Frage der genetischen Herkunft eingeführter Bestände behandelt. Dieser Autor untersuchte die genetische Variation im Chloroplastengenom mit Hilfe von fünf PCR-RFLP Markern. Seine 66 untersuchten europäischen Herkünfte bestehen aus insgesamt 300 Individuen, deren Provenienz nicht überliefert ist. Sie beinhalten unter anderem 22 Proben aus 5 deutschen Beständen. Er vergleicht diese europäischen Proben mit 290 Individuen aus 66 Beständen im natürlichen Verbreitungsgebiet der Roteiche, welche an denselben Markern untersucht wurden (Magni Diaz 2004; Magni et al. 2005). Im Vergleich zwischen den nordamerikanischen und europäischen Populationen zeigen seine Ergebnisse ähnliche Werte für die Diversität. Weder nordamerikanische- noch europäische Populationen weisen geografische Verteilungsmuster in der genetischen Vielfalt auf (Magni et al. 2005). Der einzige Unterschied zwischen den beiden Gruppen sei eine geringere Differenzierung zwischen europäischen Populationen im Vergleich zur Differenzierung zwischen Populationen im natürlichen Verbreitungsgebiet, wo die genetische Differenzierung nach Norden hin zunimmt. Magni Diaz (2004) führt dies auf die mehrfache Einführung von Saatgut und in der Folge künstliche Beimischung in europäischen Beständen zurück. Eine genaue Angabe über die mögliche Herkunft europäischer Bestände ist nicht zu finden. Dennoch beinhaltet die Arbeit interessante Ergebnisse eines Vergleichs zwischen historisch französischen und historisch deutschen Populationen (letztere beinhalten neben einem Bestand in Baden-Württemberg auch 11 weitere in Elsass-Lothringen). Sowohl die gesamte Diversität h_T, als auch die Diversität innerhalb der Populationen h_S ist in historisch deutschen Beständen etwa doppelt so hoch wie in historisch französischen (h_T = 0,62 bzw. 0,30; h_S = 0,44 bzw. 0,26). Auch die Differenzierung zwischen Populationen G_{ST} ist in historisch deutschen Beständen deutlich höher als in historisch französischen (G_{ST} = 0,29 bzw. 0,11). Magni Diaz (2004) führt dies auf mögliche Unterschiede in der Import- und Forstpolitik beider Länder zurück.

Liesebach und Schneck (2011) untersuchten amerikanische und deutsche Roteichen-Herkünfte in vier Herkunftsversuchen in Schleswig-Holstein, Hessen, Brandenburg und Niedersachsen (wobei letzterer kurz nach der Einrichtung wieder aufgegeben werden musste). Von Interesse war die Variation in der Überlebensrate, biotischen und abiotischen Schäden

sowie dem Höhen- und Durchmesserzuwachs. Die Ergebnisse zeigen genetische Unterschiede im Austriebszeitpunkt sowie der herbstlichen Laubverfärbung zwischen den untersuchten Herkünften. Aufgrund der allgemein besseren Wuchsleistung europäischer- gegenüber nordamerikanischen Herkünften halten die Autoren zusätzliche Saatgutimporte aus dem natürlichen Verbreitungsgebiet nur zur Prävention von genetischer Verarmung für sinnvoll (Liesebach und Schneck 2011).

Im Jahr 2017 veröffentlichten Merceron et al. (2017) eine weitere Studie zum Thema der Herkunft europäischer Eichenarten, in denen sie einen Großteil der bereits von Magni Diaz (2004) verwendeten Populationen mit Hilfe von nuklearen SNP (Einzelnukleotid-Polymorphismus) Markern analysierten. Insgesamt untersuchten sie 883 Individuen aus 73 nordamerikanischen- (11 zusätzliche) und 38 europäischen Populationen (inklusive je einer Probe aus 4 deutschen Beständen). Im Gegensatz zu Magni Diaz (2004) verwendeten Merceron et al. (2017) ausschließlich 69 nukleare SNP Marker zur Untersuchung der genetischen Struktur und der Identifizierung von potentiellen Herkunftsgebieten. Die Clusteranalyse zeigt das Vorkommen von drei geografischen Clustern im natürlichen Verbreitungsgebiet (im Süden, im Norden sowie in der nördlichen Mitte und im Nord-Westen), während in Europa nur zwei von ihnen nachgewiesen werden konnten. Merceron et al. (2017) schließen daraus, dass der südliche genetische Cluster entweder nicht nach Europa eingeführt wurde oder bereits wieder verschwunden ist. Sie vermuten daher, dass europäische Bestände ihren genetischen Ursprung im nördlichen Teil des natürlichen Verbreitungsgebietes haben (Merceron et al. 2017).

Pettenkofer et al. (2019) untersuchten kürzlich insgesamt 432 Individuen aus 39 deutschen- und 8 nordamerikanischen Populationen. An fünf Chloroplasten-Mikrosatelliten-Genorten konnten über alle Bestände insgesamt 13 verschiedene Haplotypen identifiziert werden (Pettenkofer et al., 2019). Diese gruppieren in drei Hauptclustern um die Haplotypen A, B und G, welche keine (A-B; A-G) oder nur ein Allel teilen (B-G). Insgesamt zeigt sich ein uniformes Haplotypenmuster in deutschen Beständen, mit A als dominanten Haplotyp (>80 %). Auch in Nordamerika ist A der vorherrschende Haplotyp, allerdings nicht ganz so häufig (>50 %). Die nächsthäufigeren Haplotypen sind sowohl in Deutschland als auch in Nordamerika Haplotyp B (12 % bzw. 13,3 %) und C (6 % bzw. 12,6 %). Zwei weitere Besonderheiten ergeben sich aus den Häufigkeiten sowie der geografischen Verteilung der Haplotypen: In nordamerikanischen Populationen wurden Haplotypen eines dritten Clusters gefunden, welche

in Deutschland nicht nachgewiesen werden konnten (Haplotypen G, H und L). Zudem wurde Haplotyp B ausschließlich in Populationen im nördlichen Teil des Verbreitungsgebietes gefunden, was die Ergebnisse von Merceron et al. (2017) unterstützt (Pettenkofer et al. 2019).

Eine Hauptkoordinatenanalyse zeigte überdies die Gruppierung von deutschen Beständen, während nordamerikanische Bestände weit verstreut liegen (Pettenkofer et al. 2019). Hinsichtlich der gesamten Haplotypendiversität weisen die deutschen Bestände deutlich niedrigere Werte auf als die nordamerikanischen Referenzpopulationen (Pettenkofer et al. 2019). Auch die Haplotypendiversität innerhalb von Populationen zeigt deutliche Unterschiede: Sie ist in Nordamerika geringer als in Deutschland. In der Folge ist die genetische Differenzierung zwischen Populationen in Nordamerika im Gegensatz zu Deutschland ausgesprochen hoch (Pettenkofer et al. 2019). Baden-Württemberg bildet eine Ausnahme in den deutschen Beständen: Hier ist die Haplotypendiversität (auch innerhalb der Populationen) hoch. Pettenkofer et al. (2019) erklären dies ebenfalls mit einer mehrfachen Einführung von Saatgut sowie einer Mischung von Material von verschiedenen Herkünften.

Eine aktuelle Studie beschäftigt sich mit der genetischen Variation deutscher Roteichenpopulationen ohne den Anspruch einer Herkunftsbestimmung: Hierbei wurde eine nähere Auswahl an 12 deutschen und 4 nordamerikanischen Populationen an 20 nuklearen Mikrosatelliten Genorten (davon 12 potentiell adaptive EST-SSRs) untersucht (Pettenkofer et al., 2020). Für die nordamerikanischen- ergaben sich gegenüber den deutschen Populationen signifikant höhere Werte für das Diversitätsmaß N_p (Anzahl privater Allele), F_{IS} (Inzuchtkoeffizient), sowie bei den Differenzierungsmaßen F_{ST} und Jost's D. Hinsichtlich der Anzahl privater Allele wurden in 12 deutschen Populationen insgesamt 9 private Allele gefunden (davon 5 in Baden-Württemberg). Hingegen konnten in nur 4 nordamerikanischen Populationen immerhin 19 private Allele identifiziert werden.

Alle anderen Maße (N_a, N_e, H_o, H_e) zeigten keine signifikanten Unterschiede zwischen den beiden Gruppen (Pettenkofer et al., 2020). Auch innerhalb der Gruppen zeigt kein Diversitätsmaß signifikante Unterschiede zwischen den Populationen. Wie bei der Analyse der Chloroplasten-Mikrosatelliten wurde auch für die nuklearen Genorte eine Hauptkoordinatenanalyse durchgeführt. Diese zeigt ebenfalls eine Gruppierung deutscher Bestände auf der Grundlage genetischer Ähnlichkeiten bei gleichzeitiger Zerstreuung nordamerikanischer Populationen (Pettenkofer et al., 2020).

Es konnten keine signifikanten Flaschenhalseffekte nachgewiesen werden. Ein Mikrosatellit (FIR13) im kodierenden Bereich des Gens *CONSTANS-like* konnte als potenziell selektiver Marker (Outlier-Locus) identifiziert werden. Eine nähere Betrachtung der Allelhäufigkeiten zeigt, dass dies auf die nordamerikanische Population BR1 zurückzuführen ist, die eine deutliche Abweichung der erwarteten- von der beobachteten Allelverteilung aufweist. FIR013 ist bereits von Lind-Riehl et al. (2014) als Outlier-Locus identifiziert worden.

2.3 Neue Marker für die Roteiche

Bis heute sind zahlreiche molekulare Marker speziell für *Quercus rubra* entwickelt oder von nahe verwandten Arten übertragen worden. Dazu gehören neben neutralen und potentiell adaptiven nuklearen Mikrosatelliten (z.B. Aldrich et al. 2002; Sullivan et al. 2013; Müller und Gailing 2018; Pettenkofer et al., 2020) auch SNPs (Konar et al. 2017; Merceron et al. 2017) sowie Mikrosatelliten und PCR-RFLPs im Chloroplastengenom der Roteiche (z.B. Romero-Severson et al. 2003; Magni et al. 2005; Zhang et al. 2015; Pettenkofer et al. 2019; Pettenkofer et al. eingereicht).

Zukünftige Studien könnten zusätzlich zu den bewährten Chloroplasten-Mikrosatelliten auf kürzlich entwickelte CAPS (engl. Cleaved Amplified Polymorphic Sequence) Marker zurückgreifen. Pettenkofer et al. (eingereicht) sequenzierten hierfür die Chloroplasten-Genome von 8 Proben, welche vier verschiedene Regionen innerhalb des natürlichen Verbreitungsgebiets repräsentieren (Pettenkofer et al., eingereicht). Mit Ausnahme des zweiten inverted Repeat konnten je nach Probe 87 % bis 97 % des Plastoms sequenziert werden. Dabei konnte eine außerordentlich hohe Sequenzierungstiefe von 4.030X–6.297X erreicht werden. Für die korrekte Anordnung der Fragmente wurde eine bereits öffentlich verfügbare Sequenz des Plastoms von *Quercus rubra* verwendet (Alexander und Woeste 2014). Insgesamt konnten 118 Einzelnukleotid-Polymorphismen (SNPs) sowie 107 Insertionen und Deletionen identifiziert werden, woraus speziell für *Quercus rubra* 15 CAPS Marker entwickelt wurden. Angewendet auf 96 Individuen aus 19 nordamerikanischen Populationen ließ sich in Kombination mit den von Pettenkofer et al. (2019) beschriebenen fünf Chloroplasten-Mikrosatelliten eine höhere Auflösung in der Identifizierung von Haplotypen erreichen. Die Kombination der beiden Markersysteme lässt in den 19 untersuchten natürlichen Beständen, in denen mit Chloroplasten-Mikrosatelliten lediglich 2 Haplotypen gefunden wurden, die Erkennung von 5 kombinierten Haplotypen zu. Insgesamt konnten 10 kombinierte Haplotypen

identifiziert werden. Die entwickelten CAPS-Marker sind darüber hinaus auch auf die eng verwandte Art *Q. ellipsoidalis* anwendbar.

3 Schlussfolgerung und Ausblick

Der Anbau der Roteiche in Trupps oder Einzelbäumen in Mischbeständen wird sicherlich auch in Zukunft noch ein Thema in Deutschland sein. Steigende mittlere Jahrestemperaturen und häufiger auftretende Trockenperioden infolge des Klimawandels machen eine Neubewertung des Anbaupotentials nichtheimischer Baumarten notwendig (Kölling 2013). In dieser Hinsicht kann die Roteiche, nicht zuletzt durch ihre breite Standortamplitude, als Nebenbaumart zur Stabilität in Mischbeständen beitragen und das vorhandene Baumartenspektrum bereichern (Klemmt et al. 2013).

Die Entwicklung von neuen molekularen Markern gestattet ein immer deutlicheres Bild von der genetischen Variation von eingeführten und natürlichen Roteichenpopulationen. Beispielsweise ermöglichen kürzlich entwickelte Chloroplasten-DNA Marker eine relativ kostengünstige und schnelle Untersuchung der Haplotypenvariation (Pettenkofer et al., 2019). Diese kann bei der Identifizierung nacheiszeitlicher Rückwanderungsbewegungen (auch eng verwandter Arten) verwendet werden, aber auch Hinweise über die Herkunft eingeführter Bestände geben.

Europäische Roteichenbestände weisen im Allgemeinen eine geringe genetische Differenzierung bei gleichzeitig ausreichend hoher genetischer Variation auf (Magni Diaz 2004; Pettenkofer et al., 2019; Pettenkofer et al., 2020). Dies weist darauf hin, dass eine mehrfache Einführung und Durchmischung von Saatgut einer genetischen Verarmung in eingeführten Beständen entgegengewirkt hat. Eine erhöhte Variation vornehmlich im Südwesten Deutschlands (Pettenkofer et al., 2019) weist darauf hin, dass nicht alle europäischen Bestände auf die gleiche Weise etabliert wurden. Vielmehr könnten jüngere Bestände aus dem Südwesten Deutschlands mit Material aus demselben Anbaugebiet (u.a. aus Beständen in Elsass-Lothringen) begründet worden sein, während Bestände weiterer deutscher Anbauregionen auf Material aus anderen Quellen zurückgehen. Allerdings ist es bisher keiner Studie gelungen eine Aussage über das genaue Herkunftsgebiet eingeführter Roteichenbestände zu machen. Die Ergebnisse von Merceron et al. (2017) und Pettenkofer et al. (2019) zeigen jedoch, dass eingeführte Bestände möglicherweise aus dem Norden des natürlichen Verbreitungsgebiets stammen.

Im Allgemeinen weisen deutsche Roteichenbestände neben einer guten Wuchsleistung auch eine hohe genetische Variation auf (Liesebach und Schneck 2011; Pettenkofer et al., 2020). Der wahrscheinliche Ursprung eingeführter Bestände aus dem Norden des natürlichen Verbreitungsgebiets könnte das Anpassungspotential unter zukünftigen Klimabedingungen einschränken. Um die Herkunft europäischer bzw. deutscher Roteichenbestände weiter einzuengen, müsste eine weitere Beprobung natürlicher Bestände in Nordamerika durchgeführt werden. Diese sollten sowohl mit den kürzlich entwickelten CAPS-, als auch mit etablierten cpSSR-Markern analysiert werden (Pettenkofer et al. 2019; Pettenkofer et al., eingereicht). Auch der cpSSR Marker D2T2 (Dumolin-Lapegue et al. 1999; Gerwein und Kesseli 2006) hat in der Vergangenheit Differenzierung gezeigt und sollte berücksichtigt werden.

Die Identifizierung von neuen Kandidatengenen (bspw. für die Artunterscheidung oder die Anpassung an bestimmte Umwelteinflüsse) kann u.a. zur genetischen Unterscheidung von eng verwandten Roteichen-Arten genutzt werden (Lind und Gailing 2013; Lind-Riehl et al. 2014). Moderne Hochdurchsatz-Sequenzierungsverfahren wie RADseq (engl. restriction site associated DNA sequencing) können die Identifizierung und Genotypisierung einer großen Anzahl von SNPs realisieren und auf diese Weise eine neue genomweite Ressource für potentiell selektive Marker bieten (Miller et al. 2007; Davey und Blaxter 2010). Auch eine lokale Eingrenzung der Herkunft eingeführter Bestände kann hierdurch ermöglicht werden, falls die verschiedenen Regionen in Nordamerika überhaupt charakteristische Differenzierungsmuster zeigen. Zudem sind weitere Analysen von Kandidatengenen für lokale Anpassung notwendig, um adaptive Unterschiede zwischen deutschen und nordamerikanischen Herkünften zu identifizieren. Darüber hinaus könnten solche Marker auch Informationen liefern, die im Rahmen von Züchtungsprogrammen, insbesondere aber zum Aufbau von Samenplantagen, genutzt werden könnten.

4 Zusammenfassung

Obwohl die Roteiche die flächenmäßig bedeutendste nichtheimische Laubbaumart in Deutschland ist, befassten sich -selbst auf europäischer Ebene- in der Vergangenheit nur wenige Studien mit der genetischen Variation oder Herkunft dieser eingeführten Art. Doch die Roteiche kann, als Trupp oder einzelbaumweise, in Mischbeständen zur Stabilität beitragen und die Massenleistung steigern. Das Ziel dieser Arbeit ist eine zusammenfassende Betrachtung der vorhandenen Arbeiten zur Herkunft und genetischen Vielfalt deutscher bzw. europäischer Roteichenpopulationen im Vergleich zu Populationen aus dem natürlichen

Verbreitungsgebiet. Sie soll relevante Erkenntnisse zusammenfassen und eine Grundlage für zukünftige Untersuchungen zu dieser Baumart bieten.

Gemessen an ihrer ökonomischen Relevanz wurden nur wenig Studien zur genetischen Variation der Roteiche in ihrem natürlichen Verbreitungsgebiet durchgeführt. Verschiedene Studien hatten aufgrund des Mangels an geografischen Barrieren und voneinander isolierten glazialen Refugien bislang Schwierigkeiten genaue nacheiszeitliche Rückwanderungsrouten zu identifizieren.

Die Gründe für die allgemein sehr geringe genetische Differenzierung und vergleichsweise hohe genetische Variation europäischer Bestände an Markern aus dem Kerngenom liegen möglicherweise darin, dass eine mehrfache Einführung und Durchmischung von Saatgut einer genetischen Verarmung in eingeführten Beständen entgegengewirkt hat. Die höhere genetische Variation in Beständen im Südwesten Deutschlands weist darauf hin, dass diese mit Material aus dem historisch deutschen Anbaugebiet (u.a. aus Beständen im Elsass-Lothringen) begründet worden sein könnten, während weitere deutsche Bestände auf Material aus anderen Quellen zurückgehen. Übereinstimmende Ergebnisse weisen darauf hin, dass eingeführte Bestände aus dem Norden des natürlichen Verbreitungsgebiets stammen.

5 Summary

Studies of the genetic variation and origin of northern red oak (*Quercus rubra* L.) in natural and introduced populations

Although Northern red oak (*Quercus rubra* L.) is, in terms of surface area, the most important non-native deciduous tree species in Germany, only few studies were conducted on the genetic variation and origin of this introduced species, even at the European level. However, Northern red oak can, in groups or as single trees, contribute to a better stability and growth performance in mixed stands. The objective of this work is to give a review on existing studies that cover the origin and genetic variation of German resp. European red oak populations in comparison to populations from the natural range. It outlines relevant findings and provides a basis for future studies on this tree species.

In the light of its economic relevance in North America, there are only few studies on the genetic variation of *Q. rubra* in its natural range. Until now, various studies had difficulties to identify particular postglacial migration routes due to the lack of geographic barriers and isolated glacial refugia. A possible consequence is the (in comparison to other oak species)

lower genetic differentiation of northern red oak in its natural range which increase towards the north.

In general, European red oak stands show sufficiently high genetic variation with at the same time only low genetic differentiation. Possible reasons may be, that multiple introductions and admixture of seeding material counteract genetic erosion in introduced populations. In one of the past studies, differences were found between historic German and historic French populations. It is suggested that differences in the import- and forest policies of the two countries may be the reason. The higher genetic variation in populations in the southwest of Germany indicates, that these stands may have been founded with material from the historic German populations (including populations in Alsace-Lorraine), while other German stands may have been founded with material of other sources. Until today, none of the reviewed studies was able to identify a particular region of origin of introduced red oak populations. However, all studies suggest, that introduced stands originate from the northern part of the natural distribution range. Furthermore, due to their sufficiently high genetic variation, the import of additional seeding material from the natural range is unnecessary in order to preserve the adaptive capacity of German red oak stands, at least under current climatic conditions.

Danksagung: Wir danken Markus Müller für die sorgfältige Überprüfung des Manuskripts. Bei dem Bundesministerium für Ernährung und Landwirtschaft (BMEL) bedanken wir uns für die finanzielle Förderung (Förderkennzeichen 22023314)

6 Literaturverzeichnis

Aldrich P, Michler C, Sun W et al (2002) Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). Molecular Ecology Notes:472–474

Alexander LW, Woeste KE (2014) Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. Tree Genetics & Genomes 10:803–812

Barrett SC, Husband BC (1990) The genetics of plant migration and colonization. In: Brown HD, Clegg MT, Kahler AL et al (eds) Plant population genetics, breeding, and genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, pp 254–277

Bauer F (1951) Die Roteiche 1950. Eine ertragskundlich-biologische und holzkundliche Untersuchung. Dissertation, Universität Göttingen

Bauer F (1954) Zur Rassenfrage der Roteiche. Allgemeine Forstzeitschrift 9:470–474

Birchenko I, Feng Y, Romero-Severson J (2009) Biogeographical distribution of chloroplast diversity in northern red oak (*Quercus rubra* L.). The American Midland Naturalist 161:134–145

Borkowski DS, Hoban SM, Chatwin W et al (2017) Rangewide population differentiation and population substructure in *Quercus rubra* L. Tree Genetics & Genomes 13:472

Branquart E, Vanderhoeven S, van Landuyt W et al (2007) Harmonia database: *Quercus rubra*. Available via http://ias.biodiversity.be/species/show/87. Accessed 26 Feb 2018

Bundesministerium für Ernährung und Landwirtschaft (BMEL) (2014) Der Wald in Deutschland. Ausgewählte Ergebnisse der dritten Bundeswaldinventur. Available via https://www.bundeswaldinventur.de/fileadmin/SITE_MASTER/content/Dokumente/Downloads/BMEL_Wald_Broschuere.pdf. Accessed 20 Mrz 2018

Chmura D (2013) Impact of alien tree species *Quercus rubra* L. on understorey environment and flora: A study of the silesian upland (southern Poland). Polish Journal of Ecology 61:431–442

Crow TR (1988) Reproductive mode and mechanisms for self-replacement of Northern red oak (*Quercus rubra*). A review. Forest Science 34:19–40

Daubree JB, Kremer A (1993) Genetic and phenological differentiation between introduced and natural populations of *Quercus rubra* L. Ann. For. Sci. 50:271s-280s

Davey JW, Blaxter ML (2010) RADSeq. Next-generation population genetics. Briefings in functional genomics 9:416–423

Dengler A (1944) Waldbau auf ökologischer Grundlage. Ein Lehr- und Handbuch. Springer-Verlag, Berlin, Heidelberg

Dey DC, Parker WC (1996) Regeneration of red oak (*Quercus rubra* L.) using shelterwood systems. Ecophysiology, silviculture and management recommendations. Forest Research Information Paper

Dlugosch KM, Parker IM (2008) Founding events in species invasions. Genetic variation, adaptive evolution, and the role of multiple introductions. Molecular Ecology 17:431–449

Dreßel R, Jäger EJ (2002) Beiträge zur Biologie der Gefäßpflanzen des herzynischen Raumes. 5. *Quercus rubra* L. (Roteiche): Lebensgeschichte und agriophytische Ausbreitung im Nationalpark Sächsische Schweiz. Hercynia:37–64

Dumolin-Lapegue S, Pemonge M-H, Gielly L et al (1999) Amplification of oak DNA from ancient and modern wood. Molecular Ecology 8:2137–2140

Gailing O, Lind J, Lilleskov E (2012) Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* E.J. Hill populations in contrasting environments. Plant Syst Evol 298:1533–1545

Gerwein JB, Kesseli R v. (2006) Genetic diversity and population structure of *Quercus rubra* (Fagaceae) in old-growth and secondary forests in southern New England. Rhodora 108:1–18

Goeze E (1916) Liste der seit dem 16. Jahrhundert eingeführten Bäume und Sträucher. Mitteilungen der Deutschen Dendrologischen Gesellschaft:129–201

Graw J (2005) Genetik, 4th edn. Springer, Berlin

Hickel R (1932) Dendrologie forestière. Lechevalier, Paris

Klemmt H-J, Neubert M, Falk W (2013) Das Wachstum der Roteiche im Vergleich zu den einheimischen Eichen. LWF aktuell:28–31

Kölling C (2013) Nichtheimische Baumarten - Alternativen im klimagerechten Waldumbau? LWF aktuell:4–11

Konar A, Choudhury O, Bullis R et al (2017) High-quality genetic mapping with ddRADseq in the non-model tree *Quercus rubra*. BMC Genomics 18:417

Kriebel HB, Bagley WT, Deneke FJ et al (1976) Geographic variation in *Quercus rubra* in North Central United States plantations. Silvae Genetica 25:118–122

Kriebel HB, Merritt C, Stadt T (1988) Genetics of growth rate in *Quercus rubra*: provenance and family effects by the early third decade in the North Central USA. Silvae Genetica 37:193–198

Kuehne C, Nosko P, Horwath T et al (2014) A comparative study of physiological and morphological seedling traits associated with shade tolerance in introduced red oak (*Quercus rubra*) and native hardwood tree species in southwestern Germany. Tree Physiology 34:184–193

Liesebach M, Schneck V (2011) Entwicklung von amerikanischen und europäischen Herkünften der Roteiche in Deutschland. Forstarchiv 82:125–133

Lind JF, Gailing O (2013) Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genetics & Genomes 9:707–722

Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a CONSTANS-like gene between two red oak species. Annals of Botany 113:967–975

Magni CR, Ducousso A, Caron H et al (2005) Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. Molecular Ecology 14:513–524

Magni Diaz CR (2004) Reconstitution de l'introduction de *Quercus rubra* L. en Europe et conséquences génétiques dans les populations allochtones. Dissertation, Ècole Nationale du Génie Rural, des Eaux et des Forêts (in French)

Merceron NR, Leroy T, Chancerel E et al (2017) Back to America. Tracking the origin of European introduced populations of *Quercus rubra* L. Genome 60:778–790

Miller MR, Dunham JP, Amores A et al (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome research 17:240–248

Möllerová J (2005) Notes on invasive and expansive trees and shrubs. Journal of Forest Science 51:19–23

Müller M, Gailing O (2018) Characterization of 20 new EST-SSR markers for northern red oak (*Quercus rubra* L.) and their transferability to *Fagus sylvatica* L. and six oak species of section *Lobatae* and *Quercus*. Annals of Forest Research 61:211–222

Nagel R-V (2015) Roteiche (*Quercus rubra* L.). In: Vor T, Spellmann H, Bolte A et al (eds) Potenziale und Risiken eingeführter Baumarten. Baumartenportraits mit naturschutzfachlicher Bewertung. Univ.-Verl. Göttingen, Göttingen, pp 219–267

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Nicolescu V-N, Vor T, Mason WL et al (2018) Ecology and management of northern red oak (*Quercus rubra* L. syn. *Q. borealis* F. Michx.) in Europe: a review. Forestry: An International Journal of Forest Research 59:249

Niinemets Ü, Valladares F (2006) Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. Ecological Monographs 76:521–547

Petit RJ, Kremer A, Wagner DB (1993) Geographic structure of chloroplast DNA polymorphisms in European oaks. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 87:122–128

Pettenkofer T, Burkardt K, Ammer C et al (2019) Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison with reference native North American populations. European Journal of Forest Research 138:275-285

Pettenkofer T, Finkeldey R, Müller M et al (eingereicht) Development of 15 novel CAPS markers based on the sequenced chloroplast genome of *Quercus rubra.*

Pettenkofer T, Finkeldey R, Müller M et al (2020) Genetic variation of introduced red oak (*Quercus rubra*) stands in Germany in comparison to North American populations. European Journal of Forest Research (doi:10.1007/s10342-019-01256-5)

Riepšas E, Straigytė L (2008) Invasiveness and ecological effects of red oak (*Quercus rubra* L.) in Lithuanian Forests. Baltic Forestry 14:122–130

Romero-Severson J, Aldrich P, Feng Y et al (2003) Chloroplast DNA variation of northern red oak (*Quercus rubra* L.) in Indiana. New Forests 26:43–49

Schlarbaum SE, Adams RP, Bagley WT et al (1982) Postglacial migration pathways of *Quercus rubra* L., northern red oak, as indicated by regional genetic variation patterns. Silvae Genetica 31:150–158

Sork VL, Huang S, Wiener E (1993) Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. Ann. For. Sci. 50:261s-270s

Stratmann J, Warth H (1987) Die Roteiche als Alternative zu Eiche oder Buche in Nordwestdeutschland. Allgemeine Forstzeitschrift 42:40–41

Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. Molecular Ecology 17:351–360

Sullivan AR, Lind JF, McCleary TS et al (2013) Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. Plant Mol Biol Rep 31:231–239

Vor T, Lüpke Bv (2004) Das Wachstum von Roteiche, Traubeneiche und Rotbuche unter verschiedenen Lichtbedingungen in den ersten beiden Jahren nach der Pflanzung. Forstarchiv 75:13–19

Zhang R, Hipp AL, Gailing O (2015) Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations. Botany 93:691–700

Paper 2

Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison with reference native North American populations

Published

European Journal of Forest Research

16 January 2019

doi:10.1007/s10342-019-01167-5

Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison with reference native North American populations

Tim Pettenkofer[1] · Katharina Burkardt[2] · Christian Ammer[2] · Torsten Vor[2] · Reiner Finkeldey[3] · Markus Müller[1] · Konstantin Krutovsky[1,4,5,6,8] · Barbara Vornam[1] · Ludger Leinemann[1] · Oliver Gailing[1,7,8]

Corresponding authors: Oliver Gailing[1], E-mail: ogailin@gwdg.de; Ludger Leinemann[1], E-mail: lleinem@gwdg.de, Tel: +49 551 3933536

[1] University of Goettingen, Faculty of Forest Sciences and Forest Ecology, Forest Genetics and Forest Tree Breeding, Büsgenweg 2, 37077 Göttingen, Germany

[2] University of Goettingen, Faculty of Forest Sciences and Forest Ecology, Silviculture and Forest Ecology of the Temperate Zones, Büsgenweg 1, 37077 Göttingen, Germany

[3] University of Kassel, Mönchebergstraße 19, 34125 Kassel, Germany

[4] Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 660036 Krasnoyarsk, Russia

[5] Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119333, Russia

[6] Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, USA

[7] Michigan Technological University, School of Forest Resources and Environmental Science, Houghton, MI 49931, United States

Keywords: chloroplast microsatellites, cpSSRs, *Quercus rubra*, haplotype diversity, origin, provenances

Abstract

Northern red oak (*Quercus rubra*) was introduced to Europe in the late seventeenth century and has since become the most important deciduous non-native tree species in Germany. Despite its importance, little is known about the origin and patterns of genetic variation in German red oak stands. To be able to make recommendations regarding the adaptive potential of red oak stands, which might be related to their origin and the selection of provenances, with respect to climate change, a better understanding of the genetic diversity and structure of German red oak stands is needed. Individuals from 62 populations in Germany and North America were genotyped at five chloroplast microsatellite loci to characterize chloroplast haplotype diversity and geographic structure. Compared to reference populations from the natural distribution range, German red oak stands demonstrated a relatively low genetic differentiation among populations and represented only a fraction of the haplotype diversity found in North America. For several stands located in southern Germany, considerably higher haplotype diversity than in other German stands was found. While most German stands showed signatures of founder effects, the diversity of stands in southern Germany might have been increased due to admixture and multiple introductions of different North American provenances. Overall, we conclude that German stands originated from a limited geographic area, possibly located in the northern part of the native distribution range.

1 Introduction

Since the rediscovery of the Americas in the end of the fifteenth century, thriving trade routes connecting the ports of the old and new worlds also transported alien plants and animals. While numbers of established alien species were relatively low until 1800, introductions increased significantly at the turn of the century (Hulme 2009). Introduced populations of non-native species often contained only a fraction of the genetic information compared to the natural range they originated from (Nei et al. 1975; Barrett and Husband 1990). This founder effect, thus, could result in the establishment of new populations with their own specific gene pool. Founder populations are usually small and hence more affected by random drift (Graw 2005). As a consequence, the adaptive potential of founder populations could be limited. However, multiple introductions seem to increase diversity over a longer period (Dlugosch and Parker 2008). Suarez and Tsutsui (2008) even suggested that multiple introductions can enhance the adaptive potential by providing a source of variation important for adaptation, which can be critical for successful establishment and spreading. A different situation arises when seeding

material originates from a limited geographic area within a species' natural range. Limited sampling induces a genetic bottleneck, whereby the newly founded population does not represent the full range of the species' variation.

Northern red oak was first introduced to Europe at the end of the seventeenth century, planted for ornamental reasons in parks and botanical gardens until the middle of the eighteenth century (Bauer 1951; Nagel 2015). Today, it is the most important foreign deciduous tree species for wood production in Germany (Bundesministerium für Ernährung und Landwirtschaft (BMEL), 2014). Mainly, this is due to a shorter rotation period of only 80–120 years compared to more than 140 years in the native oak species, combined with lower water and nutrient availability (Nagel 2015). In some European countries, *Q. rubra* is considered invasive (Möllerová 2005; Riepšas and Straigytė 2008; Chmura 2013). However, this is not the case in Germany, because it is less shade tolerant than the main tree species *Fagus sylvatica* and only little more shade tolerant than the native *Q. robur* (Vor and Lüpke 2004; Niinemets and Valladares 2006; Nagel 2015). Furthermore, it is subject to heavy browsing and can be easily controlled by tending measures (Vor 2005; Nagel 2015). Natural regeneration is not the most important way of establishing red oak in Germany. Most stands are founded by planting or seeding, formerly in pure stands, now as groups in mixed stands (Nagel 2015). *Q. rubra* belongs to section *Lobatae*, which is restricted to North America, and it does not hybridize with native white oak species of the section *Quercus* (Magni Diaz 2004; Nagel 2015).

Nagel (2015) characterized *Q. rubra* as a species growing in a wide range of climatic (annual precipitation between 600 and 2000 mm, average temperature between 4 and 15 °C) and soil conditions. Even outside its natural range, *Q. rubra* shows good to satisfactory growth despite a level of annual precipitation that, in some cases in Germany, does not reach the minimum requirement for red oak (Dreßel and Jäger 2002; Magni Diaz 2004). Thus, under climate change conditions it can be expected that *Q. rubra* will be able to increase forest adaptability and productivity in particular sites in Germany, especially in mixed forest communities, if suitable provenances are chosen. In order to evaluate the adaptive potential of German red oak stands, a detailed understanding of the geographic origin of the provenances used for the establishment of German red oak stands is required. This is especially important for stands that will be part of tree improvement strategies and serve as seed orchards.

In this study, chloroplast microsatellite (simple sequence repeat or cpSSR) markers were used to analyse haplotype diversity. Due to maternal inheritance of the chloroplast organelle in

oaks and lack of recombination, cpDNA haplotypes usually have relatively low variation within oak populations (Zhang et al. 2015) and are more affected than nuclear DNA by stochastic processes, such as genetic drift and founder events (Alexander and Woeste 2014). Haplotype differentiation within one species can be high across both populations and regions, and significant geographic variation can be found even within mixed populations of hybridizing species (Petit et al. 1993; Grivet et al. 2006; López de Heredia et al. 2007; Zhang et al. 2015). Chloroplast DNA markers have been successfully used to trace post-glacial recolonization routes of white oak and other species in Europe (Petit et al. 2002; Palmé et al. 2003; Heuertz et al. 2004) and to reveal the haplotype composition of autochthonous populations (Gailing et al. 2009). Identified haplotypes and haplotype variation can be compared with reference populations within the natural range to infer their origin.

While the analysis of genetic variation patterns in northern red oak populations within their natural range was subject of several past studies (e.g. Daubree and Kremer 1993; Romero-Severson et al. 2003; Magni et al. 2005; Zhang et al. 2015; Borkowski et al. 2017), it has rarely been studied in Europe (Magni Diaz 2004; Merceron et al. 2017). It was found that within the natural range of red oak, genetic differentiation increased northwards (Borkowski et al. 2017). In contrast, no geographic pattern was found in Europe (Magni Diaz 2004), where most of the populations were established in France and Germany and a few others also in the Netherlands, Belgium, Spain, Italy and Romania. This observation was explained by multiple introductions and admixture of material within Europe (Magni Diaz 2004), a suggestion also partly shared by Daubree and Kremer (1993). While identifying the region of origin of the introduced populations was one of the objectives in the Magni Diaz's (2004) study, it was not succeeded due to the absence of a clear phylogeographic structure within the natural range at a smaller scale. Merceron et al. (2017) studied SNP markers in both North American and European red oak populations with focus on France. Samples from Germany, the Netherlands, Belgium, Spain, Italy and Romania were also included in this study, but in smaller quantities. Merceron et al. (2017) found three main genetic clusters in the red oaks' natural range, with only two of these clusters observed in Europe. This likely suggests that European populations originated from the northern parts of the natural range. These findings are in line with other studies on this topic (Bauer 1954; Magni Diaz 2004; Nagel 2015). Furthermore, Merceron et al. (2017) reported a continuous, predominantly latitudinal gradient and only a weak phylogeographic structure in the natural range. One explanation could be that, unlike the situation of European white oaks during post-glacial recolonization, northern red oak did not survive the glacial

period in separated refugia, but instead in a wide longitudinal range in the south of its natural range (Petit et al. 2002; Magni Diaz 2004).

This study aims to assess how the genetic variation in German red oak stands differs from variation found within the species' natural range. CpSSR markers have been used in our study for the first time to analyse and compare chloroplast haplotype diversity and structure in *Q. rubra* plantations across both German stands and North American reference populations. Our objectives were to assess (i) to what extent the haplotype diversity found in the natural range is represented in Germany and (ii) to what extent haplotype diversity varies between the different regions within Germany.

We hypothesized that (1) the introduction of *Q. rubra* to Germany resulted in reduced overall haplotype diversity due to bottleneck effects and (2) the introduction of seed material from unknown sources led to an artificial differentiation of haplotypes across Germany.

2 Materials and methods

2.1 Plant material

Buds or green leaves were collected from 432 trees in total: 385 trees representing 39 stands in Germany of unknown origin (8–10 samples per stand, Supplementary Table 1S and Fig. 1S) plus material from 47 trees of 8 populations of known North American origin (5–7 samples per population, Supplementary Fig. 2S) from a provenance trial in Northern Germany (Supplementary Table 2S) (Liesebach and Schneck 2011). In addition, published data from 8 natural populations (8 samples per population) as well as unpublished data from 3 natural populations from the northern distribution range of the species were included (Supplementary Table 3S and Fig. 2S) (Lind and Gailing 2013; Lind-Riehl et al. 2014). Samples were taken randomly within each population. All North American populations served as reference having known origin, unlike German stands.

To cover a wide geographic range, sample stands were selected in 5 different federal states of Germany: 10 stands in Lower Saxony, 10 in North-Rhine Westphalia, 6 in Brandenburg, 7 in Thuringia and 6 in Baden-Wuerttemberg. Sample stands were chosen to match the following criteria: (1) they should be pure *Q. rubra* stands; (2) 50–80 year old; (3) feature a rectangular shape for easier data acquisition; and (4) should be in locations, where they will be cultivated in the future as well (according to the forest administration), as not only the present, but the expected future climatic conditions would match their autecological properties.

2.2 DNA isolation

DNA was extracted from about 1 cm^2 leaf tissue per tree with the DNeasy™ 96 Plant Kit from Qiagen (Hilden, Germany). Depending on the availability, we used either a small piece of the fresh leaf or 1–2 whole buds from a fresh twig.

2.3 Chloroplast microsatellites

Eight different universal cpSSR markers for angiosperms (ccmp1, ccmp2, ccmp3, ccmp4, ccmp5, ccmp6, ccmp7 and ccmp10) (Weising and Gardner 1999) and three cpSSRs developed for oaks (ucd4, udt1 and udt4) (Deguilloux et al. 2003) were tested for amplification and variation in 43 red oak samples from 11 different populations. Five of these cpSSR markers showed variation and were thus used in the study: two universal (ccmp2 and ccmp4) and three oak-specific markers (udt1, udt4 and ucd4, see Supplementary Table 4S for further information).

We used a touchdown PCR program for all markers in a Biometra TProfessional thermocycler (Jena, Germany). The PCR protocol started with 15 min for initial denaturation at 95 °C, followed by 10 cycles of 1 min denaturation at 94 °C, 1 min annealing at 60 °C (− 1 °C per cycle) and 1 min extension at 72 °C. This first set of cycles was then followed by another 25 cycles of 1 min denaturation at 94 °C, 1 min annealing at 50 °C and 1 min extension at 72 °C. The PCR ended with a final 20 min extension step.

For each single primer pair, PCRs were conducted in a 14 µl volume containing 1 µl of genomic DNA (about 0.6 ng/ µl), 6.8 µl ddH_2O, 1.5 µl PCR buffer (containing 0.8 M Tris–HCl and 0.2 M $(NH_4)_2SO_4$), 1.5 µl $MgCl_2$ (25 mM), 1 µl of each dNTP (2.5 mM), 1 µl primer (forward, 5 pM/µl), 1 µl primer (reverse, 5 pM/µl) and 1 U HOT FIREPol® Taq Polymerase from Solis BioDyne (Tartu, Estonia). For multiplexing of markers ccmp2 and ccmp4, we used the following PCR mix: 1 µl of genomic DNA (about 0.6 ng/µl), 4.8 µl ddH_2O, 1.5 µl PCR buffer (containing 0.8 M Tris–HCl and 0.2 M $(NH_4)_2SO_4$), 1.5 µl $MgCl_2$ (25 mM), 1 µl of each dNTP (2.5 mM), 2 × 1 µl primer (forward, 5 pM/µl), 2 × 1 µl primer (reverse, 5 pM/µl) and 1 U HOT FIREPol® Taq Polymerase (Solis BioDyne; Tartu, Estonia).

The PCR products were tested in an agarose gel electrophoresis to determine the ideal dilution ratio for the capillary electrophoresis. The gel electrophoresis was carried out at 90 v for 20 min in a 1.5 % agarose gel with TAE as a running buffer (1X working solution). The DNA was stained with Roti-Safe GelStain from Roth (Karlsruhe, Germany). Samples were diluted according to the intensity of amplification products on the gel.

For the separation of the cpDNA fragments, we performed a capillary electrophoresis on an ABI Prism Genetic Analyzer 3130xl (Applied Biosystems, Foster City, USA). The fragment sizes were scored using the GeneMapper software version 3.7 (Applied Biosystems, Foster City, USA).

2.4 Data analyses

Chloroplast DNA haplotypes were based on all genotyped cpSSR markers. The software PermutCpSSR (https://www6.bordeaux-aquitaine.inra.fr/biogeco/Production-scientifique/Logiciels/Contrib-Permut/Permut; Pons and Petit 1996; Burban et al. 1999) was used to determine the total haplotypic diversity H_T, the average expected within-population haplotypic diversity H_S and the genetic differentiation among populations G_{ST}. Assuming a stepwise mutational model, the software was also used to compute R_{ST} that takes into account allele size and is computed as the ratio of the variance between populations and the total variance of allele size in terms of number of repetitive motifs (Slatkin 1995). In contrast to G_{ST}, R_{ST} accounts better for the relatively low mutation rates occurring at chloroplast SSR markers. The haplotype network was generated using Arlequin version 3.5 (Excoffier and Lischer 2010). The software computes a matrix of pairwise distances between all haplotypes, using the sum of squared size differences.

The software package GENECLASS2 (Piry et al. 2004) was used to tentatively assign German red oak stands to North American reference populations. It uses a Bayesian method introduced by Rannala and Mountain (1997) to assign German stands to their three most probable reference populations.

Table 1 Description of *Quercus rubra* chloroplast haplotypes (A-O) and their frequencies (ƒ)

Haplotype	Five cpSSR markers and their allele fragment sizes, bp					Germany		Reference		Overall	
	ccmp2	*ccmp4*	*ucd4*	*udt1*	*udt4*	ƒ	ƒ, %	ƒ	ƒ, %	ƒ	ƒ, %
A	228	116	99	86	145	309	80.3	74	54.8	383	73.7
B	227	115	98	85	146	46	12.0	18	13.3	64	12.3
C	228	116	99	87	145	23	6.0	17	12.6	40	7.7
D	228	117	99	86	145	1	0.3			1	0.2
E	228	116	97	86	146	3	0.8			3	0.6
F	227	116	98	85	146	1	0.3			1	0.2
G	226	117	97	85	146			4	3.0	4	0.7
H	226	117	97	84	146			2	1.5	2	0.4
I	227	115	98	84	146			1	0.7	1	0.2
K	228	116	99	85	145			12	8.9	12	2.3
L	226	118	98	85	146			7	5.2	7	1.3
M	227	115	98	85	145	1	0.3			1	0.2
O	228	115	98	85	146	1	0.3			1	0.2

To analyse a possible correlation between genetic and geographic distances, Mantel tests were performed separately for German stands and North American reference populations. The correlation value R_{XY} ranges from −1 to 1. A positive correlation would imply that two populations become increasingly genetically distant the further away they are located from each other. A negative correlation would imply the contrary. First, pairwise genetic distances were calculated between all populations (Bruvo et al. 2004). To weigh them by the number of individuals in each population, the pairwise genetic distances between individual trees in each population pair were summed, and the sum was divided by the number of pairwise distances between the two populations. Secondly, a matrix of the geographic distance (in km) was computed. Lastly, the software package GenAlEx 6.5 (Peakall and Smouse 2006, 2012) was used to perform both a Mantel test (9999 permutations) and a Principle Coordinates Analysis (PCoA). The PCoA is a cluster analysis, which can visualize individual and/or population differences based on the genetic distance between them by assigning for each individual or population a location in a multidimensional space. In this study, the Bruvo distance between populations was used. The PCoA tries to find the main axes through a distance matrix that explain most of the genetic differentiations between individuals and/or populations. Usually, these are the first two or three main axes in a multidimensional space (Peakall and Smouse 2012). In our case, the first axis explained 68.10 % and the second axis 8.77 % (together 76.87 %) of the differentiation.

3 Results

3.1 Chloroplast haplotype distribution

In total, 13 different chloroplast haplotypes were found among all studied populations. Five of them (D, E, F, M and O) were found only in Germany, while 5 others (G, H, I, K and L) were found solely in North America (Table 1).

Most of these haplotypes were rare. Haplotype K (8.9 %) was found in the upper Midwest and southern Canada, and haplotype L (5.2 %) was found only in the Great Lakes region of North America. Haplotypes G and H were identified only in samples representing southern regions of the natural range. Unlike private haplotypes in North America, private haplotypes in German stands were found rarely. While haplotypes D and E are more similar to haplotype A, haplotypes F, M and O appear to be closely related to haplotype B.

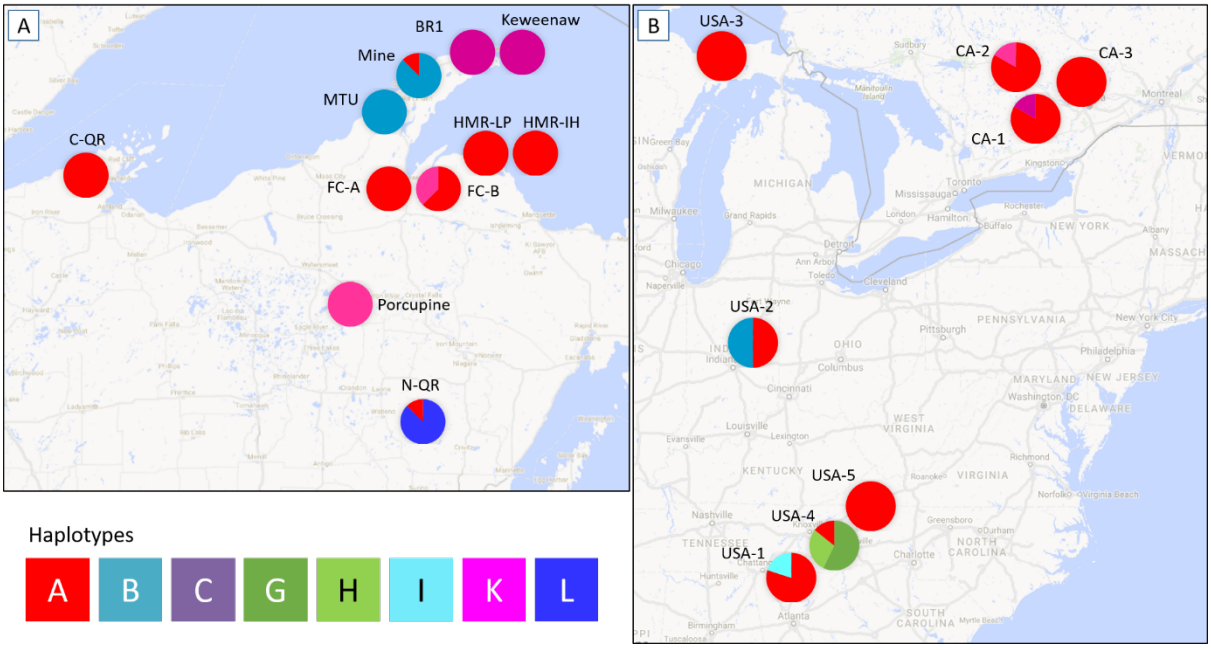


Figure 1 A: Chloroplast haplotype distribution south of the Great Lakes. Populations were sampled and genotyped in earlier studies at gSSRs (Lind and Gailing 2013; Lind-Riehl et al. 2014; Google Maps 2017b). Data from populations Keweenaw, Mine and Porcupine have not been published before. **B**: Distribution of *Quercus rubra* chloroplast haplotypes in North America. Samples were partly obtained from a provenance trial in Lübeck, Germany (Liesebach and Schneck 2011; Google Maps 2017c).

Among the common types, haplotype A was dominant in both North America and Germany, although a much higher relative frequency was found in Germany (80.3 %) compared to North America (54.8 %). Haplotype B was the next most frequent haplotype in Germany (12.0 %) with a similar frequency in North American samples (13.3 %). Interestingly, this haplotype is different from haplotype A in each of the five loci (Table 1). While haplotype

B was found in Germany in each of the five federal states, it was predominantly located in the centre and northern regions of the natural distribution range in North American samples (Fig. 1). Haplotype C, which is closely related to haplotype A, was found only in the northern regions of the natural range (on the tip of the Keweenaw Peninsula and in Ontario). In German stands, haplotype C was usually found at low frequency, only in single samples, except Baden-Wuerttemberg, where it was found at relatively high frequencies (57 %), although only 6 stands were sampled there (Fig. 2).

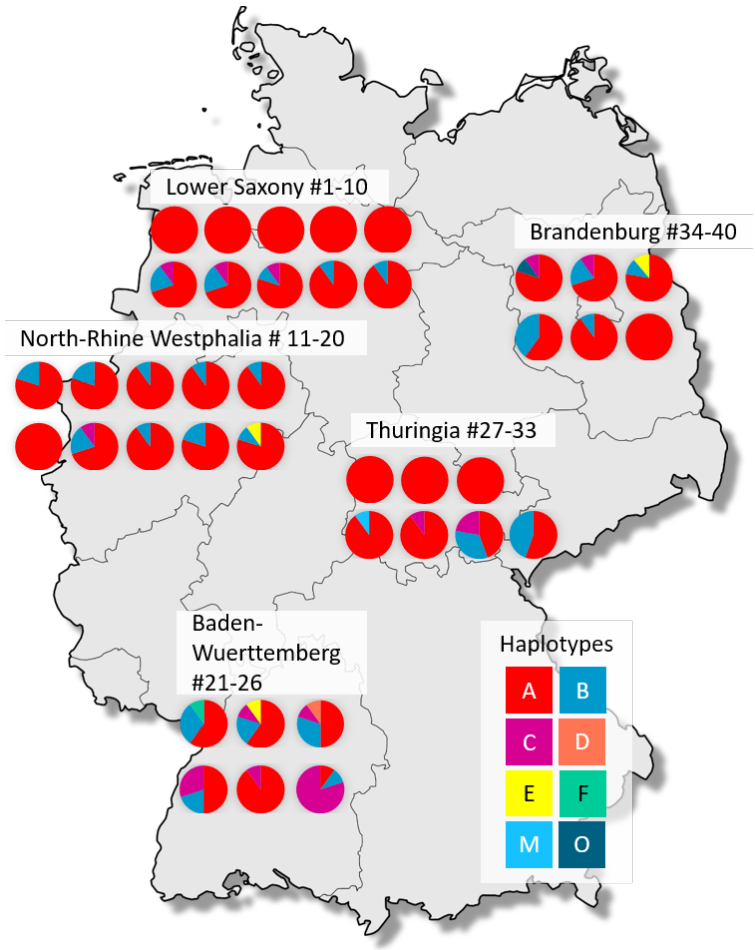


Fig. 2 Distribution of *Quercus rubra* chloroplast haplotypes in Germany (stand #36 of Brandenburg was removed from the study due to insufficient number of samples)

3.2 Haplotype diversity and structure

The minimum spanning tree in Fig. 3 shows the presence of three main lineages or clusters and their related haplotypes (A, B and H). In our samples representing North America, lineage A was found in all regions, but predominantly in the north and north-east; B was found in one population in the south, but also in the north-central and the north-west of the species' natural distribution; H was mainly found in the south. Haplotype A is different from B and H in all the five loci, whereas B and H have one locus in common.

Table 2 *Quercus rubra* chloroplast haplotype diversity within populations (H_S), total haplotype diversity (H_T), differentiation among populations (G_{ST}), the genetic differentiation based on the stepwise mutation model (R_{ST}), the number of haplotype (N_a), and the number of private haplotypes (N_p)

Data	H_S	H_T	G_{ST}	R_{ST}	N_a	N_p
Germany	0.291	0.337	0.137	0.047	8	5
Lower Saxony	0.180	0.187	0.036	-0.020	3	0
North-Rhine Westphalia	0.260	0.246	-0.056	-0.071	3	0
Baden-Wuerttemberg	0.537	0.655	0.181	0.097	6	2
Thuringia	0.240	0.313	0.234	0.190	4	1
Brandenburg	0.340	0.354	0.040	0.039	5	1
North America	0.177	0.654	0.729	0.772	8	5
North America (Great Lakes) (Zhang et al. 2015)	0.346	0.436	0.206	0.253		

Due to the fixation of the same one haplotype per site, haplotype diversity within North American populations was low (H_S = 0.177), despite total haplotype diversity being high (H_T = 0.652) accounting for a relatively high number of different haplotypes (Table 2). High genetic differentiation was found among all North American populations (G_{ST} = 0.729, R_{ST} = 0.772). In Germany, the average haplotype diversity within populations was higher (H_S = 0.291), especially in stands from Southwest Germany (H_S = 0.537).

In contrast to North American populations, markers differ mostly by only a single nucleotide motif among German stands. As a result, the R_{ST} value is smaller (i.e. demonstrating less differentiation) than the calculated G_{ST}.

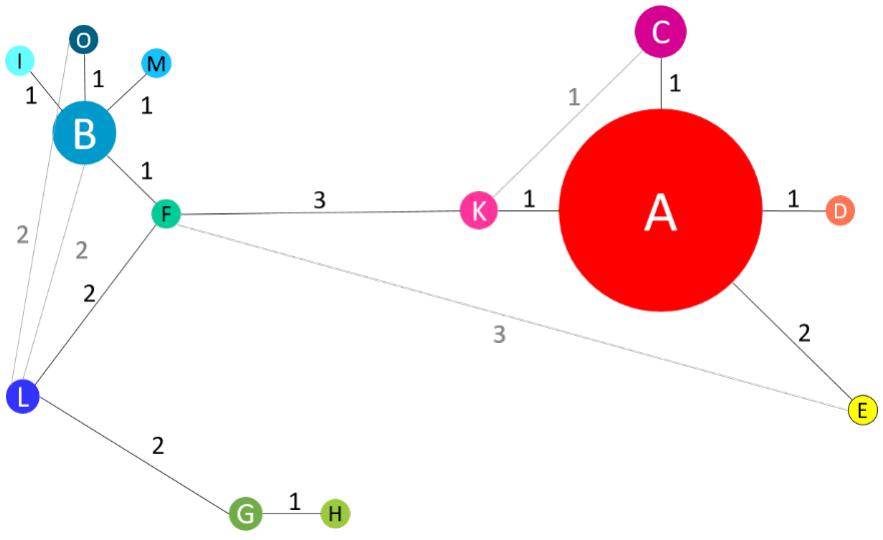


Fig. 3 Minimum spanning tree representing the chloroplast haplotype network (Excoffier and Lischer 2010) of *Quercus rubra* populations. Numbers next to the lines indicate the number of markers which differ between two haplotypes.

3.3 Relationship between German and North American populations

The PCoA showed that all German stands densely cluster with reference populations from most areas of North America (Fig. 4). Within this large group, there is further grouping for stands from North-Rhine Westphalia, Lower Saxony and Brandenburg. Stands from Thuringia and Baden-Wuerttemberg were linearly distributed along Coord. 1 (see detailed representation in Supplementary Fig. 3S). Overall, the North American reference populations were more differentiated than the German ones (Fig. 4). NQ-E, NQ-R and Nantahala (USA-4) of the upper Midwest and the south, which all contained rare haplotypes of lineage H, clustered together. MTU and Mine consisted mainly of haplotype B, while BR1 and Keweenaw on the upper end of the distribution mainly contained haplotype C. These four populations are located on the Keweenaw Peninsula at Lake Superior. The PCoA analyses showed that some German stands were similar in haplotype composition to North American populations and were consequently assigned to these reference populations with the GENECLASS2 analyses (Supplementary Table 5S). This is particularly the case for the German stands 23, 24, 30, 31 and 38, which all were assigned to the North American population Anderson in the north-central part of the natural range. In general, most German stands could be assigned to populations from the northern part of the natural range. However, most assignments were apparently inaccurate, because only a low score for each assigned population was calculated.

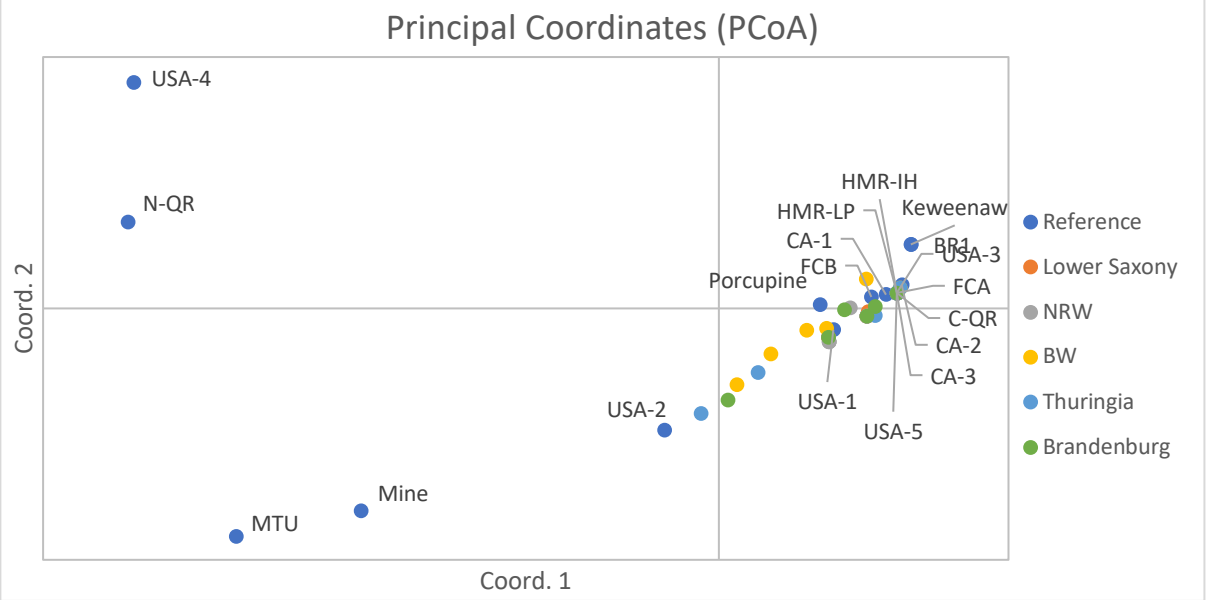


Fig. 4 Principal Coordinate Analysis (PCoA) based on *Quercus rubra* chloroplast markers for all populations (reference populations are labelled)

3.4 Relationship between genetic and geographic distances

While the Mantel test showed no significant correlation (R_{XY} = 0.148, p = 0.180) between genetic and geographic distances for North American reference populations (Supplementary Fig. 4S), it revealed a slightly positive significant correlation (R_{XY} = 0.284, p = 0.001) for the stands in Germany (Supplementary Fig. 5S). In addition, while still very low, the R^2 value in Germany was four times higher (R^2 = 0.08) than for North American reference populations (R^2 = 0.02).

4 Discussion

4.1 German plantations originated from a limited geographic range in North America

Northern red oak stands in Germany seem to originate from a geographically restricted region in the northern part of North America, covering two lineages, one of which is found all over North America, but mainly in the north and north-east (lineage A) and one that was found mainly in the north-central and north-western parts (lineage B, see Figs. 1, 3). Both lower haplotype diversity and lower differentiation among German stands point to founder effects, suggesting that German stands represent only a fraction of the diversity found in North America. This is supported by the multivariate PCoA analysis where populations from North America are widely scattered while German stands form a compact group.

Likewise, the lower differentiation among populations (G_{ST} = 0.137) also suggests that German stands originate from a restricted geographic range. In addition, the absence of haplotypes G, K and L (27.0 % in North America) supports the conclusion that German *Q. rubra* stands originated from a limited geographic range in North America. In accordance with our data, genetic differentiation at cpDNA markers among *Q. rubra* populations from a restricted region in the Great Lakes region was considerably lower (G_{ST} = 0.206, Zhang et al. 2015) as compared to differentiation described in studies that covered a range exceeding the Great Lakes region (G_{ST} = 0.73; this study; G_{ST} = 0.58, populations in the north-western part of the natural range, Birchenko et al. 2009; G_{ST} = 0.46, populations covering the whole natural range, Magni et al. 2005). In agreement with previous findings by Magni et al. (2005), no clear geographic genetic structure of haplotypes was found within the natural distribution area of *Q. rubra* in this study. The absence of geographic genetic structures within the natural range, as indicated also by the Mantel test, impedes the detailed identification of the geographic origin of German plantations (e.g. with GENECLASS2). Although their exact origin cannot be determined, the presence of haplotype B with restricted distribution in North America (two

populations on the Upper Peninsula of Michigan, one population in Indiana) would also point to the origin of German stands from the northern part of the natural distribution range. A range-wide high-density characterization of haplotypes in North American *Q. rubra* populations, as it was done for European white oak species (Petit et al. 2002), might allow to narrow down the origin of German red oak stands. The fully sequenced chloroplast genome of *Q. rubra* (Alexander and Woeste 2014), including locations for intraspecific polymorphisms, can serve as a reference for the assembly of chloroplast genomes of red oak samples from all over the natural range.

Our findings conform with a recent study analysing SNP markers in North American red oak populations as well as in European populations that are located mainly in France: Merceron et al. (2017) found that from three identified genetic clusters (G1, G2, G3) within the natural range, only two were observed in Europe. From these three clusters, one occurs mainly in the south (G2), one mainly in the north-east (G1), and the third cluster (G3), which supposedly diverged from the first two clusters, is mainly located in the north-central and north-west of the natural distribution (Merceron et al. 2017). Furthermore, G3 is reported to occur more evenly distributed over all regions. In their study, the genetic cluster mainly found in the south of the natural distribution could have been largely extirpated in or have never been introduced to Europe in the first place. Merceron et al. (2017) stated that this could be due to the source populations for the European gene pool being located in the Northern part of the range. Likewise, an early study addressing the taxonomy of *Q. rubra* variants by means of phenotypic traits came to the conclusion that German red oak stands comprise mainly *Q. rubra* var. *rubra*, a variety that is characterized by a shallower cupule and bigger acorns and is predominant in the North of the natural distribution area (Bauer 1954).

Although our sampling in North America was limited and not designed to analyse haplotype diversity across the whole natural distribution area, the geographic distribution of the three cpDNA lineages found in this study seems to match the distribution of the three genetic clusters derived from SNP markers found by Merceron et al. (2017). According to both studies, only the two northern lineages or clusters were introduced to Europe, one that occurs mainly in north-central and north-west regions (G3 or lineage B) and one that occurs mainly in the north and north-east regions (G1 or lineage A) of the natural distribution area. In addition, haplotype B from lineage B was found in all regions in Germany but was present only in the north and central part of the natural distribution range in North America.

When a species migrates to an unoccupied geographic area (e.g. an island or another continent), its gene pool hardly represents its full range of varieties. Merely a fraction of the species' variation provides the foundation for the future population, being challenged not only by new, but also constantly changing environmental conditions (Mayr 1954; Meimberg et al. 2006). Northern red oak seems to find itself in exactly this position in Germany: sampling from a limited geographic range has limited the genetic diversity at chloroplast markers in most German stands, reflecting founder events far away from the original distribution area. Although *Q. rubra* is currently well adapted to conditions in Germany (Roloff and Grundmann 2008), putative limited diversity at nuclear genes might lower the capacity to react to changing environmental conditions in the future. To describe the adaptive potential of red oak stands in Germany, future studies should focus on the assessment of genetic diversity at nuclear SSRs and genic EST-SSRs and at candidate genes in representative samples of selected German and North American reference populations.

4.2 Admixed material at least for plantations from Southwest Germany

While stands in Germany generally show low overall haplotype diversity and low genetic differentiation among populations, a different pattern is found for the stands probed in Baden-Wuerttemberg. Here, higher haplotype diversity and numbers of private haplotypes were found compared to all other examined regions of Germany (e.g. the more frequent occurrence of haplotype C in Baden-Wuerttemberg, while this haplotype is rare in other federal states). This suggests that imported seeding material for the establishment of the stands in Baden-Wuerttemberg was admixed to consist of material from different regions within the natural distribution range (e.g. due to multiple introductions). In fact, northern red oak went through several periods of cultivation since its first introduction at the end of the seventeenth century. After first unsuccessful trials in the middle of the eighteenth century, there were intensified efforts to establish *Q. rubra* in the second half of the nineteenth century and in the middle of the twentieth century (Nagel 2015). The material for the establishment of the new stands could partly have been collected from red oak stands of the first two generations, but as France and the Netherlands were also making efforts to establish northern red oak (Nagel 2015), the material could also have been obtained from there. Furthermore, additional material could have been brought directly from populations within the natural distribution area. In either case, the positive correlation between the genetic and geographic distances for German red oak stands

found in this study (Supplementary Fig. 5S) is in accordance with the introduction of different gene pools at different regions in Germany.

4.3 Future perspective

Our study revealed that although chloroplast haplotype variation in red oak in Germany is relatively low, some evidence for admixture and multiple introductions of this species are reflected in higher haplotypic diversity and genetic differentiation in stands in Baden-Wuerttemberg. Unlike North America where it has a limited geographic distribution (mainly in the northern part of the natural distribution range), the presence of haplotype B in all regions in Germany suggests an origin of German stands from the northern part of the natural range. These results support the similar conclusion drawn by Merceron et al. (2017) that red oak populations in Europe originated from the north of the natural distribution area. Further analyses at microsatellite and SNP markers in adaptive genes should be performed to confirm the genetic divergence in two lineages within Europe. If the number of sampled populations in both North America and Germany and the number of cpDNA and other markers are increased, the origin of German red oak stands in the north of the natural distribution area could potentially be narrowed down further based on the occurrence of rare and geographically restricted variants. The level of genetic variation and adaptive potential of the species should also be characterized at nuclear markers including candidate genes with potential role in local adaptation.

Acknowledgments: We thank Alexandra Dolynska, Andreas Parth, Oliver Caré, Katrin Burger and Maximilian Boersch for their technical assistance. We also thank two anonymous reviewers for their helpful comments on earlier drafts and Edward Jones for attentive English editing. The study was supported by the German Federal Ministry of Food and Agriculture (Funding code 22023314).

5 References

Alexander LW, Woeste KE (2014) Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. Tree Genetics & Genomes 10:803–812

Barrett SC, Husband BC (1990) The genetics of plant migration and colonization. In: Brown HD, Clegg MT, Kahler AL et al (eds) Plant population genetics, breeding, and genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, pp 254–277

Bauer F (1951) Die Roteiche 1950. Eine ertragskundlich-biologische und holzkundliche Untersuchung. Dissertation, Universität Göttingen

Bauer F (1954) Zur Rassenfrage der Roteiche. Allgemeine Forstzeitschrift 9:470–474

Birchenko I, Feng Y, Romero-Severson J (2009) Biogeographical distribution of chloroplast diversity in northern red oak (*Quercus rubra* L.). The American Midland Naturalist 161:134–145

Borkowski DS, Hoban SM, Chatwin W, Romero-Severson J (2017) Rangewide population differentiation and population substructure in *Quercus rubra* L. Tree Genetics & Genomes 13:472

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Molecular Ecology 13:2101–2106

Bundesministerium für Ernährung und Landwirtschaft (BMEL) (2014) Der Wald in Deutschland. Ausgewählte Ergebnisse der dritten Bundeswaldinventur. Available via https://www.bundeswaldinventur.de/fileadmin/SITE_MASTER/content/Dokumente/Downloads/BMEL_Wald_Broschuere.pdf. Accessed 20 Mar 2018

Burban C, Petit RJ, Carcreff E, Jactel H (1999) Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. Molecular Ecology 8:1593–1602

Chmura D (2013) Impact of alien tree species *Quercus rubra* L. on understorey environment and flora: A study of the silesian upland (southern Poland). Polish Journal of Ecology 61:431–442

Daubree JB, Kremer A (1993) Genetic and phenological differentiation between introduced and natural populations of *Quercus rubra* L. Ann. For. Sci. 50:271s-280s

Deguilloux M-F, Dumolin-Lapègue S, Gielly D, Grivet D, Petit RJ (2003) A set of primers for the amplification of chloroplast microsatellites in *Quercus*. Molecular Ecology Notes:24–27

Dlugosch KM, Parker IM (2008) Founding events in species invasions. Genetic variation, adaptive evolution, and the role of multiple introductions. Molecular Ecology 17:431–449

Dreßel R, Jäger EJ (2002) Beiträge zur Biologie der Gefäßpflanzen des herzynischen Raumes. 5. *Quercus rubra* L. (Roteiche): Lebensgeschichte und agriophytische Ausbreitung im Nationalpark Sächsische Schweiz. Hercynia:37–64

Excoffier L, Lischer HLE (2010) Arlequin suite ver 3.5. A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567

Gailing O, Wachter H, Heyder J, Rogge M, Finkeldey R (2009) Chloroplast DNA analyses of very old, presumably autochthonous *Quercus robur* L. stands in North Rhine-Westphalia. Allgemeine Forst- und Jagdzeitung 180:221–227

Google Maps (2017a) Map. Populations in Germany. Available via https://www.google.de/maps/@50.8871706,7.9740269,7z/data=!3m1!4b1!4m2!6m1!1s18cXo5V8opHdmODvdDhOnbGxKt3Q1t6S0?hl=de. Accessed 15 Dec 2017

Google Maps (2017b) Map. Populations south of the great lakes. Available via https://www.google.de/maps/@42.6170767,-82.2325253,7z/data=!4m2!6m1!1s1CaBHN6sHvvBYSIyeuU2vk4sRkzs?hl=de. Accessed 12 Dec 2017

Google Maps (2017c) Map. USA & Canada (East). Available via https://www.google.de/maps/@42.6170767,-82.2325253,7z/data=!4m2!6m1!1s1CaBHN6sHvvBYSIyeuU2vk4sRkzs?hl=de. Accessed 12 Dec 2017

Graw J (2005) Genetik, 4th edn. Springer, Berlin

Grivet D, Deguilloux M-F, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. Molecular Ecology 15:4085–4093

Heuertz M, Fineschi S, Anzidei M, Pastorelli R, Salvini D, Paule L, Frascaria-Lacoste N, Hardy OJ, Vekemans X, Vendramin GG (2004) Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. Molecular Ecology 13:3437–3452

Hulme PE (2009) Trade, transport and trouble. Managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46:10–18

Liesebach M, Schneck V (2011) Entwicklung von amerikanischen und europäischen Herkünften der Roteiche in Deutschland. Forstarchiv 82:125–133

Lind JF, Gailing O (2013) Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genetics & Genomes 9:707–722

Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a CONSTANS-like gene between two red oak species. Annals of Botany 113:967–975

Little EL (1999) Range of *Quercus rubra*. U.S. Geological Survey

López de Heredia U, Carrión JS, Jiménez P, Collada C, Gil L (2007) Molecular and palaeoecological evidence for multiple glacial refugia for evergreen oaks on the Iberian Peninsula. J Biogeography 34:1505–1517

Magni CR, Ducousso A, Caron H, Petit RJ, Kremer A (2005) Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. Molecular Ecology 14:513–524

Magni Diaz CR (2004) Reconstitution de l'introduction de *Quercus rubra* L. en Europe et conséquences génétiques dans les populations allochtones. Dissertation, Ècole Nationale du Génie Rural, des Eaux et des Forêts (in French)

Mayr E (1954) Change of the genetic environment and evolution. In: Huxley J, Hardy AC, Ford EB (eds) Evolution as a process. Allen & Unwin, London, pp 157–180

Meimberg H, Hammond JI, Jorgensen CM, Park TW, Gerlach JD, Rice KJ, McKay JK (2006) Molecular evidence for an extreme genetic bottleneck during introduction of an invading grass to California. Biol Invasions 8:1355–1366

Merceron NR, Leroy T, Chancerel E, Romero-Severson J, Borkowski DS, Ducousso A, Monty A, Porté AJ, Kremer A (2017) Back to America. Tracking the origin of European introduced populations of *Quercus rubra* L. Genome 60:778–790

Möllerová J (2005) Notes on invasive and expansive trees and shrubs. Journal of Forest Science 51:19–23

Nagel R-V (2015) Roteiche (*Quercus rubra* L.). In: Vor T, Spellmann H, Bolte A et al (eds) Potenziale und Risiken eingeführter Baumarten. Baumartenportraits mit naturschutzfachlicher Bewertung. Univ.-Verl. Göttingen, Göttingen, pp 219–267

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Niinemets Ü, Valladares F (2006) Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. Ecological Monographs 76:521–547

Palmé AE, Su Q, Rautenberg A, Manni F, Lascoux M (2003) Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. Molecular Ecology 12:201–212

Peakall R, Smouse PE (2006) GenAlEx 6. Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288–295

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

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, Vries SGM de, Ziegenhagen B, Beaulieu J-L de, Kremer A (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management 156:49–74

Petit RJ, Kremer A, Wagner DB (1993) Geographic structure of chloroplast DNA polymorphisms in European oaks. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 87:122–128

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2. A software for genetic assignment and first-generation migrant detection. The Journal of Heredity 95:536–539

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144:1237–1245

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences 94:9197–9201

Riepšas E, Straigytė L (2008) Invasiveness and ecological effects of red oak (*Quercus rubra* L.) in Lithuanian Forests. Baltic Forestry 14:122–130

Roloff A, Grundmann B (2008) Klimawandel und Baumarten-Verwendung für Waldökosysteme. Technische Universität Dresden. Available via http://waldundklima.de/klima/klima_docs/swin_roloff_klam_2008.pdf. Accessed 26 Feb 2018

Romero-Severson J, Aldrich P, Feng Y, Sun W, Michler C (2003) Chloroplast DNA variation of northern red oak (*Quercus rubra* L.) in Indiana. New Forests 26:43–49

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

Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. Molecular Ecology 17:351–360

Vor T (2005) Natural regeneration of *Quercus rubra* L. (Red Oak) in Germany. In: Nentwig W, Bacher S, Cock MJW et al (eds) Biological invasions - From ecology to control, pp 111–123

Vor T, Lüpke Bv (2004) Das Wachstum von Roteiche, Traubeneiche und Rotbuche unter verschiedenen Lichtbedingungen in den ersten beiden Jahren nach der Pflanzung. Forstarchiv 75:13–19

Weising K, Gardner RC (1999) A set of conserved PCR Primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Zhang R, Hipp AL, Gailing O (2015) Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations. Botany 93:691–700

6 Supplementary Files

Fig. 1S Study sites in Germany; different colours represent federal states (Google Maps 2017a)

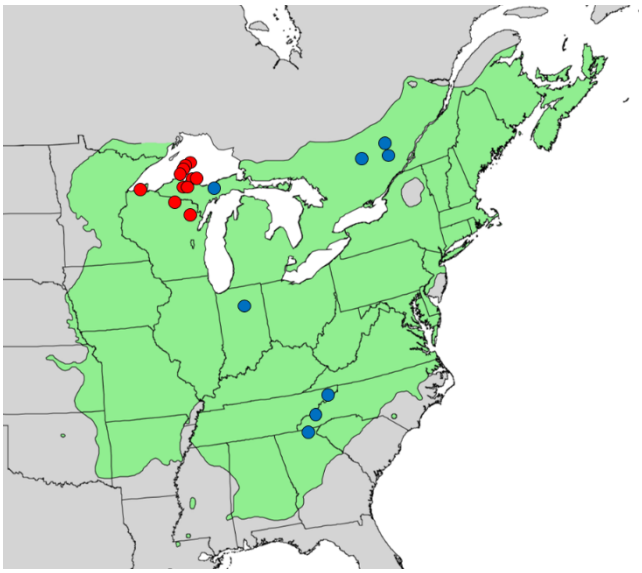


Fig. 2S Natural range and reference populations in North America (red and blue sites are listed in Tables 2S and 3S, respectively) (Little 1999)

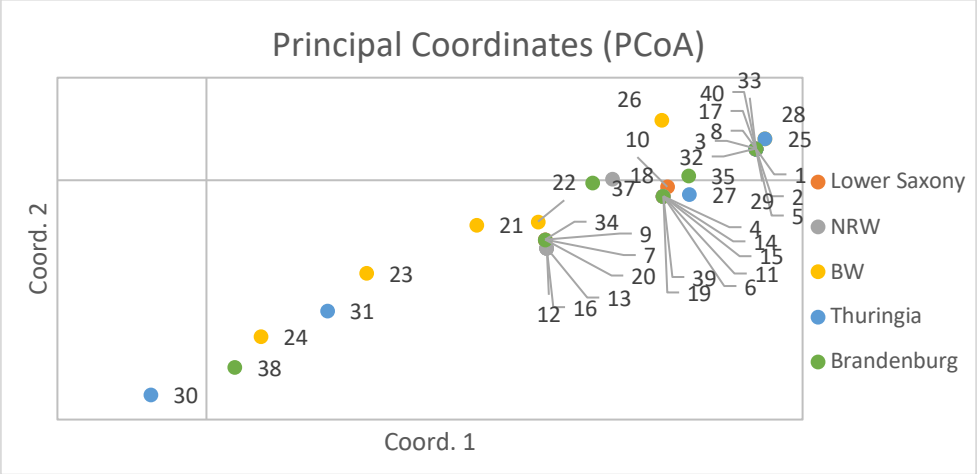


Fig. 3S Principal Coordinate Analysis (PCoA) of German stands (see Table 1S) based on *Quercus rubra* chloroplast markers (Table 1)

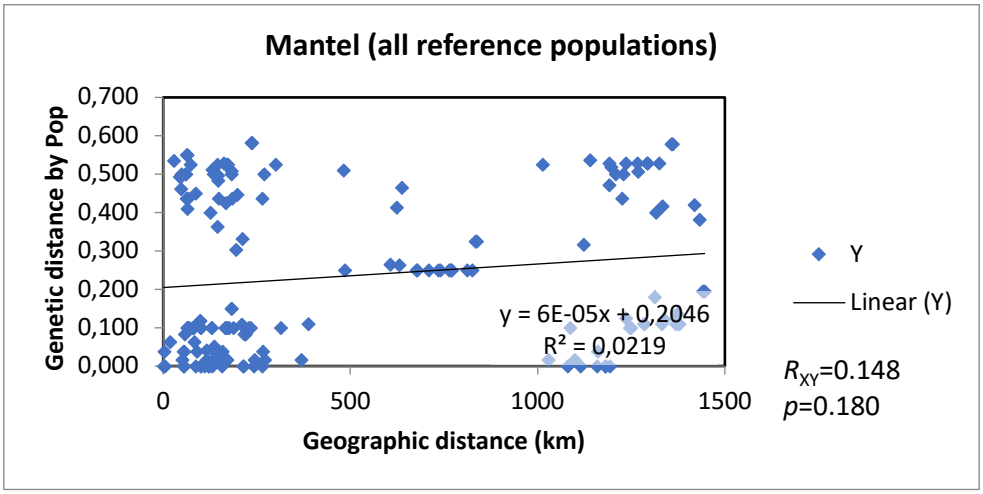


Fig. 4S Mantel test for all reference populations of *Quercus rubra*

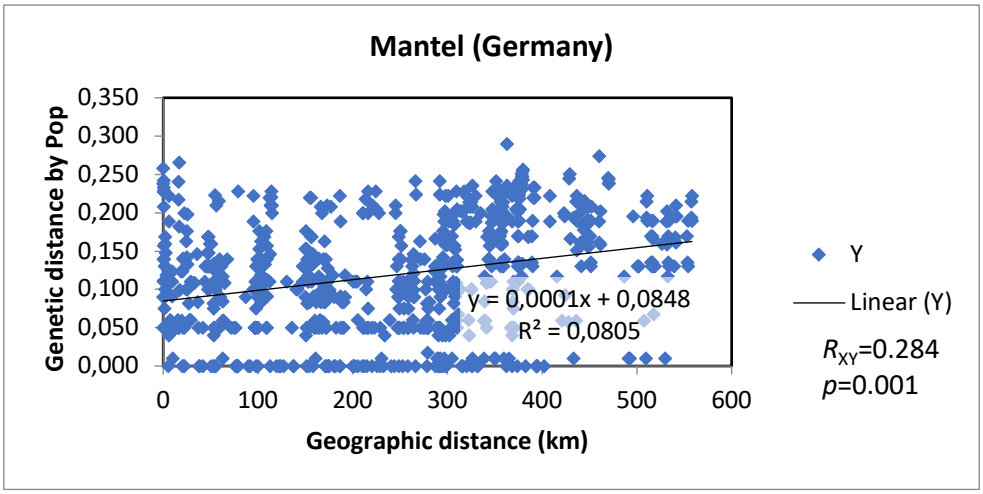


Fig. 5S Mantel test for German populations of *Quercus rubra*

Table 1S Study sites in Germany

#	Region	District	N	Latitude	Longitude	Altitude
1	Dassel	Sievershausen	10	51.77971	9.58899	524 m
2	Rotenburg	Spange 1	10	53.00454	9.34475	84 m
3	Rotenburg	Spange 2	10	53.00273	9.26215	80 m
4	Fuhrberg	Beerbusch	10	52.40619	9.83055	73 m
5	Rotenburg	Thörenwald	10	53.34354	9.35411	57 m
6	Rotenburg	Luhne	10	53.14049	9.37014	54 m
7	Rotenburg	Diensthoop	10	52.94624	9.36675	69 m
8	Rotenburg	Hepstedt 1	8	53.18048	9.18389	58 m
9	Göhrde	Zienitz	10	53.10025	10.8545	94 m
10	Rotenburg	Hepstedt 2	10	53.24972	9.18638	53 m
11	Rhein-Sieg-Erft	Knechtstenden	10	51.03911	6.80227	72 m
12	Niederrhein	Schwalm-Nette 1	10	51.15004	6.19473	104 m
13	Niederrhein	Schwalm-Nette 2	10	51.16385	6.18172	97 m
14	Niederrhein	Leucht 1	10	51.54538	6.48695	53 m
15	Niederrhein	Leucht 2	10	51.53801	6.50687	52 m
16	Niederrhein	Leucht 3	10	51.55844,	6.50824	73 m
17	Niederrhein	Leucht 4	10	51.53854	6.51056	57 m
18	Niederrhein	Leucht 5	10	51.53741	6.49099	68 m
19	Niederrhein	Leucht 6	10	51.54422	6.48661	56 m
20	Niederrhein	Leucht 7	10	51.53946	6.49804	72 m
21	Offenburg	Schutterwald 1	10	48.45871	7.86786	158 m
22	Offenburg	Schutterwald 2	10	48.45346	7.85918	148 m
23	Offenburg	Schutterwald 3	10	48.46007	7.86184	152 m
24	Offenburg	Schutterwald 4	10	48.45299	7.85969	148 m
25	Achern	Großweiher	10	48.65421	8.03769	143 m
26	Oberkirch	Renchen	10	48.57306	7.99869	171 m
27	Weida	Treben 1	10	51.06314	12.42265	166 m
28	Weida	Leina	10	50.97307	12.54363	222 m
29	Sonderhausen	Volkenroda	10	51.29994	10.66687	315 m
30	Neustadt	Strößwitz 1	9	50.76691	11.72482	382 m
31	Neustadt	Strößwitz 2	9	50.76442	11.7242	392 m
32	Jena	Holzland	10	50.77549	11.64021	372 m
33	Weida	Treben 2	10	51.05847	12.50526	190 m
34	Wünsdorf	Großbeeren	10	52.36366	13.35163	46 m
35	Potsdam	Güterfelde 1	10	52.45505	13.0795	47 m
36	Potsdam	Güterfelde 2	*	52.41546	13.0043	45 m
37	Rathenow	Kater	9	52.55401	12.2304	44 m
38	Potsdam	Güterfelde 3	10	52.38356	13.00608	38 m
39	Lehnin	Brandenburg	10	52.41631	12.46474	37 m
40	Baruth	Merzdorf	10	51.98894	13.37158	143 m

*Stand #36 was removed due to insufficient number of samples.

Table 2S Geographic origin in North America for samples collected from a provenance trial in Lübeck, Germany (Liesebach and Schneck 2011)

Population	Abbr.	Country	State	Location	N
Atomic Energy	CA-1	Canada	ON	Renfrew, Atomic Energy, Chalk River	6
Constance Bay	CA-2	Canada	ON	Ottawa, Constance Bay	6
Pl. de Kazabazua	CA-3	Canada	ON	Gatineau, Plaines de Kazabazua, Basse-Lièvre	6
Chattahoochee	USA-1	USA	GA	Fannin, Chattahoochee, Toccoa	5
Anderson	USA-2	USA	IN	Madison, Anderson	6
Hiawatha	USA-3	USA	MI	Chippewa, Hiawatha, Soo	6
Nantahala	USA-4	USA	NC	Clay, Nantahala, Tusquitee	7
Cherokee	USA-5	USA	TN	Sullivan, Cherokee, Watauga	5

Note: The origin of these populations in North America is shown in blue colour in Fig. 2S. State: ON (Ontario); GA (Georgia); IN (Indiana); MI (Michigan); NC (North Carolina); TN (Tennessee)

Table 3S Study sites in North America

Abbreviation	Region	N	Latitude	Longitude	Altitude
FC-A[1]	Ford Forestry Center, MI	8	46.65261	-88.50193	415 m
FC-B[1]	Ford Forestry Center, MI	8	46.67442	-88.52427	393 m
BR1[1]	Brockway Mountain, MI	8	47.46616	-87.91671	355 m
MTU[1]	Michigan Tech Trails, MI	8	47.10055	-88.5475	270 m
HMR-IH[1]	Huron Mountain Res, MI	8	46.85357	-87.84522	256 m
HMR-LP[1]	Huron Mountain Res, MI	8	46.84994	-87.83022	289 m
N-QR[2]	Nicolet NF, WI	8	45.34805	-88.38805	345 m
Keweenaw[3]	Brockway Mountain, MI	8	47.4401	-87.85658	321 m
Mine[3]	Calumet Township, MI	8	47.25398	-88.42676	372 m
Porcupine[3]	Kentuck Lake, WI	8	46.00003	-88.9996	532 m
C-QR[2]	Chequamegon NF, WI	8	46.715	-91.03555	328 m

Note: MI - Michigan; WI – Wisconsin; [1]Lind and Gailing 2013; [2]Lind-Riehl et al. 2014; [3]unpublished. The origin of these population in North America is shown in Figs 2 and 3.

Table 4S Chloroplast microsatellite markers (cpSSRs) used in this study

cpSSR	PCR primer		Repeat motif	Size, bp	n_a	Location
	Direction	**Nucleotide sequences (5'-3')**				
ccmp2[1]	Forward	GATCCCGGACGTAATCCTG	$(A)_{11}$	233-234	2	5' to *trnS*
	Reverse	ATCGTACCGAGGGGTTCGAAT				
ccmp4[1]	Forward	AATGCTGAATCGAYGACCTA	$(T)_{13}$	126	3	*atpF* intron
	Reverse	CCAAAATATTBGGAGGACTCT				
ucd4[2]	Forward	TTATTTGTTTTTGGTTTCACC	$(T)_{12}$	97	3	intergenic
	Reverse	TTTCCCATAGAGAGTCTGTAT				ycf6-psbM
udt1[2]	Forward	ATCTTACACTAAGCTCGGAA	$(A)_{11}$	86	3	intergenic
	Reverse	TTCAATAACTTGTTGATCCC				trnE-trnT
udt4[2]	Forward	GATAATATAAAGAGTCAAAT	$(A)_9$	147	3	Intergenic
	Reverse	CCGAAAGGTCCTATACCTCG				trnE-trnT

Note: [1]Weising and Gardner 1999; [2]Deguilloux et al. 2003

Table 5S Populations assignment with GENECLASS2 based on the Bayesian method (Rannala and Mountain 1997). Assignments to the south, centre (yellow) and north (green) of the natural distribution area. Populations are marked white, if clear assignment to one of the three areas was impossible.

			Rank 1	Score	Rank 2	Score	Rank 3	Score
Assigned sample			Assigned Ref	%	Assigned Ref	%	Assigned Ref	%
1	Dassel	Sievershausen	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
2	Rotenburg	Spange 1	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
3	Rotenburg	Spange 2	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
4	Fuhrberg	Beerbusch	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
5	Rotenburg	Thörenwald	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
6	Rotenburg	Luhne	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
7	Rotenburg	Diensthoop	Chattahoochee	49.1	Anderson	17.3	Atomic Energy	9.9
8	Rotenburg	Hepstedt 1	C-QR	16.1	HMR-LP	16.1	HMR-IH	16.1
9	Göhrde	Zienitz	Chattahoochee	49.1	Anderson	17.3	Atomic Energy	9.9
10	Rotenburg	Hepstedt 2	Atomic Energy	16.3	Constance Bay	16.3	Cherokee	8.5
11	Rhein-Sieg-Erft	Knechtstenden	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
12	Niederrhein	Schwalm-Nette 1	Chattahoochee	50.0	Anderson	16.0	Atomic Energy	11.0
13	Niederrhein	Schwalm-Nette 2	Chattahoochee	50.0	Anderson	16.0	Atomic Energy	11.0
14	Niederrhein	Leucht 1	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
15	Niederrhein	Leucht 2	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
16	Niederrhein	Leucht 3	Chattahoochee	50.0	Anderson	16.0	Atomic Energy	11.0
17	Niederrhein	Leucht 4	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
18	Niederrhein	Leucht 5	Chattahoochee	17.5	Atomic Energy	17.2	Cherokee	10.3
19	Niederrhein	Leucht 6	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
20	Niederrhein	Leucht 7	Chattahoochee	49.1	Anderson	17.3	Atomic Energy	9.9
21	Offenburg	Schutterwald 1	Chattahoochee	50.0	Anderson	33.2	Atomic Energy	5.2
22	Offenburg	Schutterwald 2	Chattahoochee	47.5	Anderson	20.7	Constance Bay	11.3
23	Offenburg	Schutterwald 3	Anderson	89.8	Chattahoochee	9.5	Atomic Energy	0.3
24	Offenburg	Schutterwald 4	Anderson	98.5	Chattahoochee	1.4	Atomic Energy	0.0
25	Achern	Großweiher	Constance Bay	16.5	HMR-LP	13.7	HMR-IH	13.7
26	Oberkirch	Renchen	Keweenaw	48.8	BR1	48.8	Constance Bay	1.5
27	Weida	Treben 1	Atomic Energy	17.5	C-QR	10.1	HMR-LP	10.1
28	Weida	Leina	Constance Bay	16.5	HMR-LP	13.7	HMR-IH	13.7
29	Sonderhausen	Volkenroda	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
30	Neustadt	Strößwitz 1	Anderson	99.6	Chattahoochee	0.4	Mine	0.0
31	Neustadt	Strößwitz 2	Anderson	95.5	Chattahoochee	4.2	Atomic Energy	0.1
32	Jena	Holzland	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
33	Weida	Treben 2	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6
34	Wünsdorf	Großbeeren	Chattahoochee	49.1	Anderson	17.3	Atomic Energy	9.9
35	Potsdam	Güterfelde 1	Atomic Energy	16.7	Constance Bay	16.7	Pd Kazabazua	8.4
37	Rathenow	Kater	Chattahoochee	24.7	Atomic Energy	16.6	Cherokee	9.8
38	Potsdam	Güterfelde 3	Anderson	99.1	Chattahoochee	0.9	Atomic Energy	0.0
39	Lehnin	Brandenburg	Atomic Energy	17.4	Pd Kazabazua	9.5	Hiawatha	9.5
40	Baruth	Merzdorf	C-QR	16.6	HMR-LP	16.6	HMR-IH	16.6

Paper 3

Development of novel *Quercus rubra* chloroplast genome CAPS markers

Submitted

Annals of Forest Research

23 January 2019

Development of novel *Quercus rubra* chloroplast genome CAPS markers

Tim Pettenkofer[1] • Reiner Finkeldey[2] • Markus Müller[1] • Konstantin V. Krutovsky[1,4,5,6,7] • Barbara Vornam[1] • Ludger Leinemann[1] • Oliver Gailing[1,3,7]

Corresponding authors: Oliver Gailing[1,3,7], E-mail: ogailin@gwdg.de; Ludger Leinemann[1], E-mail: lleinem@gwdg.de, Tel: +49 551 3933536

[1] University of Göttingen, Faculty of Forest Sciences and Forest Ecology, Forest Genetics and Forest Tree Breeding, Büsgenweg 2, 37077 Göttingen, Germany

[2] University of Kassel, Mönchebergstraße 19, 34125 Kassel, Germany

[3] Michigan Technological University, School of Forest Resources and Environmental Science, Houghton, MI 49931, United States

[4] Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 660036 Krasnoyarsk, Russia

[5] Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia

[6] Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, USA

[7] University of Göttingen, Center for Integrated Breeding Research (CiBreed), Göttingen, Germany

Keywords: *Quercus rubra*, oak, chloroplast genome, sequencing, SNP, indel, CAPS, SSR

Abstract

Our main objective was to generate cost-effective chloroplast (cp) DNA markers that are easy to apply and to score. In combination with already published cpSSR markers they should increase haplotype resolution in populations. To discover new cpDNA markers, we sequenced 87-97 % of the entire chloroplast genome (except the second inverted repeat) of 8 trees representing different regions of the *Quercus rubra* L. natural range with 4,030X–6,297X coverage and assembled the genome sequences using the publicly available chloroplast genome of *Quercus rubra* L. as a reference. In total, 118 single nucleotide polymorphisms (SNPs) and 107 insertions or deletions (indels) were detected, and 15 cleaved amplified polymorphic sequence (CAPS) markers were developed for *Q. rubra*. Using these new markers together with five chloroplast microsatellite or simple sequence repeat (cpSSR) markers, we identified 10 haplotypes in our diversity panel of 19 *Q. rubra* populations. Specifically, two haplotypes based only on the cpSSR markers could now be separated in five haplotypes. These markers are useful to assess haplotype diversity with high resolution and are also transferable to a closely related species, *Quercus ellipsoidalis* E.J.Hill.

1 Introduction

Northern red oak (*Quercus rubra* L.) is a dominant deciduous tree species covering a large natural area in eastern North America (Borkowski et al. 2014; Fig. 1). It grows best on mesic sites with deep sandy loam, but can also be found on pug, loam or gravelly, and sandy soils, and thus covers a wide ecological range (Desmarais 1998; Nagel 2015).

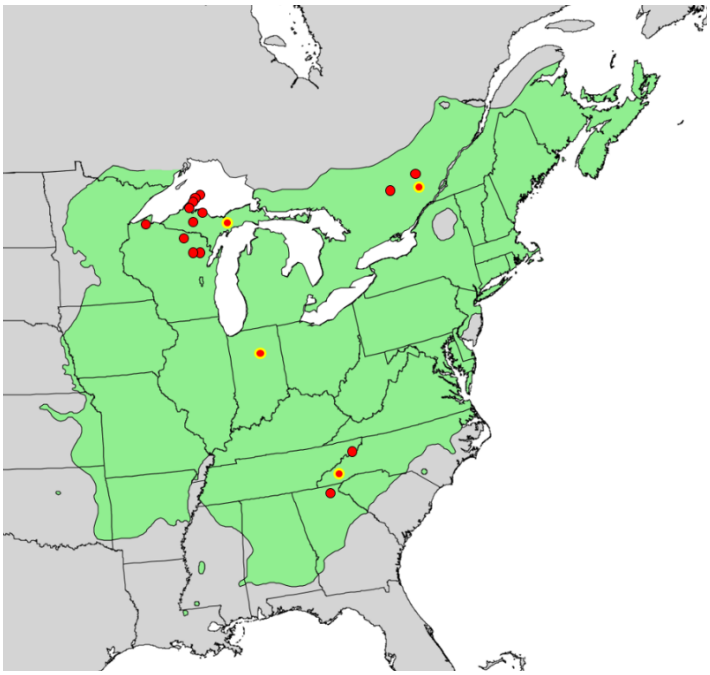


Figure 1 Natural range of *Quercus rubra* (Little 1999). Populations in the first set are marked with yellow circles (Liesebach und Schneck 2011), populations in the second set are marked in red (Lind und Gailing 2013; Lind-Riehl et al. 2014)

In contrary to European white oak species (Petit et al. 2002), *Q. rubra* is characterized by a relatively low geographic structure of chloroplast variation, making it difficult to distinguish between stands and geographic regions (Magni et al. 2005; Birchenko et al. 2009; Alexander and Woeste 2014). However, at larger distances between populations, significant geographical variation can be found in oak species (Petit et al. 1993; Grivet et al. 2006; Zhang et al. 2015). Within populations, chloroplast DNA (cpDNA) haplotypes usually show relatively low variation supposedly due to lack of recombination and its maternal inheritance (Finkeldey and Gailing 2013; Zhang et al. 2015). In the past, cpDNA markers have successfully been used to reveal the haplotype composition of autochthonous oak populations in Germany (Gailing et al. 2009) and to trace post-glacial recolonization routes of white oaks in Europe (Petit et al. 2002). By comparing identified haplotypes and haplotype variation with reference populations within the natural range, information of their origin can be inferred.

In this study, first, the red oak chloroplast genome was sequenced almost entirely (except the second inverted repeat region) for 8 samples representing different geographic regions. Then, the sequences were assembled using the publicly available chloroplast genome of *Q. rubra* as a reference (Alexander and Woeste 2014) to search for cpDNA polymorphisms. In total, 118 single nucleotide polymorphisms (SNPs) and 107 insertions or deletions (indels)

were identified. We used them to develop population specific cleaved amplified polymorphic sequence (CAPS) markers by selecting only those enzyme restriction sites that were present in one of the provenances but absent in all others. We developed PCR primers that amplify fragments with a single restriction site containing SNPs. We tested 23 primer-enzyme combinations in total, and 15 showed good results. These 15 CAPS markers can be genotyped by analysing digested fragment sizes in a simple agarose gel electrophoresis and do not need much DNA and expensive laboratory equipment (Kaundun and Matsumoto 2003). They can be used for relatively easy and cost-effective determination of chloroplast variation of oak populations. In case of cultivated stands of unknown origin, this tool can be applied to determine the origin of seeds or plants used to establish them if appropriate reference information for assignment is available. In addition to chloroplast microsatellite or simple sequence repeat (cpSSR) markers (Zhang et al. 2015; Pettenkofer et al. 2019), CAPS markers are expected to increase resolution of chloroplast variation.

Finally, the fully sequenced chloroplast genomes of trees representing all major regions within the natural range will also provide a good resource for the identification of new polymorphic markers for the characterization of intraspecific differentiation in the future.

Our main objectives were to a) introduce a novel set of easy-to-use and cost-effective CAPS markers applicable for *Q. rubra* and closely related species such as *Q. ellipsoidalis*, and b) reveal more chloroplast haplotypes by combining CAPS with cpSSR markers.

2 Material and methods

2.1 Plant material

We used two sets of samples (Table 1). Bud samples in the first set were collected from eight trees representing four different regions (two trees per region) within the natural distribution range *Q. rubra* (Table 1, Fig. 1) and were obtained from a provenance trial located in Luebeck, Germany. This provenance trial was established from seeds collected by Antoine Kremer in 1988 and then distributed to research institutes in France, Germany, Belgium, Great Britain and the USA (Liesebach and Schneck 2011). This set was used for amplicon-based sequencing of the chloroplast genome. The second set consisted of 96 samples collected from 19 different populations within the natural distribution range in North America and was used to develop and test CAPS markers (Table 1). This set included samples from populations described by Liesebach and Schneck (2011), Lind and Gailing (2013) and Lind-Riehl et al. (2014).

Table 1 Study sites in North America

Stand	Set	Country	Location	N	Lat.	Long.
Atomic Energy[1]	2	Canada, ON	Renfrew, Atomic Energy	5	-	-
Constance Bay[1]	1, 2	Canada, ON	Ottawa, Constance Bay	4	-	-
Pl. de Kazabazua[1]	2	Canada, ON	Gatineau, Plaines de Kazabazua, Basse-Lièvre	6	-	-
Chattahoochee[1]	2	USA, GA	Fannin, Chattahoochee, Toccoa	4	-	-
Anderson[1]	1, 2	USA, IN	Madison, Anderson	6	-	-
Hiawatha[1]	1, 2	USA, MI	Chippewa, Hiawatha, Soo	6	-	-
Nantahala[1]	1, 2	USA, NC	Clay, Nantahala, Tusquitee	6	-	-
Cherokee[1]	2	USA, TN	Sullivan, Cherokee, Watauga	5	-	-
FC-A[2]	2	USA, MI	Ford Forestry Center	8	46.65261	-88.50193
FC-B[2]	2	USA, MI	Ford Forestry Center	3	46.67442	-88.52427
BR1[2]	2	USA, MI	Brockway Mountain	3	47.46616	-87.91671
MTU[2]	2	USA, MI	Michigan Tech Trails	3	47.10055	-88.5475
HMR-LP[2]	2	USA, MI	Huron Mountain Res	8	46.84994	-87.83022
N-QE[3]*	2	USA, WI	Nicolet NF	4	45.32194	-88.33138
N-QR[3]	2	USA, WI	Nicolet NF	7	45.34805	-88.38805
Keweenaw[4]	2	USA, MI	Brockway Mountain	4	47.4401	-87.85658
Mine[4]	2	USA, MI	Calumet Township	2	47.25398	-88.42676
Porcupine[4]	2	USA, WI	Kentuck Lake	4	46.00003	-88.9996
C-QR[3]	2	USA, WI	Chequamegon NF	8	46.715	-91.03555

Note: MI - Michigan, WI - Wisconsin, GA - Georgia, IN - Indiana, NC - North Carolina, TN - Tennessee, ON - Ontario. [1] Liesebach und Schneck 2011, [2] Lind und Gailing 2013, [3] Lind-Riehl et al. 2014, [4] unpublished. The location of these populations in North America is shown in Fig. 1. *Population was genetically identified as *Q. ellipsoidalis* (Lind und Gailing 2013; Lind-Riehl et al. 2014), a species which is closely related to *Q. rubra*. Both species are interfertile and hybridize with each other in contact zones.

2.2 DNA isolation

Bud samples were collected in 2013 as fully grown, live buds from 10-20 cm long twigs. DNA was isolated subsequently with the DNeasyTM 96 Plant Kit from Qiagen (Hilden, Germany) following the instructions.

2.3 Amplicon sequencing

The red oak chloroplast genome (NCBI GenBank accession number JX970937.1) is 161304 bp long and contains a large single copy (LSC), a short single copy (SSC) and two highly conserved inverted repeat (IR) regions (Alexander and Woeste 2014). The second inverted repeat region is the reverse compliment of the first one. Excluding the second inverted repeat region due to lack of variation (Alexander and Woeste 2014), a total of 32 primer pairs (Table 2) were designed using BioEdit v. 7.2.6.1 (Hall 1999) and OligoCalc v. 3.27 (Kibbe 2007) software to amplify 132042 bp (96.87 %) of the genome (see Table 2 and Fig. 2).

The targeted cpDNA regions were PCR amplified in a 19 µl volume containing 1 µl of the template genomic DNA (~0.6 ng/µl), 10 µl HotStarTaq Master Mix from Qiagen (containing 5 units / µl HotStarTaq DNA Polymerase, PCR buffer with 3 mM $MgCl_2$ and 400 µM of each dNTP) (Hilden, Germany), 2.3 µl Q-Solution from Qiagen (Hilden, Germany), 1.1 µl $MgCl_2$ (25 mM), 2.3 µl of each forward and reverse primers (5 pM/µl).

PCR reactions were conducted in a MJ Research PTC-200 thermocycler (St. Bruno, Canada). The PCR profile started with 15 min initial denaturation at 95 °C, followed by 38 cycles of 1 min denaturation at 94 °C, 30 sec annealing at 55 °C and 6 min extension at 68 °C, and ended with a final 20 min extension step at 72 °C.

The PCR products were checked in an agarose gel electrophoresis at 90 volts for 20 to 30 minutes using a 1 % agarose gel with TAE as running buffer. The DNA was stained with Roti-Safe GelStain from Roth (Karlsruhe, Germany). Then, amplicon bands were cut out of the gel under UV light. DNA was extracted from the gel by using the innuPREP Gel Extraction Kit from Analytik Jena (Jena, Germany). The concentration of the eluted DNA was measured with the NanoDrop 2000 from Thermo Fisher Scientific (Waltham, USA).

Table 2 PCR primer sequences designed in the study and used to amplify the chloroplast genome regions (except the second inverted repeat region)

Primer pair	Forward primer (5’ -3’)	Reverse primer (5’ -3’)	Nucleotide position in the reference genome		Amplicon length, bp
			start	end	
1	GTGGATTCACAATCCACTGCCT	AAACCCAAGCATTGCGATACTTT	47	5057	5010
1.1*	GTGGATTCACAATCCACTGCCT	CCCTTGGATTGCTGTTGCATATTC	47	1181	1134
1.2*	CCCCCACTCCCTGTATGTAGT	AAAGGAGCAATGTCAACCCTCTTG	97	496	399
2	CCCAGCAAATCCGACGAATCCATC	GTAAGGCAACGGGTTTTGGTCCCG	4722	8177	3455
3	GGTATGCTCTGGGACGGAAGG	TCCGCCCATAACATCTATGTCAGC	8139	11057	2918
3.1	CTCATTAGTCTCCTTGGCCAAAGC	TGATCCTTCCAAACAAGCACAGGC	10353	12592	2239
3.2*	CCGGCGGTCTTCGTAATAGAAGAG	GAAGGACTGATTGGAGAATGGGGA	12554	15093	2539
4	GGGAAGGAAGAAGGCGAGTC	AGAGAATTGAGGGTTGGAACAAG	15062	19248	4186
4.1*	TGATGACCCTCCATGGATTCGCCT	TTGCAAACATATCTGGGCGGGATC	16670	17890	1220
5	TTGCTAAAACTCGAGCAGTTTCTT	GGTCCCTCCCTTTCATTACAT	18875	23831	4956
5.2*	CCATGAGTAGGACTTCGGCCA	CGGGAGCTAGAGGAAATGCATCTC	21853	22378	525
6	TCGAAGCACGATGTTGTCTAATTA	CCAGAGACTGAATCTTGATATACC	23746	28272	4526
7	GCCCATAGTGACTAGGATGGA	GGGAGGTTCGAACTTACCTTTT	27838	33018	5180
8*	GAAGATCCTTTATCCATACCGAA	TCCCCAGACATATTCCTCTCG	32795	36929	4134
8.1*	CCTTGAAAGAGAGATGTCCTGAAC	AAGCCCCTTATCGGATTTGAACCG	34539	35674	1135
8.2*	CCGTTTTGAGTTTACCCATAATCAC	CATTCGCCACTATGAGTCTAGTGC	35372	35565	193
9	AGTGCGTTGCTTGCTTTGAACCGC	CGTCGATCCCCCAAAAAAGAA	36686	42009	5323
10	CGGTGGGGATTGTAACTTTCC	GTCAAAATAGAAGGTCCTCCGG	41946	47387	5441
11	GAGTGCTCCTATTCGAAACGCCTTG	GATTGGGTATGGCCTCGACTATTCCC	47541	51141	3600
12	GGCTTTTCCGGCAATACGAACG	GTTTATGAAAGAGCCCAATGCAC	51231	56897	5666
13	GATAAAGAGATGGCTCCGTGTGCTCGG	CCAACGAAATCTAGTGCAGTGCG	56965	59963	2998
14	GCGGATCAATAAAAGTCTTGAG	TTGTCATAAAATGGATACCTGG	60077	64991	4914
15	ATAAAATATGTGATGTGGTTTCCCCCG	CATTAAATCCCATCCCTCTTTTGCGG	65200	69687	4487
16	GATCATTAAGAACTCAACGGG	CATTCCAAAATAACTGTTACTCG	69896	74012	4116
17	GCTAATAAGGAATCAAAGAGG	CATGGAAATACCCCTTTAATCAACG	74060	78977	4917

Primer pair	Forward primer (5’ -3’)	Reverse primer (5’ -3’)	Nucleotide position in the reference genome		Amplicon length, bp
			start	end	
17.1*	CCATGAGCTTGGGCTTCTGTTGCT	GGCCAAGAGGTTGATAGCGATCTC	76010	77115	1105
18	TATCGTGTTCATACCGTCGTATTG	CCAACTTAGCTCTTAATACAG	79014	83810	4796
19	TTATTATGGAGAACCCGCATGG	GAAAGGAATATCTTATCGAGGTAATCG	83850	88101	4251
20	GTTTACGGAATCTGGCTCTTTTGG	CAGACAAAAGTATTCGGTTATTG	88144	92771	4627
21	CTTCCCTATGAGTTTTAGTCTCAATAAG	AGAATTCAGCTCAGATGTAGG	92833	97328	4495
22	AGTTGCTCTGGAACAATTAGGAG	CCAAGAAATAACCCCTCACGTGCG	97372	101386	4014
23	GATAGAAACAACGCTCGTAAGGAG	CGCGGGCTTTACGCAATCGATCGG	101454	105630	4176
24	CGGAACCGGGGAAAGTATACAG	GGCAAGTCTTTGTGAAATAACTCCG	105661	109545	3884
25	GCACACTTGGAGAGCGCAGTACACCGG	AGGGGACACCAAAGGCCTCTG	109596	114020	4424
26	GCCTTTATTCGTTCATGGTTCGATATTCTG	ATGGTTCATACCTTTCATTCCAG	114093	118847	4754
26.1*	GCGGTGACGATACTGTAGGGGAGG	TCTTCGGCGCACGTGGGCTTTTCC	113976	118571	4595
27	GCTCGAAAAAAGAGTGCACCTAATTG	TCCTGATGTTAGCAATGTATAGCG	119283	123777	4494
28	GAATCTCGAGTAACCGGCCAAG	TAAGTGGTCGAATTCGGGGCCG	123815	127591	3776
29*	ATTTGCTGAGTAATAACTTTGAG	GAGTCTGGATATTCTGATGGTCCG	127696	131625	3929
30	GGATTTCTCAGTCTAAGCAGGAG	CTTATTAGATCTAATAAGTACCTTGTG	131688	136055	4367
30.1*	AAGCCGTTTTCGGGGATAACC	CTGACGAGGTAGCTTTGATCCG	133645	134411	766

Note: *used also for CAPS analysis (see also Fig. 2).

The collected amplicons representing 32 cpDNA fragments were pooled for each sample and sequenced by the Transcriptome and Genome Analysis Laboratory (TAL) at the University of Göttingen using the Illumina MiSeq sequencing system. Amplicon sizes ranged from 2239 to 5666 bp (average 4295 bp).

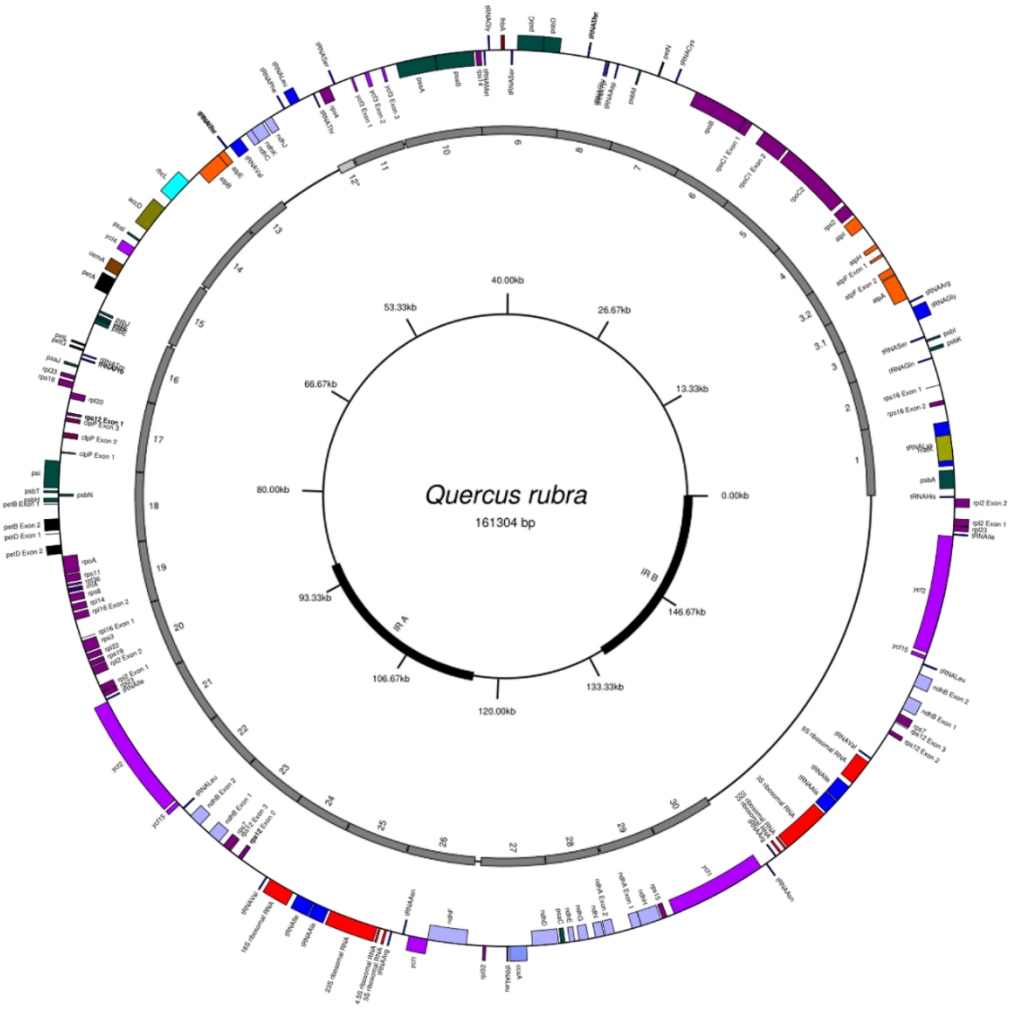


Figure 2 Chloroplast genome of *Quercus rubra* assembled and annotated by Alexander and Woeste (2014). Bold lines in the inner circle represent inverted repeat regions. The middle circle indicates positions of primers and gaps. The second inverted repeat region was excluded from amplicon sequencing. Modified after Alexander und Woeste (2014)

2.4 Sequence alignment and detection of SNPs and indels

Sequence alignment and detection of SNPs and indels were accomplished with the CLC Genomics Workbench v. 9.0.1 software (Qiagen, Aarhus, Denmark). Sequence reads were mapped to the *Q. rubra* chloroplast reference genome available from the NCBI GenBank nucleotide database (accession number JX970937.1). The BioEdit v. 7.2.6.1 (Hall 1999) and

OligoCalc v. 3.27 (Kibbe 2007) software were used to identify restriction sites and develop CAPS markers. The identified restriction enzymes were subsequently tested using 96 DNA samples representing 19 North American populations from different regions within the natural distribution area. The samples were selected to include different chloroplast haplotypes already identified in northern red oak populations in Germany and North America (Pettenkofer et al. 2019). Among 96 samples, 57 had cpSSR haplotype A, each of the haplotypes B, C, K and L was represented by eight samples, five samples had haplotype G, and two samples haplotype H. Except for the last two haplotypes, each haplotype was represented by samples from different locations of the natural range.

Table 3 Mapping of sequence reads to the *Q. rubra* chloroplast reference genome

Population[1]	Sample ID	Reference covered by reads, %	Total number of reads mapped	Mean read length, bp	Total read length, bp	GC, %	Average coverage
Nantahala	3	92	3,325,327	226.75	754,029,642	36	5,488
	4	94	3,298,374	222.61	734,237,502	36	5,335
Anderson	7	97	3,336,145	204.31	681,614,046	36	4,965
	8	96	2,557,784	215.95	552,357,621	36	4,030
Hiawatha	10	93	3,018,802	214.73	348,233,509	36	4,718
	12	87	3,052,790	222.67	679,766,646	36	4,959
Constance	13	93	3,881,031	222.43	863,256,125	36	6,297
Bay	14	94	3,374,986	207.92	701,724,192	36	5,103

Note: Total length of genome used as reference was 136299 bp. [1] The locations of these populations in North America are presented in Fig. 1.

2.5 Restriction digestion

Directly after PCR amplification, 1 µl of restriction enzyme and 2 µl of associated buffer were added to 8 µl of the PCR reaction. Then, the mixtures were incubated following the recommended protocol provided by the manufacturer (Thermo Fisher Scientific, Waltham, USA). After incubation, an agarose gel electrophoresis was performed at 90 volts for 20 to 30 minutes on a 1 %-, 1.5 %- or 2.5 % agarose gel (depending on fragment sizes) with TAE as running buffer.

2.6 Identification of haplotypes

Chloroplast microsatellite haplotypes for each sample were identified in Pettenkofer et al. (2019). They were analysed together with the CAPS haplotypes identified for each sample in this study. Haplotypes based on both cpSSR and CAPS polymorphisms were inferred using MS Excel.

2.7 Data analyses

The cpSSR haplotype data were converted into a binary 0/1-matrix using MS Excel to combine them with CAPS markers for haplotype network analysis. The Arlequin v. 3.5 software (Excoffier and Lischer 2010) was used to compute the haplotype network based on both cpSSR and CAPS markers.

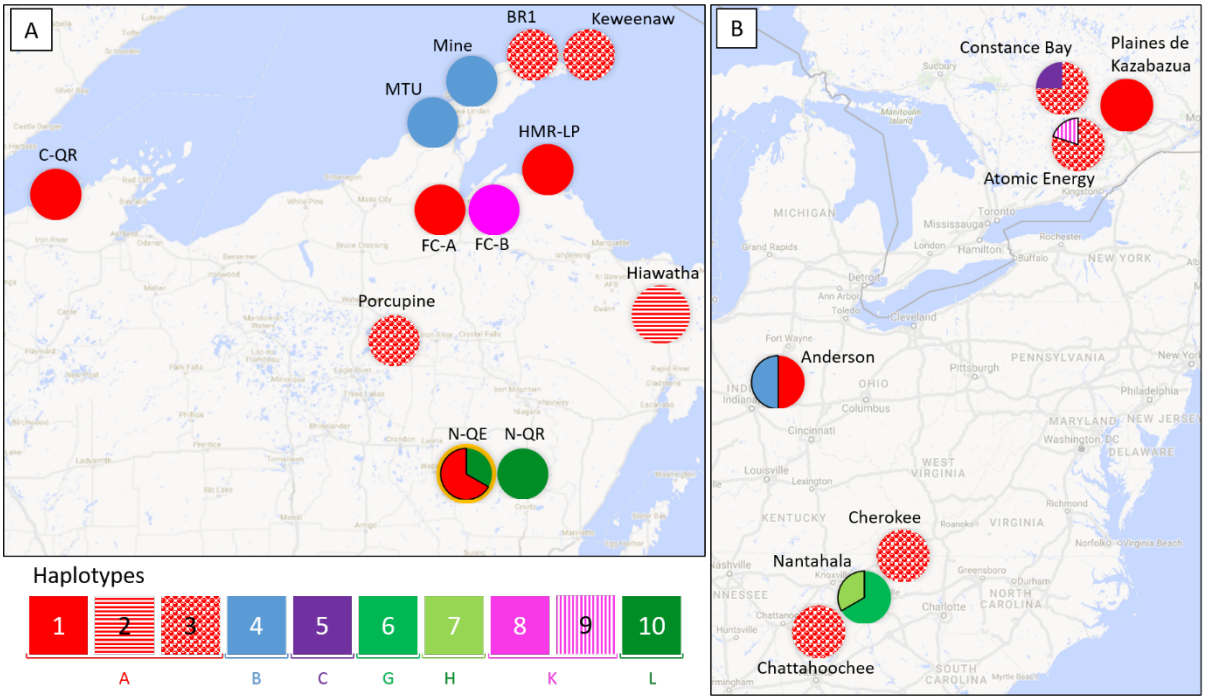


Figure 3 Haplotypes identified in the studied populations of *Quercus rubra*. Numbers indicate haplotypes based on both cpSSR and CAPS markers (see also Fig. 3 and Table 4). Letters indicate cpSSR-based haplotypes. **A** –Haplotypes identified in the area of the Great Lakes. Population of *Q. ellipsoidalis* is marked with a yellow circle. **B** - Haplotypes identified in North American populations (see Tables 1 and 4 for details). The cpSSR haplotype A was further separated into haplotypes 1, 2, and 3, and the cpSSR haplotype K was separated into haplotype 8 and 9 (see also Table 4).

3 Results

Depending on the sample, between 87 % and 97 % of the reference genome was sequenced with coverage between 4,030X and 6,297X (Table 3). In total, 118 SNPs and 107 indels were identified, which were subsequently used to identify restriction sites and design CAPS markers. In this study, we selected only restriction sites that were present in one provenance while absent in the others. The PCR primers were designed to target only a single restriction site to simplify the scoring procedure. From 23 primer-enzyme combinations tested initially, 15 showed good results and were validated in a larger dataset (Table 4). These 15 CAPs markers together with cpSSR markers developed by Pettenkofer et al. (2019) increased haplotype resolution within the natural range of *Q. rubra* and allowed to discover 10 different haplotypes (Table 5). The most frequent cpSSR haplotype A splits up into three different haplotypes, while haplotype K into two haplotypes (Figs 3 and 4, Table 5). All developed CAPS markers were successfully amplified also in samples of *Q. ellipsoidalis* (population N-QE).

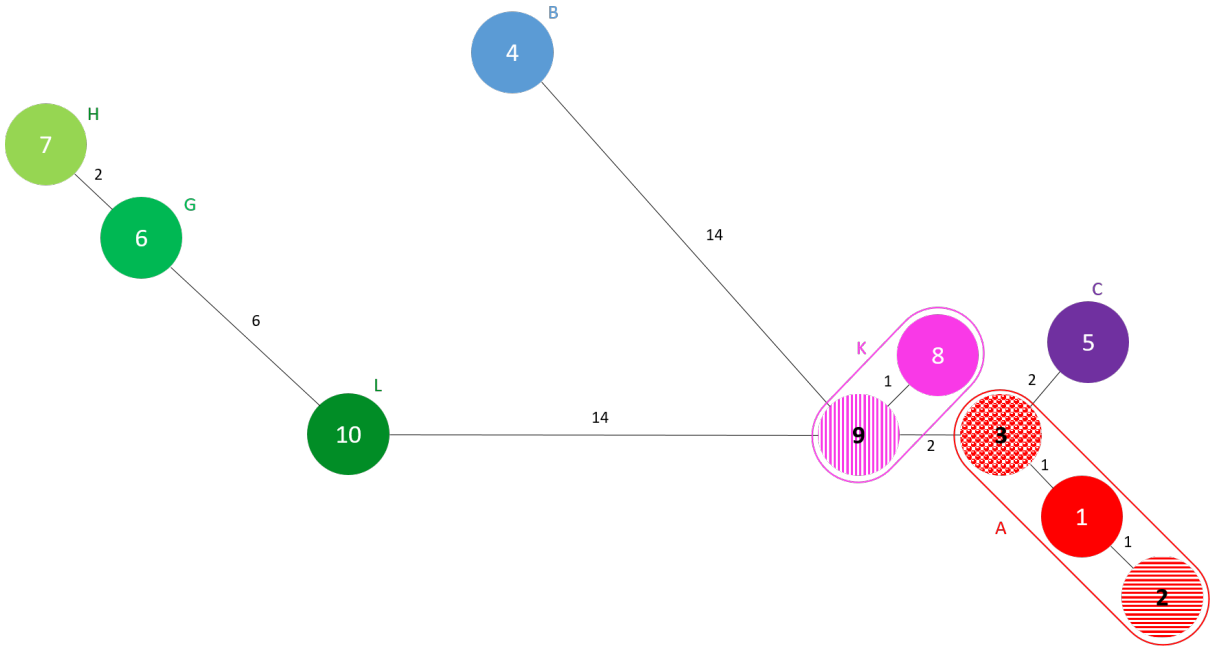


Figure 4 Minimum spanning tree of combined haplotypes based on both cpSSR and CAPS markers (indicated by numbers within circles). CpSSR haplotypes are indicated by letters. Numbers next to lines indicate the number of differences between two haplotypes.

Table 4 Primer-enzyme combinations designed and tested in the study

#	Primer pair	Nucleotide position in the reference genome[1]		Amplicon size, bp	Restriction enzyme	SNP position in the reference genome	Number of fragments after cutting	Sense sequence (5'-3')	Antisense sequence (3'-5')
		start	end						
1	1.1	47	1181	1134	*HinP1*I	557	2	-	*psbA* exon
2	1.2	97	496	399	*Aci*I	200	2	-	intergenic *trnH-psbA*
3	3.2	12554	15093	2539	*BmgB*I	12918	2	-	*atpA* exon
4[2]	4.1	16670	17890	1220	*Hae*III	17341	2[3]	-	*atpI* exon
5[2]	5.2	21853	22378	525	*Aci*I	22029	2	-	*rpoC2* exon
6	8	32795	36929	4134	*Hha*I	35513	2	intergenic *petN-trnT*	intergenic *trnS-trnE*
7	8.1	34539	35674	1135	*Cla*I	34889	2	intergenic *petN-trnT*	intergenic *trnS-trnE*
8	8.2	35372	35565	193	*Mse*I	35426	2	intergenic *petN-trnT*	intergenic *trnS-trnE*
9	11	47541	51141	3600	*EcoR*V	48925	2	-	*ycf3* intron
10[2]	17.1	76010	77115	1105	*Apa*I	76425	2	*rps12*	*clpP* intron, *rps12*
11	26.1	113976	118571	4595	*Swa*I	117302	2	*rps12*	*ndhF* exon
12	26.1	113976	118571	4595	*Pac*I	114509	2	intergenic *trnR-ycf1*	-
13	29	127696	131625	3929	*Eag*I	129513	2	*rps12*	*ndhH* exon
14	29	127696	131625	3929	*Bgl*II	130647	2 or 3	*rps12*	intergenic *ycf1-rps15*
15	30.1	133645	134411	766	*Dra*I	133915	2	*rps12*	*ycf1* exon

Note: [1] NCBI GenBank accession number JX970937.1. [2] Combinations that together with cpSSR markers were sufficient for identification of all 10 haplotypes (see also Table 5). [3] Three fragments were observed in samples from FC-B.

Table 5 Restriction fragments and cpSSR alleles associated with 10 haplotypes

Haplotype			CAPS[1]															cpSSRs[2], bp				
			1.1	1.2	3.2	4.1*	5.2	8	8.1	8.2	11	17.1	26.1	26.1	29	29*	30.1					
CAPS+ cpSSRs	CAPS	cpSSRs	*HinP1*I	*Aci*I	*BmgB*I	*Hae*III	*Aci*I	*Hha*I	*Cla*I	*Mse*I	*EcoR*V	*Apa*I	*Swa*I	*Pac*I	*Eag*I	*Bgl*II	*Dra*I	*ccmp2*	*ccmp4*	*ucd4*	*udt1*	*udt4*
1	1	A	2	1	2	2	1	1	2	1	1	1	2	2	1	2	1	228	116	99	86	145
2	2	A	2	1	2	2	1	1	2	1	1	**2**	2	2	1	2	1	228	116	99	86	145
3	3	A	2	1	2	2	**2**	1	2	1	1	1	2	2	1	2	1	228	116	99	86	145
4	5	B	2	1	2	1	1	1	1	2	1	1	2	2	2	**3**	1	227	115	98	85	146
5	3	C	2	1	2	2	2	1	2	1	1	1	2	2	1	2	1	228	116	99	87	145
6	4	G	1	2	1	2	2	2	2	1	2	1	1	1	1	2	2	226	117	97	85	146
7	4	H	1	2	1	2	2	2	2	1	2	1	1	1	1	2	2	226	117	97	84	146
8	6	K	2	1	2	**3**	2	1	2	1	1	1	2	2	1	2	1	228	116	99	85	145
9	3	K	2	1	2	2	2	1	2	1	1	1	2	2	1	2	1	228	116	99	85	145
10	7	L	1	2	2	2	2	2	2	1	1	1	1	1	1	2	2	226	118	98	85	146

Note: [1] Numbers indicate numbers of fragments observed after digestion by restriction enzymes. [2] Described in Weising und Gardner 1999; Deguilloux et al. 2003). * 4.1-HaeIII and 29-BglII had more than one restriction site and were genotyped as two loci to allow analysis with the Arlequin software. CAPs markers 1.1, 4.1, 5.2, 11, 17.1 (coloured) are sufficient to identify all seven CAPS-based haplotypes. CAPs markers 4.1, 5.2 and 17.1 together with cpSSR markers are sufficient to identify all 10 haplotypes

4 Discussion

Alexander and Woeste (2014) found only six SNPs and 45 indels in the chloroplast genome of four red oak individuals that represented mostly the south and north-east of the natural range. We found much more markers (118 SNPs and 107 indels) in a larger and more diverse sample. This number is rather large considering the conservation of the chloroplast genome. Generally, the number of SNPs increases with taxonomic divergence (Alexander and Woeste 2014; van Nguyen et al. 2018).

As expected, the seven CAPS haplotypes included the main haplotypes previously identified with cpSSR markers (Table 5; Pettenkofer et al. 2019). However, together with cpSSR markers a higher haplotype resolution can be obtained. Some CAPS markers identify the same haplotypes due to the tight linkage and linkage disequilibrium. Therefore, it is important to notice that only three CAPS markers in addition with five cpSSR markers are needed to detect all 10 haplotypes found in our sample (see Table 5). The successful application of all primer-enzyme combinations also in a population of *Q. ellipsoidalis* shows the transferability of these markers. The newly developed CAPS markers can thus be used as a cost-effective and easy way of determining the chloroplast variation in populations of these two oak species and potentially other closely related species.

This opens new possibilities for the study of post-glacial migration routes and for tracing seeding material or identifying the origin of established stands in case if reference information for assignment is available. The obtained genome sequences of trees representing four different regions can also be used to develop additional cpSSR, indel and CAPS markers, and thus contribute to future studies of genetic diversity and structure in oak populations.

Acknowledgments: We thank Alexandra Dolynska for technical assistance and Mirko Liesebach (Thünen Institute) for providing plant material. The study was supported by the German Federal Ministry of Food and Agriculture (BMEL, grant number 22023314).

5 References

Alexander LW, Woeste KE (2014) Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. Tree Genetics & Genomes 10:803–812

Birchenko I, Feng Y, Romero-Severson J (2009) Biogeographical distribution of chloroplast diversity in northern red oak (*Quercus rubra* L.). The American Midland Naturalist 161:134–145

Borkowski DS, McCleary T, McAllister M et al (2014) Primers for 52 polymorphic regions in the *Quercus rubra* chloroplast, 47 of which amplify across 11 tracheophyte clades. Tree Genetics & Genomes 10:885–893

Deguilloux M-F, Dumolin-Lapègue S, Gielly D et al (2003) A set of primers for the amplification of chloroplast microsatellites in *Quercus*. Molecular Ecology Notes 3:24–27

Desmarais KM (1998) Northern red oak regeneration: biology and silviculture, University of New Hamshire

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5. A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567

Finkeldey R, Gailing O (2013) Genetics of chloroplasts. In: Maloy S, Hughes K (eds) Brenner's Encyclopedia of Genetics. Elsevier, pp 525–527

Gailing O, Wachter H, Heyder J et al (2009) Chloroplast DNA analyses of very old, presumably autochthonous *Quercus robur* L. stands in North Rhine-Westphalia. Forst- und Jagdzeitung 180:221–227

Grivet D, Deguilloux M-F, Petit RJ et al (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. Mol Ecol 15:4085–4093

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

Kaundun SS, Matsumoto S (2003) Development of CAPS markers based on three key genes of the phenylpropanoid pathway in tea, *Camellia sinensis* (L.) O. Kuntze, and differentiation between *assamica* and *sinensis* varieties. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 106:375–383

Kibbe WA (2007) OligoCalc. An online oligonucleotide properties calculator. Nucleic Acids Research 35:W43-6

Liesebach M, Schneck V (2011) Entwicklung von amerikanischen und europäischen Herkünften der Roteiche in Deutschland. Development of American and European provenances of northern red oak in Germany. Forstarchiv 82:125–133

Lind JF, Gailing O (2013) Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genetics & Genomes 9:707–722

Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a CONSTANS-like gene between two red oak species. Annals of Botany 113:967–975

Little EL (1999) Range map of *Quercus rubra*. U.S. Geological Survey

Magni CR, Ducousso A, Caron H et al (2005) Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. Mol Ecol 14:513–524

Nagel R-V (2015) Roteiche (*Quercus rubra* L.). In: Vor T, Spellmann H, Bolte A et al (eds) Potenziale und Risiken eingeführter Baumarten. Göttingen University Press, Göttingen

Petit RJ, Brewer S, Bordács S et al (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management 156:49–74

Petit RJ, Kremer A, Wagner DB (1993) Geographic structure of chloroplast DNA polymorphisms in European oaks. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 87:122–128

Pettenkofer T, Burkardt K, Ammer C et al (2019) Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison to reference native North American populations. Eur J Forest Res 10

van Nguyen B, Linh Giang VN, Waminal NE et al (2018) Comprehensive comparative analysis of chloroplast genomes from seven Panax species and development of an authentication system based on species-unique single nucleotide polymorphism markers. Journal of Ginseng Research

Weising K, Gardner RC (1999) A set of conserved PCR Primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Zhang R, Hipp AL, Gailing O (2015) Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations. Botany 93:691–700

Supporting information: All SNPs and indels identified in the *Quercus rubra* chloroplast genome, as well as tested CAPS markers are listed in the supplemental Microsoft Excel file.

Paper 4

Genetic variation of introduced red oak (*Quercus rubra*) stands in Germany in comparison to North American populations

Published

European Journal of Forest Research

18 January 2020

doi:10.1007/s10342-019-01256-5

Genetic variation of introduced red oak (*Quercus rubra*) stands in Germany in comparison to North American populations

Tim Pettenkofer[1] • Reiner Finkeldey[2] • Markus Müller[1] • Konstantin V. Krutovsky[1,4,5,6,7] • Barbara Vornam[1] • Ludger Leinemann[1] • Oliver Gailing[1,3,7]

Corresponding authors: Oliver Gailing[1,3,7], E-mail: ogailin@gwdg.de; Ludger Leinemann[1], E-mail: lleinem@gwdg.de, Tel: +49 551 3933536

[1] University of Göttingen, Faculty of Forest Sciences and Forest Ecology, Forest Genetics and Forest Tree Breeding, Büsgenweg 2, 37077 Göttingen, Germany

[2] University of Kassel, Mönchebergstraße 19, 34125 Kassel, Germany

[3] Michigan Technological University, School of Forest Resources and Environmental Science, Houghton, MI 49931, United States

[4] Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 660036 Krasnoyarsk, Russia

[5] Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia

[6] Department of Ecosystem Science and Management, Texas A&M University, 2138 TAMU, College Station, TX 77843-2138, USA

[7] University of Göttingen, Center for Integrated Breeding Research (CiBreed), 37075 Göttingen, Germany

Keywords: *Quercus rubra*, introduced species, nuclear microsatellites, EST, genetic variation

Abstract

Although Northern red oak (*Quercus rubra* L.) is the most important introduced deciduous tree species in Germany, only little is known about its genetic variation. For the first time, we describe patterns of neutral and potentially adaptive nuclear genetic variation in Northern red oak stands across Germany. For this purpose, 792 trees were genotyped including 611 trees from 12 stands in Germany of unknown origin and 181 trees from four populations within the natural distribution area in North America. Our marker set included 12 potentially adaptive (expressed sequence tag-derived simple sequence repeat = EST SSR) and 8 putatively selectively neutral nuclear microsatellite (nSSR) markers. Our results showed that German stands retain comparatively high levels of genetic variation at both EST-SSRs and nSSRs, but are more similar to each other than to North American populations. These findings are in agreement with earlier chloroplast DNA analyses which suggested that German populations originated from a limited geographic area in North America. The comparison between potentially adaptive and neutral microsatellite markers did not reveal differences in the analyzed diversity and differentiation measures for most markers. However, locus FIR013 was identified as a potential outlier locus. Due to the absence of signatures of selection in German stands, we suggest that introduced populations were established with material from provenances that were adapted to environmental conditions similar to those in Germany. However, we analyzed only a limited number of loci which are unlikely to be representative of adaptive genetic differences among German stands. Our results suggest that the apparent introduction from a limited geographic range in North America may go along with a reduced adaptive potential.

1 Introduction

The introduction of Northern red oak (*Quercus rubra* L.) from its natural range in North America to Europe dates back to the end of the seventeenth century, when it was likely first brought to France (Houba 1887; Hickel 1932). Until the middle of the eighteenth century, *Q. rubra* was planted in parks and botanical gardens for ornamental purposes (Bauer 1951; Nagel 2015). After two major waves of cultivation in the second half of the nineteenth century and at the turn of the nineteenth to the twentieth century, it is now the most important non-native deciduous tree species for wood production in Germany (Bauer 1951; Bundesministerium für Ernährung und Landwirtschaft (BMEL) 2014). Nagel (2015) described *Q. rubra* as a species covering a wide range of soil and climatic conditions with annual precipitation between 600

and 2000 mm and mean annual temperatures between 4 and 15 °C and featuring a shorter rotation period (~ 80–120 years) than native white oak species (~ 140 years), as well as a lower demand for nutrients and water.

Unlike other European countries (Möllerová 2005; Riepšas and Straigytė 2008; Chmura 2013), *Q. rubra* is not considered invasive in Germany. It is less shade-tolerant than the main native tree species Fagus sylvatica L. and only little more shade tolerant than the native white oak species *Q. robur* L. (Vor and Lüpke 2004; Niinemets and Valladares 2006; Nagel 2015). Moreover, it is easily controllable by containment measures (Vor 2005; Nagel 2015).

Introduced species are usually expected to experience a founder effect and genetic bottlenecks that greatly promote genetic drift and can result in decreased genetic variation, especially if only a limited numbers of trees served as seed source and/or only a small part of the species' natural range was sampled (Nei et al. 1975; Barrett and Husband 1990). In the past, several studies were conducted to reveal patterns of genetic variation within Northern red oak's natural range (e.g., Daubree and Kremer 1993; Romero-Severson et al. 2003; Magni et al. 2005; Zhang et al. 2015; Borkowski et al. 2017). However, only few studies focussed on the impact of the introduction on genetic variation of Northern red oak populations in Europe (Magni Diaz 2004; Merceron et al. 2017; Pettenkofer et al. 2019).

Borkowski et al. (2017) found that the genetic differentiation between Northern red oak populations increased from south to north within the natural range, reflecting its postglacial migration movement, but without revealing distinct pathways. Magni Diaz (2004) used chloroplast (cp) DNA-based PCR–RFLP (restriction fragment length polymorphism) markers to analyze the genetic variation of introduced stands mainly located in France and Germany with only a few samples from the Netherlands, Belgium, Spain, Italy, and Romania. Interestingly, Magni Diaz (2004) found greater total and within-population diversity as well as differentiation between populations for historical German populations in comparison with historical French populations potentially as a result of different import and forest policies in these countries. However, for all European populations, Magni Diaz (2004) could not identify a geographic pattern. Recently, Pettenkofer et al. (2019) used maternally inherited chloroplast DNA markers to analyze the genetic variation of German stands in comparison with North American populations. Most German stands showed very similar haplotype frequencies and low haplotype diversity, pointing to a limited number of seed sources. A considerably higher cpDNA haplotype diversity was detected in Southern Germany. Multiple introductions and

admixture of material within Europe or prior to its introduction to Europe were suggested as possible reasons for high haplotype diversity within certain regions (Magni Diaz 2004; Pettenkofer et al. 2019). While it was not possible to narrow down the geographic origin to specific regions in North America, patterns of genetic variation suggested an origin from a limited geographic area, likely in the northern part of the species' distribution range (Magni Diaz 2004; Pettenkofer et al. 2019).

Merceron et al. (2017) studied single-nucleotide polymorphism (SNP) markers randomly distributed across the genome in European populations, mainly in France, but also including samples from Germany, the Netherlands, Belgium, Spain, Italy, and Romania, and compared them with data for North American populations. Three main genetic clusters were identified in North America: in the south, the northeast, and the northwest, respectively. However, only trees representing the two northern clusters were found in Europe. Merceron et al. (2017) stated that trees representing the southern cluster were either never introduced to Europe or vanished eventually. They suggested that European populations may thus originate from the northern part of the natural range, a conclusion supported by other studies on this topic (Bauer 1954; Magni Diaz 2004; Nagel 2015; Pettenkofer et al. 2019).

This study used some of the same populations that were analyzed earlier with chloroplast SSR markers (Pettenkofer et al. 2019). In contrast to chloroplast SSR markers, which are maternally inherited in *Q. rubra* and therefore useful to track migration routes (Petit et al. 1997; Alexander and Woeste 2014), nSSR markers are biparentally inherited and much more polymorphic. However, nSSR markers require a considerably larger set of samples per population. For the first time, we describe patterns of neutral and potentially adaptive nuclear genetic variation in Northern red oak stands across Germany.

The main objective of the present study was to assess the genetic variation of German Northern red oak stands by analyzing variable potentially adaptive (expressed sequence tag-derived simple sequence repeats = EST SSR) and putatively selectively neutral nuclear microsatellite (nSSR) markers and to compare German stands with selected North American populations.

We hypothesize that (1) the *Q. rubra* gene pool introduced to Germany represents only a fraction of the North American gene pool, and (2) potentially adaptive genetic markers show different variation patterns compared to neutral markers.

2 Material and Methods

2.1 Plant Material

In total, 792 trees were genotyped in this study including 611 trees from 12 stands in Germany of unknown origin (48–57 samples per stand; Table 1, Fig. 1) and 181 trees from four populations within the natural distribution area in North America (40–47 samples per population; Table 2, Fig. 2). The selection of the study sites was based on the chloroplast haplotype frequencies identified earlier in these populations based on smaller sample sizes (5–10 samples per population) (Pettenkofer et al. 2019). Due to the strict maternal inheritance and lower variation of cpDNA, these sample sizes allowed to characterize 39 stands in Germany and 19 populations in the USA and Canada. In the study carried out by Pettenkofer et al. (2019), most German stands had the most common haplotype "A." The next frequent haplotype "B" was found exclusively in the northern parts of the natural range. Two of the three haplotype

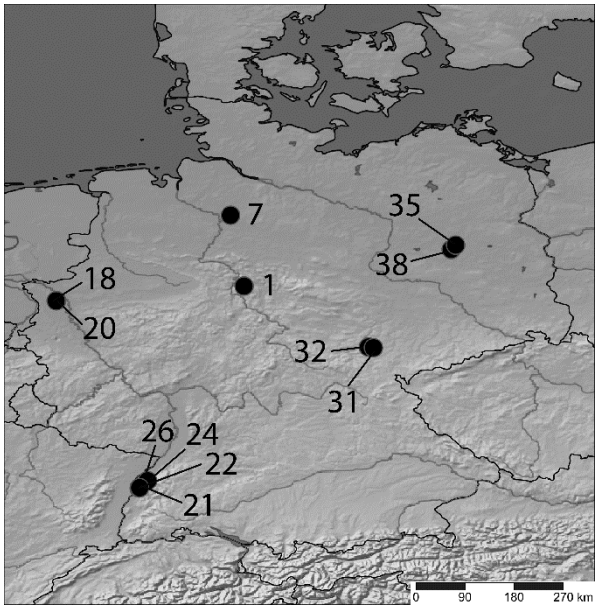


Fig. 1 Study sites in Germany

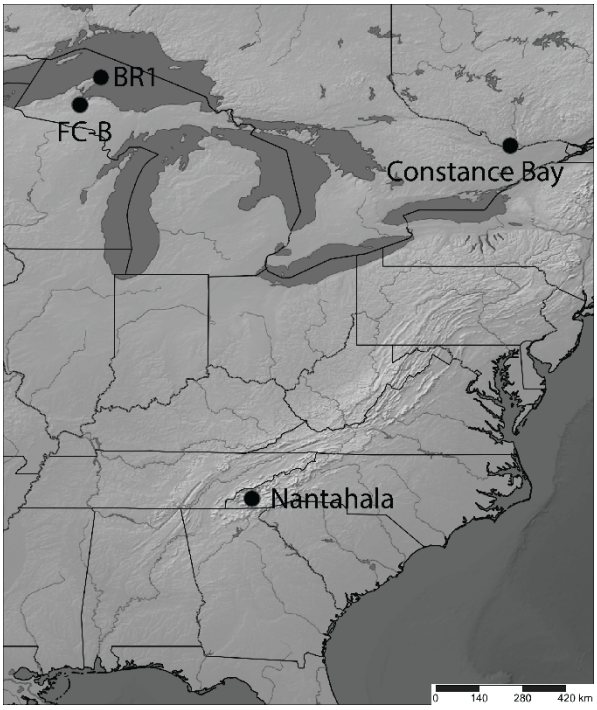


Fig. 2 Populations in North America

clusters identified in the natural range were also found in German populations. Stands located in the southwest of Germany had a higher haplotype variation compared to other sampled stands. The populations within the natural range used in this study were selected from the northern and southern parts of the range that were different based on the chloroplast haplotypes, and samples were available in sufficient numbers for them. Samples were taken randomly in each population as buds, green leaves, or cambium probes.

Table 1 Study sites in Germany as described in Pettenkofer et al. (2019)

#	Land	Region	District	*N*	Latitude	Longitude	Altitude, m
1	Lower Saxony	Dassel	Sievershausen	57	51.77971	9.58899	524
7		Rotenburg	Diensthoop	53	52.94624	9.36675	69
18	North Rhine-	Niederrhein	Leucht 5	52	51.53741	6.49099	68
20	Westphalia		Leucht 7	49	51.53946	6.49804	72
21	Baden-	Offenburg	Schutterwald 1	58	48.45871	7.86786	158
22	Württemberg		Schutterwald 2	49	48.45346	7.85918	148
24			Schutterwald 4	45	48.45299	7.85969	148
26		Oberkirch	Renchen	54	48.57306	7.99869	171
31	Thuringia	Neustadt	Strößwitz 2	48	50.76442	11.7242	392
32		Jena	Holzland	49	50.77549	11.64021	372
35	Brandenburg	Potsdam	Güterfelde 1	48	52.45505	13.0795	47
38			Güterfelde 3	49	52.38356	13.00608	38

Within Germany, sample stands were selected in five different federal states of Germany (Table 1): two stands in each of Lower Saxony, North Rhine-Westphalia, Brandenburg, and Thuringia and four stands in Baden-Wuerttemberg, where Pettenkofer et al. (2019) discovered the highest chloroplast variation. All German stands were (1) pure Q. rubra stands, (2) 50–80 years old, (3) featured a rectangular shape for easier data acquisition, and (4) both the present and expected future climate conditions matched their autecological properties.

2.2 DNA isolation

The DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) was used to extract the DNA from either about 1 cm2 leaf tissue of the fresh leaf or 1–2 whole buds from a fresh twig per tree.

Extraction of DNA from cambium samples was performed with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) using the respective manufacturer's protocol with the following changes: The incubation of cambium samples took place overnight at 65 °C in a lysate containing the buffer AP1 and 9.4% polyvinylpyrrolidone (PVP) in the final solution.

Table 2 Study sites in North America

Abbreviation	Region	N	Latitude	Longitude	Altitude, m
BR1[a]	Brockway Mountain, MI	48	47.46616	-87.91671	355
FC-B[a]	Ford Forestry Center, MI	48	46.67442	-88.52427	393
Constance Bay[b]	Ottawa, Constance Bay	48	45.49094*	-76.07482*	60*
Nantahala[b]	Clay, Nantahala, Tusquitee	48	35.27477*	-83.54827*	630*

Note: MI - Michigan; WI – Wisconsin; [a] Lind and Gailing 2013; [b] Liesebach and Schneck 2011; * approximation

2.3 EST- and neutral nuclear SSR markers

For initial testing, we selected 25 nuclear microsatellite markers arranged in 4 multiplexes (Supplementary Table 1S). Five of them (FIR043, FIR004, quru-GA-2M04, quru-GA-2F05, and GOT037) were removed due to difficulties in amplification or allele scoring. Finally, we used 12 EST-SSR and 8 neutral nSSR markers (Aldrich et al. 2002; Durand et al. 2010; Sullivan et al. 2013; Supplementary Table 1S).

Before the preparation of the specific multiplexes, each forward and reverse primer was diluted to a concentration of 5 pM/µl. Except for multiplex 4 (as explained further below), the PCR mix contained for each sample the following components: 1 µl of genomic DNA (about 0.6 ng/µl), 1.5 µl PCR buffer (containing 0.8 M Tris–HCl and 0.2 M $(NH_4)_2SO_4$), 1.5 µl $MgCl_2$ (25 mM), 1 µl of each dNTP (2.5 mM), and 1 U HOT FIREPol®Taq-polymerase (Solis BioDyne; Tartu, Estonia). In addition, the following components were added to their specific multiplex in the denoted amount: multiplex 1–0.8 µl ddH_2O and 1 µl of each forward and reverse primer for the markers quru-GA-1P10, quru-CA-2P24, quru-CA-3A05, and quru-CA-3D15, respectively; multiplex 2–0.8 µl ddH_2O and 1 µl of each forward and reverse primer for the markers FIR043, FIR035, FIR028, and FIR013, respectively; multiplex 3—two separate PCRs were performed—(1) 4.8 µl ddH_2O and the denoted amount for each forward and reverse primer for the markers quru-GA-0E09 (1 µl), quru-GA-C06 (0.6 µl), and quru-GA-F07 (0.4 µl), (2), 3.8 µl ddH_2O and the denoted amount for each forward and reverse primer for the markers GOT021 (0.8 µl), FIR024 (0.8 µl), FIR031 (0.5 µl), and quru-GA-2M04 (0.4 µl). The products of the last two PCRs were pooled before electrophoretic separation.

Multiplex 4 was performed with the following PCR mix per sample: 8.5 µl Multiplex PCR Kit (Qiagen, Hilden, Germany) and the denoted amount for each forward and reverse primer for the markers PIE040 (1 µl), GOT037 (1 µl), quru-GA-2F05 (0.6 µl), quru-GA-0C11 (0.4 µl), FIR104 (0.4 µl), PIE125 (0.4 µl), VIT107 (0.4 µl), VIT023 (0.4 µl), and GOT040 (0.3 µl).

The same touchdown PCR program was used for all markers in a Biometra TProfessional thermocycler (Jena, Germany). The PCR protocol started with 15 min for initial denaturation at 95 °C, followed by 10 cycles of 1-min denaturation at 94 °C, 1-min annealing at 60 °C (− 1 °C per cycle), and 1-min extension at 72 °C. This first set of cycles was then followed by another 25 cycles of 1-min denaturation at 94 °C, 1-min annealing at 50 °C, and 1-min extension at 72 °C. The PCR ended with a final 20-min extension step.

Before adding to HiDi formamide, the PCR products were diluted for capillary gel electrophoresis based on band intensities on agarose gels in the ratios of 1:150 for multiplexes 1 and 2 and 1:60 for multiplexes 3 and 4. The SSR fragments were separated using capillary electrophoresis on an ABI Genetic Analyzer 3130xl (Applied Biosystems, Foster City, USA) and sized using the internal size standard GeneScan™ 500 ROX™ as reference for multiplexes 1 and 2 and GeneScan™ 500 LIZ™ as reference for multiplexes 3 and 4 (Applied Biosystems, Foster City, USA). The fragments were scored using the software package GeneMapper version 3.7 (Applied Biosystems, Foster City, USA).

2.4 Data analyses

Based on the obtained SSR genotypes, the software GenAlEx 6.5 (Peakall and Smouse 2006, 2012) was used to calculate the number of alleles N_a, the effective number of alleles N_e, the observed heterozygosity H_o, the expected heterozygosity H_e, and the genetic differentiation measures D (Jost 2008) and F_{ST} for all markers and populations. We chose to calculate both differentiation parameters because traditional F_{ST} can underestimate differentiation for such highly polymorphic markers as SSRs (Jost 2008). The fixation index F_{IS} and the test for significant differences of its values from 0 were calculated with FSTAT 2.9.3 with P-values adjusted using the sequential Bonferroni procedure (Rice 1989; Goudet 1995). The Kruskal–Wallis test was used to identify significant differences between German and North American populations for all diversity parameters (Table 3) using the software R version 3.3.2 (R Core Team 2016). Also, the Kruskal–Wallis test with multiple comparisons implemented in the R-package “pgirmess” (Giraudoux et al. 2018) was used to test for pairwise differences among all populations. The implemented multiple-comparison test determines which population pair shows significant differences. To visualize population differences, we further used GenAlEx 6.5 to perform a principal coordinates analysis (PCoA). This cluster analysis was based on the pairwise genetic distance matrix between populations and between individuals. The PCoA assigns a location for each individual or population within a multidimensional space. Two or

three first main axes that explain most of the genetic differentiation between individuals and/or populations are usually presented in a plot based on this analysis (Peakall and Smouse 2012). In this study, the genetic distance matrix was based on Jost's *D* (Jost 2008) calculated using GenAlEx 6.5.

To build a neighbor-joining tree (NJT), the genetic distance D_A by Nei et al. (1983), which is especially suited for microsatellite markers (Takezaki and Nei 1996), was calculated using the Populations 1.2.32 software (Langella 1999). The bootstrap values were based on 1000 permutations across loci, and the NJT was visualized using the online software IcyTree (Vaughan 2017).

Arlequin version 3.5 (Excoffier and Lischer 2010) was also used to perform an analysis of molecular variance (AMOVA) by applying 9999 permutations for all populations (the entire dataset), all German populations, and the four North American populations (Excoffier et al. 1992). The software computes a matrix of pairwise distances between all populations using the number of different alleles (F_{ST}-like). Further, Arlequin was also used to search for outlier loci that could be potentially under selection (Excoffier et al. 2009). Here, 50,000 simulations and 100 demes were selected as running conditions.

To make inferences about the potential population structure and number of clusters that can be identified in our genotyped samples, we used the software STRUCTURE 2.3.4 by Pritchard et al. (2000). STRUCTURE uses a model-based statistical clustering method within a Bayesian framework to assign individuals to populations and detect population structure over all sampled multilocus genotypes (Pritchard et al. 2000; Falush et al. 2003). We used 10,000 and 100,000 Markov chain Monte Carlo (MCMC) replicates for the burn-in period and further iterations, respectively. Runs were performed for 1–20 potential clusters (*K*) using 20 iterations for each test. To infer the optimal number of *K*, we applied the ΔK method by Evanno et al. (2005) using the STRUCTURE HARVESTER 0.6.94 software (Earl and Von Holdt 2012). Results were visualized using the online software CLUMPAK (Kopelman et al. 2015).

To test for genetic bottleneck effects, the software BOTTLENECK (Cornuet and Luikart 1996) was used with 1000 iterations and assuming a two-phase model (TPM). The TPM allows to assume different proportions (contributions) of the infinite allele model (IAM) and the stepwise mutation model (SMM) for all markers. Considering that dinucleotide and other perfect repeats follow the IAM while imperfect repeats rather follow the SMM, we selected the

option with 70 % IAM and 30 % SMM in accordance with our marker set (Cornuet and Luikart 1996; Cristescu et al. 2010).

3 Results

3.1 Genetic diversity

The German stands and North American populations showed similar levels of mean observed (H_o = 0.61 vs. 0.59) and expected (H_e = 0.63 vs. 0.65) heterozygosities, respectively (Table 3). These parameters were also similar across all populations (H_o = 0.57–0.62, H_e = 0.62–0.68, Table 3). The mean (N_a) and effective (N_e) number of alleles also showed no significant differences between German and North American populations (Table 3). However, the number of private alleles and their mean frequency in Germany (9 in 12 populations, N_p = 0.75) were significantly lower than those in the North American populations (19 in 4 populations, N_p = 4.75) (Tables 3 and 4). In Germany, 5 of 9 private alleles were found in stands in Baden-Wuerttemberg, while two were found in each Brandenburg and Lower Saxony (Table 4). We detected no private alleles in stands from Thuringia or North Rhine-Westphalia. In North America, most private alleles were found in BR1 (N_p = 9) and Constance Bay (N_p = 5); both populations are located in the north of the natural distribution range. Relative frequencies of private alleles range from 0.009 to 0.042 (Table 4). Among all populations and tested parameters (N_a, N_e, H_o, H_e, F_{IS}), only F_{IS} showed significant differences between the populations 18 (North Rhine-Westphalia) and BR1. No recent genetic bottleneck was detected using the BOTTLENECK software.

The search for outlier loci that could be potentially under selection identified only the EST-SSR FIR013 as a significant outlier locus (Supplementary Table 2S and Fig. 3). At this marker, the North American population BR1 shows a higher frequency (0.24) of allele 138 in comparison with other populations (0.01–0.11).

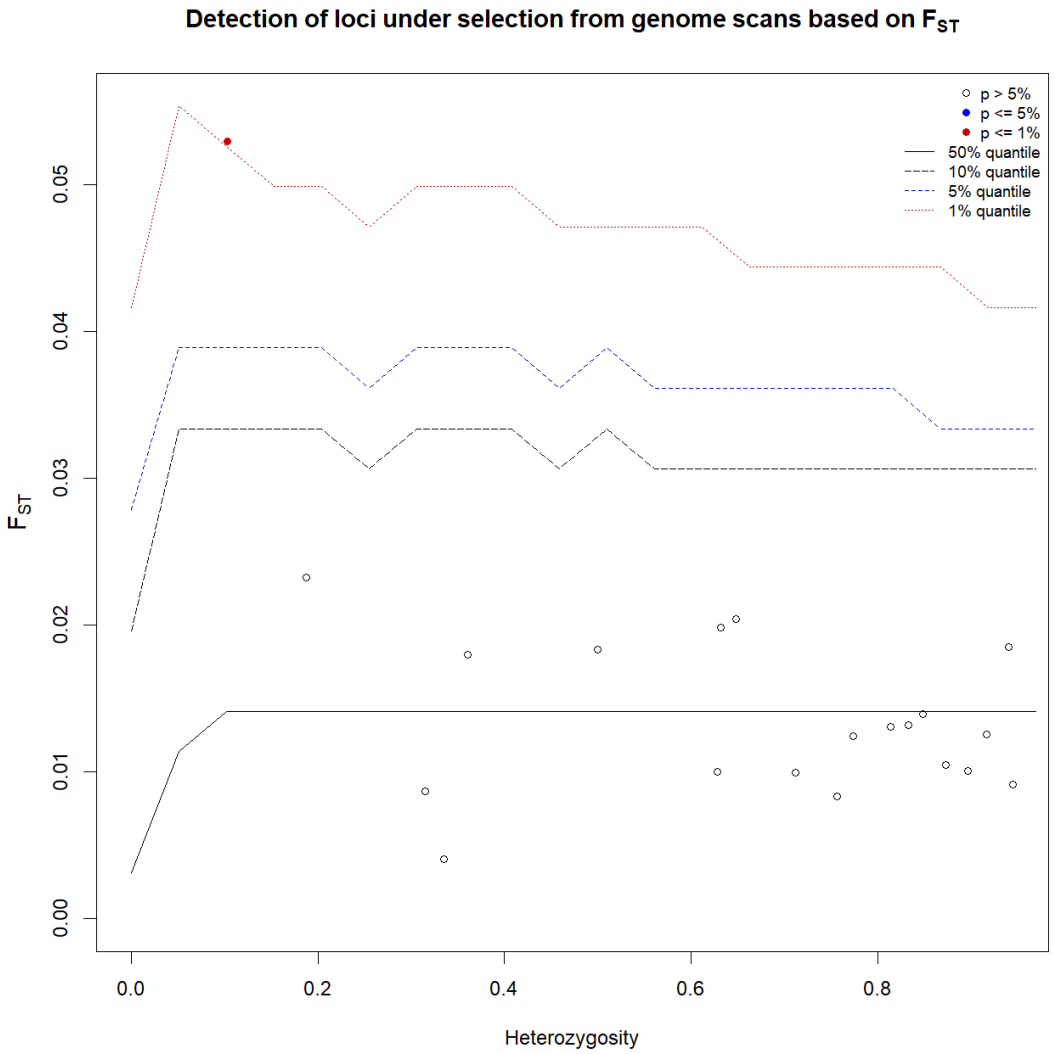


Fig. 3 Detection of outlier loci using the Arlequin software

Table 3 Mean diversity parameters for all populations based on 20 nuclear microsatellite markers (SSRs)

Population	*N*	N_a	N_e	N_p	H_o	H_e	F_{IS}	F_{ST}	Jost's *D*
1 ■	57	9.35	4.65	0	0.58	0.63	0.074	0.010	0.016
7 ■	53	9.40	5.13	2	0.61	0.63	0.040	0.012	0.020
18 ●	52	8.75	4.49	0	0.61	0.62	0.025	0.016	0.027
20 ●	49	9.10	4.57	0	0.62	0.62	0.017	0.012	0.019
21 ◆	58	9.75	4.92	0	0.61	0.64	0.053	0.013	0.021
22 ◆	49	9.35	4.35	1	0.59	0.63	0.072	0.013	0.024
24 ◆	45	9.00	4.55	2	0.61	0.64	0.059	0.009	0.016
26 ◆	54	9.35	4.89	2	0.62	0.64	0.043	0.017	0.027
31 ◆	48	9.95	5.11	0	0.61	0.64	0.055	0.008	0.013
32 ◆	49	9.25	4.74	0	0.60	0.62	0.038	0.011	0.019
35 ▲	48	9.65	5.34	1	0.61	0.66	0.077	0.011	0.018
38 ▲	49	9.90	5.38	1	0.61	0.64	0.047	0.010	0.014
Mean (Germany)	**51**	**9.40**	**4.84**	**0.75**	**0.61**	**0.63**	**0.050**	**0.012**	**0.020**
BR1 ◆	48	10.50	5.16	9	0.60	0.68	0.126	0.024	0.050
Constance Bay ◆	48	9.80	5.01	5	0.61	0.64	0.065	0.018	0.032
FC-B ◆	48	8.65	4.56	2	0.59	0.64	0.099	0.039	0.032
Nantahala ◆	48	8.80	4.55	3	0.57	0.62	0.097	0.018	0.031
Mean (North America)	**48**	**9.44**	**4.82**	**4.75**	**0.59**	**0.65**	**0.097**	**0.025**	**0.036**
P		**0.90**	**0.95**	**<0.01**	**0.08**	**0.41**	**0.011**	**<0.01**	**<0.01**

Note: *N* – number of samples, N_a – mean number of alleles, N_e – mean effective number of alleles, N_p – number of private alleles, H_o – observed heterozygosity, H_e – expected heterozygosity, F_{IS} – fixation index (all values were not significantly different from 0), F_{ST} (over all loci), Jost's *D* (over all loci); *P* – probability of difference between German and North American populations

3.2 Genetic differentiation and population structure

The PCoA based on Jost's *D* between all populations showed that German stands densely clustered between the North American population FC-B, which is located in the northwest of the natural range (Ford Forestry Center, Michigan), and Nantahala, which is located in the south of the range, in the western parts of North Carolina (Fig. 4). Constance Bay (Canada) and BR1 (Brockway Mountain, Michigan) were set apart from the German cluster. The first two principal coordinates together explained 66.32 % of the variation among populations. Also, the individual-based PCoA revealed more heterogeneous point clouds for the populations of North American origin as compared to the German plantations. Especially, the two natural populations from Northern Michigan occupy a comparatively wide space in the two-dimensional PCoA (Supplementary figures 5S and 6S).

To partition genetic variation among and within populations, a hierarchical AMOVA was performed for all populations (the entire dataset) and separately for the North American populations and all German populations (Supplementary Tables 3S, 4S, and 5S). In the entire dataset, 1.10 % of the variation was due to partition among populations, while 98.90 % was due to variation within populations (among and within individuals), with all components being

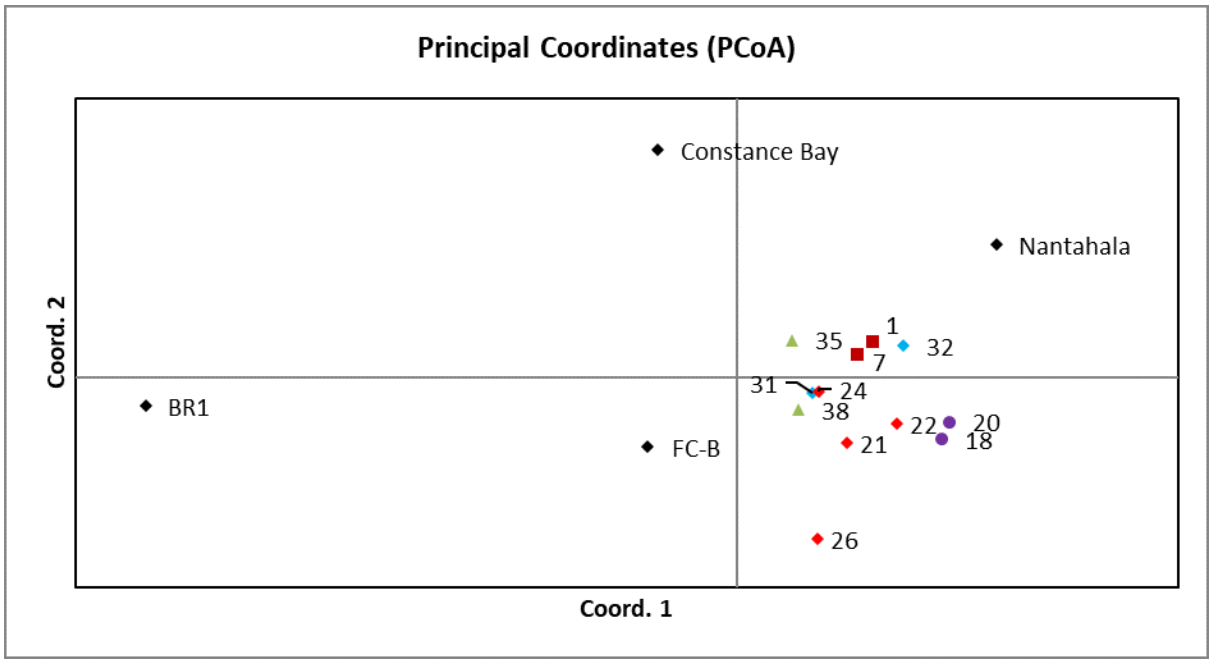


Fig. 4 Principal coordinate analysis (PCoA) of all populations (both coordinates together explain 66.32 % of the variation between populations)

highly significant. In North America, 1.88 % of the variation was due to partition among populations ($p < 0.001$), and 98.12 % of the variation was due to variation within populations. In German populations, the genetic variation among populations was lower than among North American populations (0.78 %).

The STRUCTURE and Delta *K* analyses of the entire dataset, as well as for North American and German populations, indicated no strong population structure (Supplementary figures 1S–4S).

The NJT did not reveal strong differentiation among German populations (Fig. 5) and was in agreement with the STRUCTURE and PCoA analyses. Most clusters were not well supported by bootstrap values. However, all four North American populations, BR1, FC-B, Nantahala, and Constance Bay clustered together. The German population 31 clustered close to Nantahala and Constance Bay populations, but this relationship was not well supported.

The comparison of diversity and differentiation measures between potentially adaptive EST- markers and supposedly selectively neutral nSSRs is presented in Supplementary Table 2S. Unlike nSSR markers, EST-SSRs are more conserved being located very close to functional genes which variation is very likely to be limited due to selection. Thus, they feature higher transferability across species within genera than nSSR markers but are not as polymorphic as nSSR markers (Ellis and Burke 2007; Kalia et al. 2011). Accordingly, for our dataset, mean allelic richness was lower for the potentially adaptive EST-SSR markers (N_a = 5.63) than for the neutral nSSR markers (N_a = 15.04). This pattern was also observed for the observed and expected heterozygosity (neutral: H_o = 0.798, H_e = 0.850; EST: H_o = 0.474, H_e = 0.494). Although we identified EST-SSR marker FIR013 as an outlier locus, in general, F_{ST} and Jost's *D* were similar for potentially adaptive EST-SSR and neutral nSSR markers.

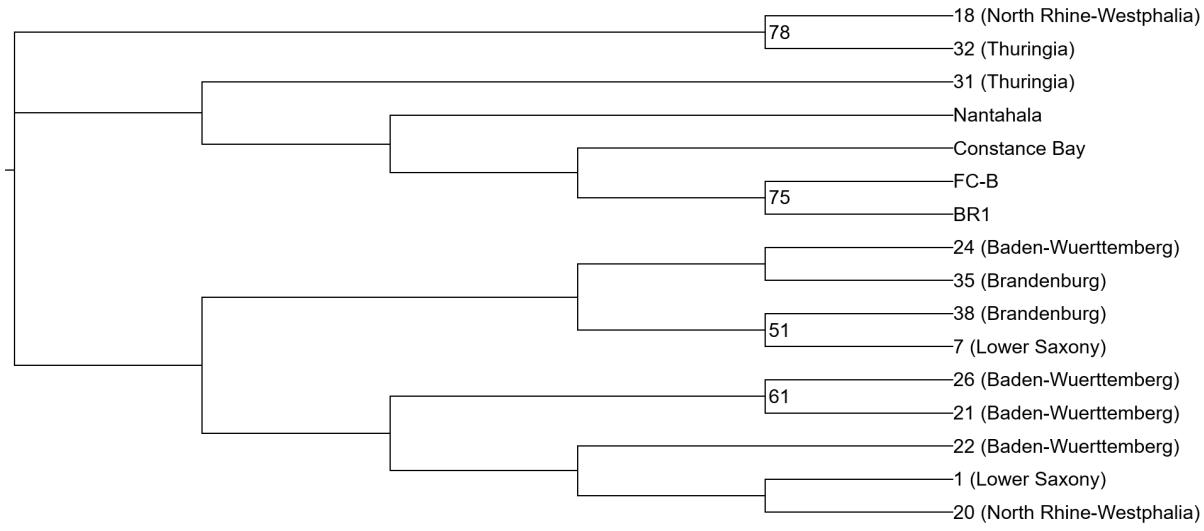


Fig. 5 Neighbor-joining tree (NJT) of the analysed populations. Only bootstrap numbers above 50 % are shown.

4 Discussion

4.1 Genetic variation patterns indicate the origin of German populations from a limited area

Our results show that 19 rare alleles (relative frequencies = 0.010–0.042), which are private in the examined North American populations, are not found in German populations. These drift-like effects can be attributed to the selection and sampling of seed sources. The occurrence of other rare alleles in German stands and the uniform values of H_o and H_e across populations show no signs of a recent reduction of the overall genetic variation in German stands. Moreover, H_o and H_e remained uniform across all populations showing no signs of recent bottleneck effects. The lack of indications for recent bottleneck effects does not exclude that European Northern red oak populations might have undergone bottleneck effects in Germany at all. A genetic bottleneck could have been induced a long time ago, for instance due to selective sampling within the native range. However, admixture of material within Europe and multiple introductions over time could have maintained the high genetic variation in German populations (Magni Diaz 2004) but could also have led to a lower differentiation between introduced populations. Concerning the overall level of genetic variation in German Northern red oak stands, additional import of seeding material seems not to be necessary at the moment. But concerning the limited number of North American populations investigated in our study and with respect to future climate change scenarios, additional import of material from other parts of the natural range might be necessary in the future.

Interestingly, the number of private alleles among German populations was especially high in Baden-Württemberg, suggesting multiple introductions or admixture of plant material before its introduction (Pettenkofer et al. 2019). In North America, 14 out of 19 private alleles were found in the northern populations BR-1 (9) and Constance Bay (5).

The AMOVA showed that only 1.10 % of the genetic variation was due to partitioning among populations. The PCoA clusters that visualized the genetic differentiation between populations (Fig. 4) and the pairwise population genetic distances supported the data of Pettenkofer et al. (2019). German populations clustered densely occupying only a small part of the two-dimensional space in the PCoA plot indicating the introduction from a limited geographic range (Fig. 4). Likewise, the distribution of point clouds in individual-based PCoA analyses and earlier chloroplast DNA analyses suggest that introduced populations originated from a limited range in the northern part of the species' distribution.

Considering also the similar levels of genetic diversity in native and introduced populations and the absence of recent genetic bottlenecks, the clustering of populations in the PCoA supported the hypothesis that German stands might have been established with material from a particular region within the species' native range.

Table 4 Relative frequency of private alleles.

Origin	Population	Locus	Allele	Rel. frequency
Lower Saxony	7	FIR104*	231	0.020
		VIT023*	113	0.028
Baden-Württemberg	22	VIT107*	129	0.010
	24	1C06	289	0.012
		1F07	295	0.012
	26	FIR013*	134	0.010
		0C11	198	0.009
Brandenburg	35	FIR028*	230	0.011
	38	0E09	257	0.010
Brockway Mountain, MI	BR1	1P10	268	0.021
		FIR028*	236	0.010
		FIR035*	166	0.010
		FIR024*	210	0.023
		GOT021*	95	0.042
		PIE040*	182	0.011
		0C11	232	0.011
		VIT107*	145	0.011
		VIT107*	151	0.011
Ottawa, Constance Bay	Const. Bay	2P24	170	0.011
		3D15	243	0.011
		0E09	243	0.011
		1C06	295	0.022
		PIE125*	148	0.021
Ford Forestry Center, MI	FC-B	0E09	253	0.038
		PIE040*	155	0.026
Clay, Nantahala, Tusquitee	Nantahala	3A05	169	0.011
		1F07	359	0.021
		PIE125*	166	0.011

Note: *EST-SSR markers

4.2 Weak genetic structure was observed across all sampled populations

Although the bootstrap support was weak for most of the clusters, the NJT showed that the two North American populations that are located close to the Lake Superior in Michigan, USA (FC-B and BR1), are clustered together. For the North American cluster, however, bootstrap

support was rather low (= 28 %), showing only weak distinction between North American and German populations.

However, there is no strong evidence for inter-populational or inter-regional differentiation within both German stands and North American populations.

In this study, only the EST-SSR marker FIR013 was identified as potentially under selection. It was the most differentiated marker with the highest F_{ST} value among all markers (Supplementary Table 2S), mostly due to the different frequencies of the FIR013 alleles in the North American population BR1 (data not shown), which was also the most differentiated population among all populations with the highest pairwise Jost's *D* and F_{ST} values. FIR013 is located in a CONSTANS-like gene and represents a region consisting of several tandemly repeated codons CAG and/or CAA encoding glutamine. This gene is known to be involved in the regulation of flowering time and photoperiodic control of growth (Lind-Riehl et al. 2014). Also, the distribution of alleles between populations of closely related *Q. rubra* and *Q. ellipsoidalis* shows that FIR013 might also serve as a marker to trace potential introgression of genes from *Q. ellipsoidalis* into *Q. rubra* (Lind-Riehl et al. 2014). Considering the overlap of ranges of both species in the Great Lakes region, genetic introgression may be one possible reason for higher differentiation of BR1. However, population BR1 is located outside the current distribution range of *Q. ellipsoidalis* (Lind-Riehl et al. 2014). The absence of signatures of local adaptation in German populations may suggest that introduced populations were established with material from provenances with similar environmental conditions as the ones prevailing in Germany (Liesebach and Schneck 2011; Pettenkofer et al. 2019). However, we only analyzed a limited number of potentially adaptive markers. Further analyses of candidate genes for local adaptation and of adaptive traits in provenance trials are needed to assess adaptive differences among provenances.

4.3 Conclusions and future perspective

Comparatively high levels of genetic variation at both EST-SSRs and nSSRs in German stands suggest that additional import of seeding material from the native range is not required for maintaining a sufficiently high level of genetic variation in Germany. Estimation of chloroplast haplotype diversity (Pettenkofer et al. 2019) further revealed that plantations in the southwest of Germany may provide a variable genetic resource. Moreover, Liesebach and Schneck (2011) state that German provenances perform better in respect of growth than

provenances from the natural range. Therefore, additional import of material from the native range is not necessary to enhance the genetic variation or growth of local Northern red oak stands. The analysis of additional candidate genes for local adaptation and of adaptive traits (e.g., drought tolerance) in provenance trials is needed to assess adaptive differences among German and North American provenances. The identification and genotyping of large numbers of SNPs can be achieved by genotyping by sequencing techniques such as restriction site-associated DNA sequencing (RADseq) and thus provide a new source of genome-wide potentially adaptive markers (Miller et al. 2007; Davey and Blaxter 2010).

Finally, North American populations representing the natural range could be characterized at nuclear and chloroplast DNA markers to narrow down the geographic origin of German plantations.

Acknowledgments: We thank Alexandra Dolynska, Andreas Parth, Katharina Burkardt, Christine Radler and Natalie Breidenbach for their valuable technical assistance and Mirko Liesebach (Thünen Institute) for providing plant material. The study was supported by the German Federal Ministry of Food and Agriculture (Funding code 22023314).

5 References

Aldrich P, Michler C, Sun W et al (2002) Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). Molecular Ecology Notes:472–474

Alexander LW, Woeste KE (2014) Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. Tree Genetics & Genomes 10:803–812

Barrett SC, Husband BC (1990) The genetics of plant migration and colonization. In: Brown HD, Clegg MT, Kahler AL et al (eds) Plant population genetics, breeding, and genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, pp 254–277

Bauer F (1951) Die Roteiche 1950. Eine ertragskundlich-biologische und holzkundliche Untersuchung. Dissertation, Universität Göttingen

Bauer F (1954) Zur Rassenfrage der Roteiche. Allgemeine Forstzeitschrift 9:470–474

Borkowski DS, Hoban SM, Chatwin W et al (2017) Rangewide population differentiation and population substructure in *Quercus rubra* L. Tree Genetics & Genomes 13:472

Bundesministerium für Ernährung und Landwirtschaft (BMEL) (2014) Der Wald in Deutschland. Ausgewählte Ergebnisse der dritten Bundeswaldinventur. Available via https://www.bundeswaldinventur.de/fileadmin/SITE_MASTER/content/Dokumente/Downloads/BMEL_Wald_Broschuere.pdf. Accessed 20 Mar 2018

Chmura D (2013) Impact of alien tree species *Quercus rubra* L. on understorey environment and flora: A study of the silesian upland (southern Poland). Polish Journal of Ecology 61:431–442

Cornuet J-M, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Cristescu R, Sherwin WB, Handasyde K et al (2010) Detecting bottlenecks using BOTTLENECK 1.2.02 in wild populations. The importance of the microsatellite structure. Conserv Genet 11:1043–1049

Daubree JB, Kremer A (1993) Genetic and phenological differentiation between introduced and natural populations of *Quercus rubra* L. Ann. For. Sci. 50:271s-280s

Davey JW, Blaxter ML (2010) RADSeq. Next-generation population genetics. Briefings in functional genomics 9:416–423

Durand J, Bodénès C, Chancerel E et al (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11

Earl DA, Holdt BM von (2012) STRUCTURE HARVESTER. A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet Resour 4:359–361

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE. A simulation study. Molecular Ecology 14:2611–2620

Excoffier L, Lischer HLE (2010) Arlequin suite ver 3.5. A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164

Giraudoux P, Antonietti J-P, Beale C et al (2018) pgirmess (R-package). Spatial analysis and data mining for field ecologists.

Goudet J (1995) FSTAT (Version 1.2). A Computer Program to Calculate F-Statistics. Journal of Heredity 86:485–486

Hickel R (1932) Dendrologie forestière. Lechevalier, Paris

Houba J (1887) Les chênes de l'Amérique septentrionale en Belgique. Leur origine, leur qualités, leur avenir. M. Ceysens (ed), Hasselt

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Molecular Ecology 17:4015–4026

Kalia RK, Rai MK, Kalia S et al (2011) Microsatellite markers. An overview of the recent progress in plants. Euphytica 177:309–334

Kopelman NM, Mayzel J, Jakobsson M et al (2015) Clumpak. A program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191

Langella O (1999) Populations. CNRS UPR 9034, Gif-sur-Yvette Cedex

Liesebach M, Schneck V (2011) Entwicklung von amerikanischen und europäischen Herkünften der Roteiche in Deutschland. Forstarchiv 82:125–133

Lind JF, Gailing O (2013) Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genetics & Genomes 9:707–722

Lind-Riehl JF, Sullivan AR, Gailing O (2014) Evidence for selection on a CONSTANS-like gene between two red oak species. Annals of Botany 113:967–975

Magni CR, Ducousso A, Caron H et al (2005) Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. Molecular Ecology 14:513–524

Magni Diaz CR (2004) Reconstitution de l'introduction de *Quercus rubra* L. en Europe et conséquences génétiques dans les populations allochtones. Dissertation, Ècole Nationale du Génie Rural, des Eaux et des Forêts (in French)

Merceron NR, Leroy T, Chancerel E et al (2017) Back to America. Tracking the origin of European introduced populations of *Quercus rubra* L. Genome 60:778–790

Miller MR, Dunham JP, Amores A et al (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome research 17:240–248

Möllerová J (2005) Notes on invasive and expansive trees and shrubs. Journal of Forest Science 51:19–23

Nagel R-V (2015) Roteiche (*Quercus rubra* L.). In: Vor T, Spellmann H, Bolte A et al (eds) Potenziale und Risiken eingeführter Baumarten. Baumartenportraits mit naturschutzfachlicher Bewertung. Univ.-Verl. Göttingen, Göttingen, pp 219–267

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. Journal of Molecular Ecolution 19:153–170

Niinemets Ü, Valladares F (2006) Tolerance to shade, drought, and waterlogging of temperate northern hemisphere trees and shrubs. Ecological Monographs 76:521–547

Peakall R, Smouse PE (2006) GenAlEx 6. Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288–295

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

Petit RJ, Pineau E, Demesure B et al (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. Proc Natl Acad Sci U S A 94:9996–10001

Pettenkofer T, Burkardt K, Ammer C et al (2019) Genetic diversity and differentiation of introduced red oak (*Quercus rubra*) in Germany in comparison with reference native North American populations. Eur J Forest Res 10

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

R Core Team (2016) R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Riepšas E, Straigytė L (2008) Invasiveness and ecological effects of red oak (*Quercus rubra* L.) in Lithuanian Forests. Baltic Forestry 14:122–130

Romero-Severson J, Aldrich P, Feng Y et al (2003) Chloroplast DNA variation of northern red oak (*Quercus rubra* L.) in Indiana. New Forests 26:43–49

Sullivan AR, Lind JF, McCleary TS et al (2013) Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. Plant Mol Biol Rep 31:231–239

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389–399

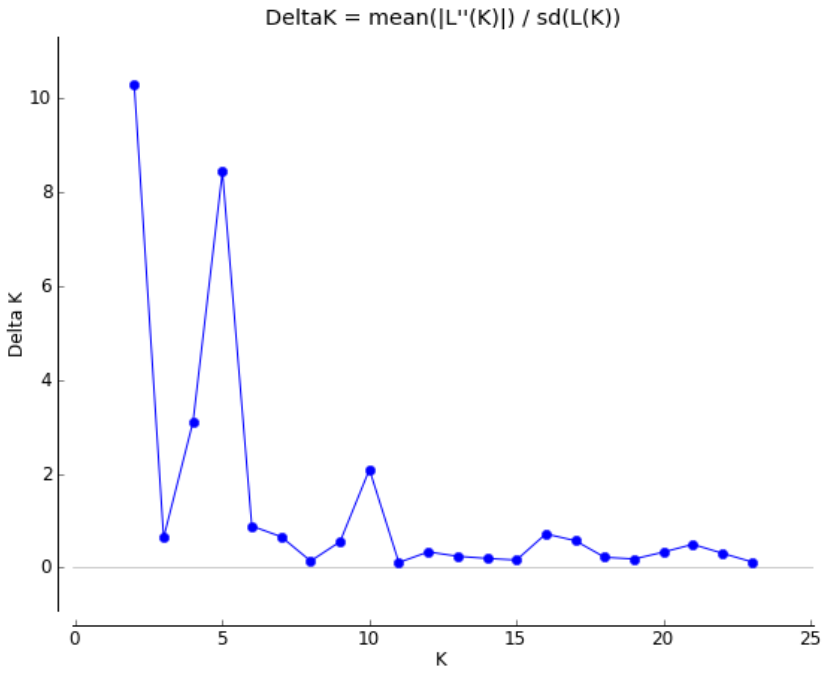
Vaughan TG (2017) IcyTree. Rapid browser-based visualization for phylogenetic trees and networks. Bioinformatics (Oxford, England) 33:2392–2394

Vor T (2005) Natural regeneration of *Quercus rubra* L. (Red Oak) in Germany. In: Nentwig W, Bacher S, Cock MJW et al (eds) Biological invasions - From ecology to control, pp 111–123

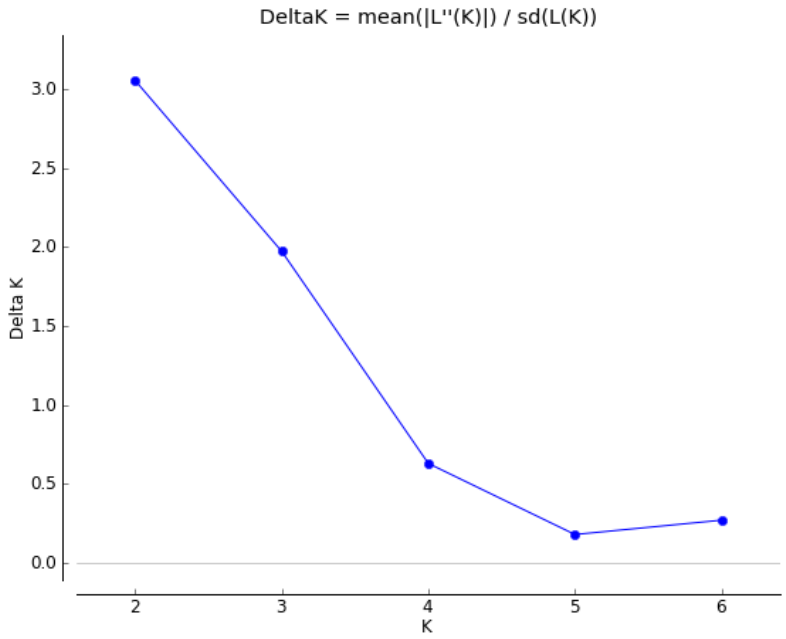
Vor T, Lüpke Bv (2004) Das Wachstum von Roteiche, Traubeneiche und Rotbuche unter verschiedenen Lichtbedingungen in den ersten beiden Jahren nach der Pflanzung. Forstarchiv 75:13–19

Zhang R, Hipp AL, Gailing O (2015) Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations. Botany 93:691–700

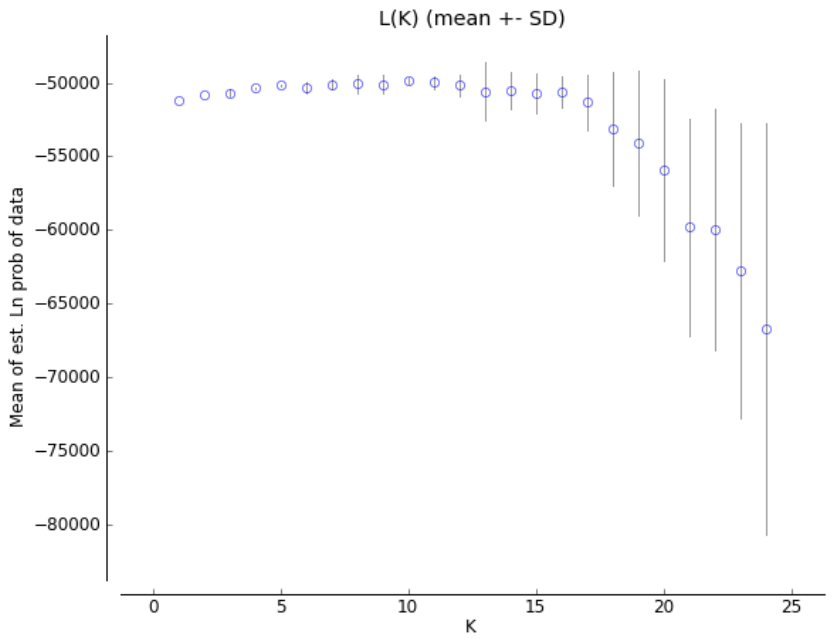
6 Supplementary Files



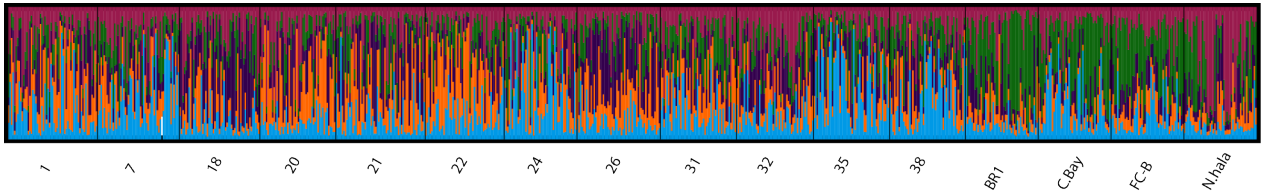
Supplementary Fig. 1S Delta K for all populations



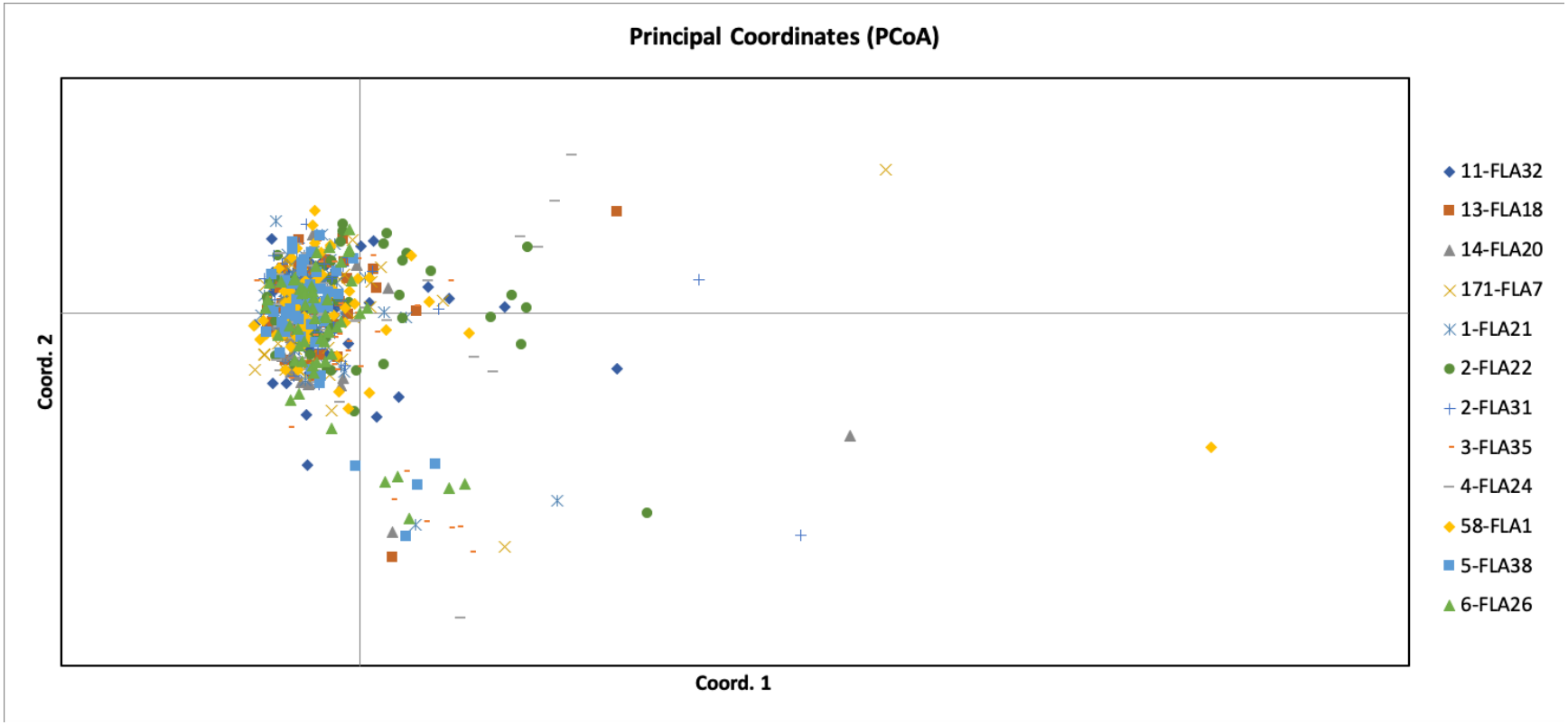
Supplementary Fig. 2S Mean of estimated Ln probability of data for all populations



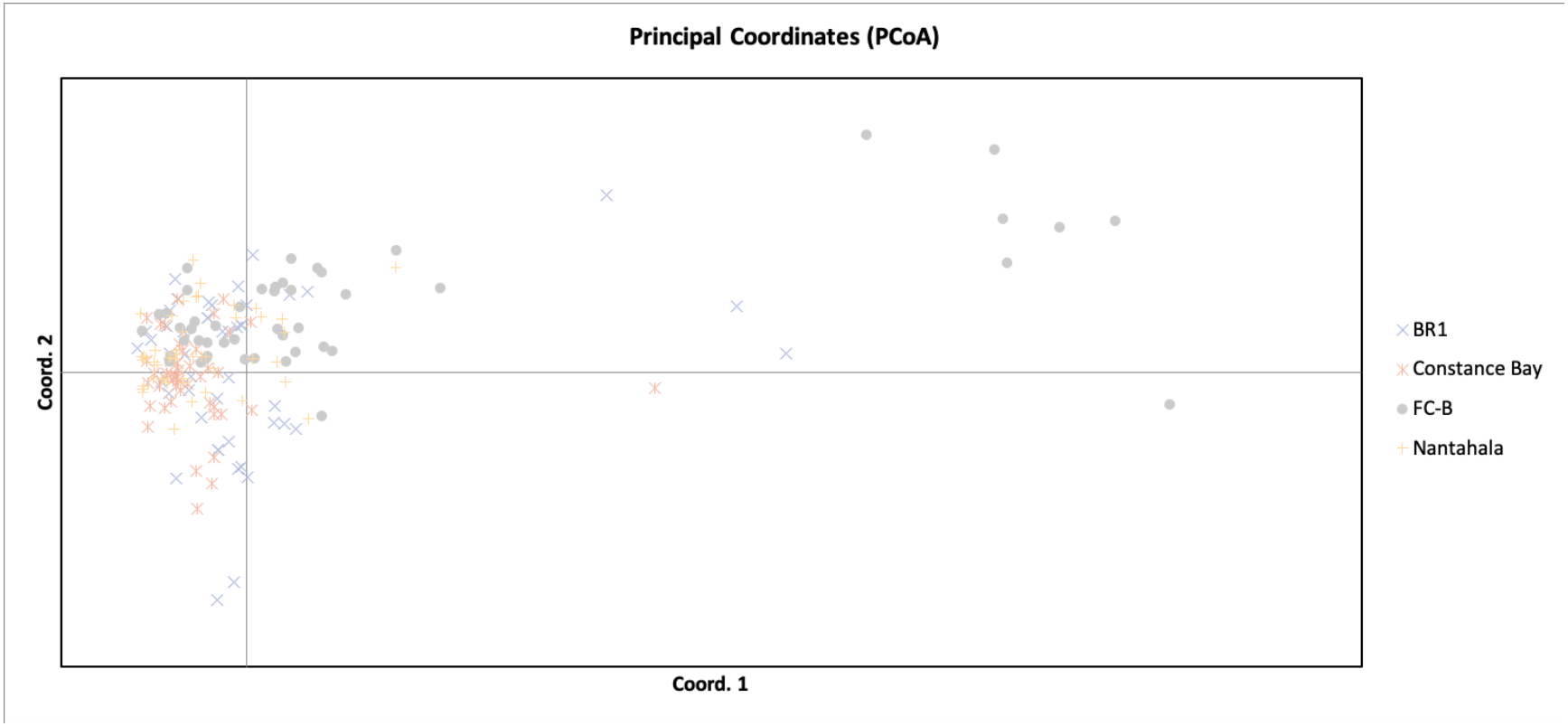
Supplementary Fig. 3S Delta K for North American populations



Supplementary Fig. 4S STRUCTURE analysis ($K = 5$) for all populations



Supplementary Fig. 5S Individual-based PCoA for German populations



Supplementary Fig. 6S Individual-based PCoA for North American populations

Supplementary Table 1S Nuclear microsatellite markers (SSRs) tested and used in this study.

Multiplex	Locus	Dye	Primer nucleotide sequence		Motif	Allele size, bp
			forward	reverse		
1	*quru*-GA-1P10[1]	FAM	ATTTCTGATGCAGGGTGTCG	TAGGCCAAGGACCAGAGACC	$(TG)_{12}GCC(TG)_3$	237-265
	quru-CA-2P24[1]	FAM	GCAAGAGATCACACACAAACTAGC	CTTTGGGTTCACCAAACAGC	$(CA)_{14}$	136–164
	quru-CA-3A05[1]	HEX	AACGTGACCTCTCTCACAGC	AGTGCTGGAGTGCTCATGG	$(CA)_{11}(CT)_2$	138–160
	quru-CA-3D15[1]	HEX	GGTGGTGGCAGATACACTGG	GACTCAGACAACCAACTTCAGG	$(CA)_{15}$	208–236
2	FIR043[2]*	HEX	TTCTCCATTTCACACGCTTC	ACGACATCGTTTTGGAGCTT	$(TC)_9$	114-146
	FIR035[2]*	HEX	GCTAAGGTTCCGTGTTCCAA	GGCCAGCAACTAAACCAAGA	$(AT)_6$	146-152
	FIR028[2]*	FAM	GGAAGAGTGTTCGGAAAGCA	CCAGCTCCTCCACAATAGCA	$(TC)_8$	201-237
	FIR013[2]*	FAM	CGGGGAGGTTGATGAGTATT	AACACTGTCACCCCCATAGC	$(CAG)_5$	133-144
3	*quru*-GA-0E09[3]	FAM	TGCCATCCCTATACACAACCA	CCTCCATCACAAAGTTGCC	$(GA)_{16}$	186-230
	quru-GA-1C06[3]	ROX	CAAATAAATATTGTGGGGTTCA	GGAGGGGATCCGGAAAA	$(GA)_{29}$	234-262
	quru-GA-1F07[3]	FAM	CCGGTCAAAGAAGTTATCAGA	GGGTGGATTGGGTTTCTACCTA	$(GA)_{22}$	306-348
	FIR004[1γ]	HEX	TCTCTCTCAGGGCAGCTTCT	AACCAAACTCAGATCCAGATTCA	$(CT)_{18}$	128–180
	GOT021[1]*	HEX	AGAAAGTTCCAGGGAAAGCA	CTTCGTCCCCAGTTGAATGT	$(AT)_{13}$	95–101
	quru-GA-2M04[3γ]	ROX	GGAGAGGACGGGATGCC	TACTATGTCAGCCGGATG	$(GA)_{20}$	182-220
	FIR031[1]*	FAM	ACGAGTCCAACGGAAGTTGT	CACAACTTCACAAGGCAAGG	$(TC)_7$	139–174
	FIR024[1]*	HEX	CGCTTCTCCTCATCCTCAAG	CTCAAAAGGCACGATTCTCC	$(CCT)_6$	214–229
4	*quru*-GA-0C11[3]	HEX	ATACCCAGCTCCCATGACCA	TCCCCAAATTCAGGTAGTGT	$(GA)_{15}$	204-222
	FIR104[1]*	FAM	TTAACTCGGTTTGCGACTCA	AGCACGTGACTCGACCTGTA	$(GGT)_7$	206–221
	PIE040[1]*	HEX	GTGAGAGAGAGAGAGACAAAGAAGAAAAA	AAATTCTCCGCCACATTGAG	$(TTC)_8$	155-174
	PIE125[1]*	FAM	AATACAAATCGCAGGAGGTG	CTAACCCATCGTTCATGGAG	$(GGAAGC)_3$	147-162
	quru-GA-2F05[3γ]	FAM	CCGCTTCGTGACGATTATTC	GAGGTTTGGAGGAGAGATCATTCT	$(GA)_{21}$	294-322
	GOT037[1*γ]	HEX	CCATCCTTTTCATTCTTTCCA	TGTTGTTGTTGCTGTTGTCG	$(CT)_{11}$	239–265
	GOT040[1]*	FAM	AAGGCACTCGTCGCTTTCTA	ACCGATTTGAAGCTCGAGAA	$(GA)_{11}$	234–254
	VIT107[1]*	HEX	TGATCACAGATTGGAGCTTAACA	CCCCCACTTAGGAAAGAAGC	$(TA)_{13}$	127-139
	VIT023[2]*	ROX	AATGCGAACGACATGAACAA	CTCTCGTCGGAGACTCAACC	$(ATA)_6$	115–118

Note: [1] Sullivan et al. (2013); [2] Durand et al. (2010); [3] Aldrich et al. (2002); * EST-SSRs; γ excluded from further genotyping due to difficulties in allele binning

Supplementary Table 2S Outlier detection using the Arlequin software

Locus	Obs, Het, BP	Obs F_{ST}	F_{ST} *P*-value	1-F_{ST} quantile
1P10	0.757	0.008	0.356	0.644
2P24	0.849	0.014	0.478	0.478
3A05	0.774	0.012	0.477	0.523
3D15	0.873	0.010	0.414	0.586
FIR013*	0.107	0.052	0.008	0.008
FIR028	0.814	0.013	0.496	0.504
FIR035	0.335	0.004	0.257	0.743
FIR024	0.632	0.020	0.316	0.316
FIR031	0.833	0.013	0.499	0.501
GOT021	0.188	0.023	0.243	0.243
0E09	0.946	0.009	0.373	0.627
1C06	0.917	0.013	0.478	0.522
1F07	0.941	0.018	0.338	0.338
FIR104	0.315	0.009	0.375	0.625
GOT040	0.628	0.010	0.406	0.594
PIE040	0.500	0.018	0.364	0.364
PIE125	0.712	0.010	0.400	0.600
0C11	0.897	0.010	0.400	0.600
VIT023	0.361	0.018	0.373	0.373
VIT107	0.648	0.020	0.302	0.302

Note: *significant outlier

Supplementary Table 3S Results of AMOVA for all populations

Source of variation	d.f.	Sum of squares	Variance components	Percentage variation	*P*
Among populations	15	191.35	0.07	1.10	<0.001
Among individuals within populations	807	4838.77	0.11	1.78	<0.001
Within individuals	823	4760.50	5.78	97.12	<0.001
Total	1645	9190.62	5.96		

Supplementary Table 4S Results of AMOVA for German populations

Source of variation	d.f.	Sum of squares	Variance components	Percentage variation	*P*
Among populations	11	120.08	0.05	0.78	<0.001
Among individuals within populations	619	3730.59	0.08	1.41	0.008
Within individuals	631	3696.00	5.86	97.81	<0.001
Total	1261	7546.67	5.99		

Supplementary Table 5S Results of AMOVA for the North American populations

Source of variation	d.f.	Sum of squares	Variance components	Percentage variation	*P*
Among populations	3	43.89	0.10	1.88	<0.001
Among individuals within populations	188	965.31	-0.03	-0.58	0.700
Within individuals	192	997.50	5.20	98.70	0.210
Total	383	2006.70	5.26		

CONTRIBUTIONS TO THE SCIENTIFIC PAPERS

My contributions to the four co-authored papers are described below.

For Paper 1, 2 and 4 I was, together with Katharina Burkardt, responsible for the identification and selection of potential populations for sampling. I took part in the fieldwork in all stands of the Brandenburg area and in 12 stands from other federal states. Also, I was responsible for the entire DNA analyses except for Paper 4, where Christine Radler and Alexandra Dolynska performed PCRs for half of the samples.

For Paper 3, I was additionally responsible for the preparation for the sequencing of the chloroplast genomes of 8 samples from North America. For this paper, Alexandra Dolynska did the testing of the primer-enzyme-combinations, which I identified beforehand.

For all of the papers, I was responsible also for the overall analysis of the data. Further, I was responsible for drafting the manuscripts and to finalise them according to the comments and improvements suggested by the co-authors.

SYNOPSIS

Northern red oak (*Quercus rubra* L.) has its natural distribution area in the eastern and north-eastern USA and south-eastern Canada and was first introduced at the end of the 17th century in France. Later, *Q. rubra* was planted in parks and botanical gardens all over Europe. In comparison to native white oak species, northern red oak can be harvested after only 80−120 years and features a comparably low demand for water and nutrients. It needs at least 600 mm precipitation per anno and can cope with hard winters and hot summers with mean temperatures between −14 and 26 °C. Currently, northern red oak is with 44.000 ha the most important non-native deciduous tree species in Germany. Its wood is used in a wide range of products, although it cannot be used in the construction of wine barrels. Northern red oak is currently regarded as qualified for cultivation in Germany and can improve the adaptation of forests to projected effects of climate change.

The first manuscript is a review of all relevant studies, that investigate the origin and genetic variation of German resp. European northern red oak stands in comparison to populations from the natural distribution range. This manuscript connects the three following papers with relevant research and provides a basis for future studies. In general, due to the lack of geographic barriers and isolated glacial refugia, various studies failed to identify particular post-glacial migration routes. A possible consequence is the -compared to other oak species- lower genetic differentiation of *Q. rubra* in its natural range. Other studies show that European red oak stands show sufficiently high genetic variation with at the same time only low genetic differentiation. This suggests that multiple introductions and admixture of seeding material might have prevented genetic erosion in introduced populations. All in all, none of the reviewed studies could identify a particular region of origin of introduced red oak stands. Yet all studies propose, that introduced stands originate from the northern part of the natural range.

The second paper addresses patterns of chloroplast variation in German red oak stands in comparison to populations in the natural distribution range. To characterize chloroplast haplotype diversity and geographic structure, individuals from 62 populations in Germany and North America were genotyped at five chloroplast microsatellite loci. In comparison to native reference populations, German red oak stands showed a relatively low genetic differentiation among populations and represented only a fraction of the haplotype diversity found in the natural distribution range. For several stands located in Baden-Wuerttemberg in south-western Germany, a considerably higher haplotype diversity compared to other German stands was

found. While most German stands showed signatures of founder effects at chloroplast markers, the diversity of stands in Baden-Wuerttemberg might have been increased due to admixture and multiple introductions of different North American provenances. Overall, German stands seem to originate from a limited geographic area, possibly located in the northern part of the natural range.

The main objective of Paper 3 was to design chloroplast DNA markers that are cost-effective, easy to apply and to score. In combination with established cpSSR markers they should increase haplotype resolution in populations. In a first step, 87-97 % of the entire chloroplast genome (excluding the second inverted repeat) of 8 individuals of different regions within the red oaks natural range were sequenced with a coverage of 4,030X–6,297X. Afterwards, the genome sequences were assembled using the publicly available chloroplast genome of *Quercus rubra* L. as a reference. 118 SNPs and 107 insertions and deletions were detected, allowing the development of 15 cleaved amplified polymorphic sequence (CAPS) markers for *Q. rubra*. By combining these markers with five established chloroplast microsatellite markers, 10 haplotypes were identified in the diversity panel of 19 red oak populations. Hereby, two cpSSR haplotypes could now be separated into five combined haplotypes. This study showed that the designed markers are useful to assess haplotype diversity with high resolution. Also, they are transferable to *Quercus ellipsoidalis*, a closely related species.

Paper 4 addresses patterns of neutral and potentially adaptive nuclear genetic variation in red oak stands across Germany. In total, 792 trees were genotyped. They included 611 trees from 12 stands in Germany of unknown origin and 181 trees from four populations within the natural range. For the analysis, 12 potentially adaptive- and 8 putatively selectively neutral nuclear microsatellite markers were selected. The results showed that German stands hold comparatively high levels of genetic variation at both EST-SSRs and nSSRs. Nevertheless, they are more similar to each other than to native populations. These findings support earlier chloroplast DNA studies which suggested that German populations originated from a limited geographic area within the natural range. Also, the absence of signatures of selection in German stands may suggest that introduced stands were founded with material from provenances that are already adapted to environmental conditions similar to those in Germany. All in all, the results suggest that for maintaining a sufficiently high level of genetic variation in Germany, an additional import of seeding material from the native range is not required.

ZUSAMMENFASSUNG

Das natürliche Verbreitungsgebiet der Roteiche (*Quercus rubra* L.) erstreckt sich über den Osten und Nordosten der USA und über den Südosten von Kanada. Sie wurde am Ende des 17. Jahrhunderts in Frankreich eingeführt und gelangte von dort in Parks und botanische Gärten nach ganz Europa. Die Roteiche weist nur einen geringen Bedarf an Nährstoffen und Wasser auf: Sie benötigt mindestens einen Jahresniederschlag von 600 mm und kann harte Winter und heiße Sommer mit mittleren Temperaturen zwischen −12 und 26 °C ertragen. Im Augenblick ist die Roteiche mit ca. 44.000 ha die flächenmäßig bedeutendste nichtheimische Laubbaumart in Deutschland. Ihr Holz kann für eine Vielzahl von Produkten eingesetzt werden, eignet sich jedoch aufgrund der der unverthyllten Gefäße nicht zur Produktion von Weinfässern. *Q. rubra* gilt in Deutschland als anbauwürdig und kann die Anpassungsfähigkeit von Wäldern an die prognostizierten Klimaveränderungen verbessern.

Das erste Manuskript ist ein Überblick über alle relevanten Studien zur Herkunft und genetischen Variation von deutschen bzw. europäischen Beständen im Vergleich zu Populationen in Nordamerika. Das Manuskript verbindet die drei folgenden Artikel mit relevanten Studien. Allgemein ist es aufgrund des Mangels an geografischen Barrieren und isolierten glazialen Refugien bislang keiner Studie gelungen ausgeprägte nacheiszeitliche Rückwanderungsrouten zu identifizieren. Eine mögliche Folge ist die vergleichsweise niedrigere genetische Differenzierung von *Q. rubra* in ihrem natürlichen Verbreitungsgebiet. Andere Studien zeigen, dass europäische Roteichenbestände eine ausreichend hohe genetische Variation bei gleichzeitig niedriger genetischer Differenzierung aufweisen. Dies weist darauf hin, dass mehrfache Einführungen und Durchmischung von Saatgut die genetische Verarmung eingeführter Bestände verhindert haben. Insgesamt konnte keine der vorgestellten Studien bestimmte Ursprungsgebiete identifizieren. Mehrere Studien deuten jedoch darauf hin, dass der Ursprung eingeführter Bestände im nördlichen Teil des natürlichen Verbreitungsgebietes liegt.

Der zweite Artikel thematisiert die räumliche Verteilung von Variationsmustern an Chloroplasten-Markern in deutschen Beständen im Vergleich zu Populationen im natürlichen Verbreitungsgebiet. Dafür wurden Proben von 62 deutschen und nordamerikanischen Populationen an fünf cpSSR-Genorten untersucht. Im Vergleich zu nordamerikanischen Populationen zeigen deutsche Bestände nur eine relativ geringe Differenzierung zwischen Populationen und weisen nur einen Bruchteil der Haplotypendiversität auf. In den meisten Beständen in Baden-Württemberg wurde eine höhere Haplotypendiversität als in anderen Regionen Deutschlands gefunden. Während die meisten deutschen Bestände Anzeichen von

Gründereffekten zeigten, könnte die Chloroplasten-Diversität im Südwesten Deutschlands durch mehrfache Einführungen und Durchmischung von verschiedenen nordamerikanischen Herkünften erhöht worden sein. Insgesamt zeigen die Ergebnisse, dass deutsche Bestände möglicherweise aus einem begrenzten geografischen Gebiet im Norden des natürlichen Verbreitungsgebietes stammen.

Das Hauptziel des dritten Artikels war die Entwicklung von Chloroplasten-Markern, welche kosteneffizient und leicht anzuwenden sind. In Kombination mit bereits entwickelten cpSSR Markern sollten sie die Haplotypen-Auflösung in Populationen erhöhen. In einem ersten Schritt wurden 87-97 % des gesamten Chloroplastengenoms von 8 Bäumen aus verschiedenen Regionen des natürlichen Verbreitungsgebietes sequenziert und die Sequenzen mit Hilfe eines bereits vorhandenen Referenzgenoms zusammengefügt. Insgesamt konnten 118 SNPs und 107 Indels identifiziert- und im nächsten Schritt 15 CAPS-Marker entwickelt werden. In Kombination mit fünf etablierten cpSSR-Markern konnten bei der Anwendung auf 19 nordamerikanische Roteichenpopulationen insgesamt 10 kombinierte Haplotypen identifiziert werden. Dabei konnten zwei cpSSR-Haplotypen in insgesamt fünf kombinierte Haplotypen aufgetrennt werden. Die Ergebnisse der Studie zeigen, dass die entwickelten Marker für die Untersuchung der Haplotypendiversität mit hoher Auflösung nützlich- und auf die eng verwandte Roteichenart *Quercus ellipsoidalis* übertragbar sind.

Der vierte Artikel beschäftigt sich mit neutralen und potenziell adaptiven nuklearen genetischen Variationsmustern in deutschen Roteichenbeständen. Dazu wurden 792 Bäume aus deutschen und nordamerikanischen Populationen untersucht. Für die Analyse wurden 12 potenziell adaptive EST- und 8 mutmaßlich selektiv neutrale nukleare Mikrosatelliten ausgewählt. Die Ergebnisse zeigen nicht nur, dass deutsche Bestände eine vergleichsweise hohe genetische Variation an beiden Markertypen aufweisen, insgesamt jedoch ähnlicher zueinander als zu nordamerikanischen Populationen sind, sondern auch, dass eingeführte Bestände aus einem begrenzten geografischen Gebiet im natürlichen Verbreitungsgebiet stammen. Die Abwesenheit Selektion in deutschen Beständen könnte darauf hinweisen, dass eingeführte Bestände mit Material begründet wurden, welche von Herkünften stammt, die bereits an Umweltbedingungen ähnlich derer in Deutschland angepasst waren. Um einen ausreichend hohen Grad an genetischer Variation in deutschen Beständen zu erhalten, seien keine zusätzlichen Saatgutimporte aus dem natürlichen Verbreitungsgebiet notwendig. Eine effektive Ergänzung des Roteichen-Genpools in Deutschland ist er auf der Grundlage weiterer Forschungen im Hinblick auf die Identifizierung besonders geeigneter Provenienzen sinnvoll.