Institut für Pflanzenzüchtung und Pflanzenschutz
der Landwirtschaftlichen Fakultät
der
Martin-Luther-Universität
Halle-Wittenberg

„Genetische Variabilität und phänotypische Plastizität der Modellpflanze *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae)“

Dissertation

zur Erlangung des akademischen Grades
doctor agricularum (Dr. agr.)

vorgelegt von

Diplomagraringenieurin
Heike Schmuths

geb. am 18.05.1978
in Magdeburg

Gutachter: Prof. Dr. W. E. Weber
Prof. Dr. K. Bachmann
Prof. Dr. W. Diepenbrock

Verteidigung am: 28.06.2004

Halle/Saale 2004

urn:nbn:de:gbv:3-000007461
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000007461]

Inhaltsverzeichnis

		Seite
1.	Zusammenfassung / Summary	2
2.	Problemstellung: genetische Variabilität und phänotypische Plastizität	6
2.1.	Definitionen und Hintergründe zur Fragestellung	6
2.2.	Single nucleotide polymorphism (SNP) - der (fast) ideale molekulare Marker?	10
3.	*Arabidopsis thaliana* als Modell für Variabilität und Plastizität	13
3.1.	Das Untersuchungsobjekt	13
3.2.	Bisherige Beschränkungen der Stichprobenauswahl bei *Arabidopsis thaliana*	15
3.3.	Populationsstruktur von *Arabidopsis thaliana*	17
3.4.	Intraspezifische Variabilität von Genomgrößen	18
3.5.	Ergebnisse	19
3.5.1.	Geographic distribution and recombination of genomic fragments on the short arm of chromosome 2 of *Arabidopsis thaliana*	21
3.5.2.	Genome size variation among accessions of *Arabidopsis thaliana*	42
3.5.3.	Phenological and morphological plasticity differ among populations of a world-wide sample of *Arabidopsis thaliana* (L.) Heynh. under different temperature treatments	51
3.5.4.	Evidence for pleistocene refugia and postglacial expansion of *Arabidopsis thaliana* from genome-wide SNP markers	65
4.	Schlussfolgerungen	89
5.	Ausblick	95
6.	Referenzen	97
6.1.	*Curriculum vitae*	97
6.2.	Publikationsliste	98
6.3.	Danksagungen	100
6.4.	Eigenständigkeitserklärung	101
7.	Literatur	102
8.	Verzeichnis der Abkürzungen und Fachtermini	113
9.	Anhang – Supplementary Data	114

1. Zusammenfassung

Arabidopsis thaliana ist eine bedeutende Modellpflanze für verschiedenste biologische Fragestellungen. Als nahe Verwandte vieler landwirtschaftlich relevanter Arten kommt ihr eine wichtige Rolle bei der Übertragung von Forschungsergebnissen auf Kulturpflanzen zu. Die zum Teil gattungsübergreifenden Sequenzhomologien einiger chromosomaler Regionen erleichtern diese Übertragung.
Die Kenntnis der natürlichen Variabilität von *A. thaliana* ist neben der durch künstliche Mutation entstanden Variabilität sehr wichtig. Die genetische, morphologische und phänologische Variation zwischen bisher untersuchten Stichproben zeigte kaum Korrelation zu geographischen oder klimatischen Eigenschaften des Herkunftsortes. Das beruht teilweise auf der noch nicht abgeschlossenen nacheiszeitlichen Ausbreitung innerhalb des Areals, bei der anthropogene Einflüsse eine zunehmend große Rolle spielen. Dass Pflanzen auch außerhalb ihres Herkunftsortes in anderen klimatischen Verhältnissen überleben können, kann zum Beispiel durch eine weite phänotypische Plastizität und/oder durch lokale Anpassung mit wenigen Genen bedingt sein.

Die vorliegende Arbeit hatte das Ziel, eine repräsentative Stichprobe des natürlichen eurasischen Areals molekularbiologisch, morphologisch und phänologisch zu untersuchen. Die Ergebnisse dieser Untersuchungen sollten untereinander korreliert werden. Weiterhin sollten geographische oder klimatische Eigenschaften des Herkunftsortes in die Korrelationen miteinbezogen werden.

Die Untersuchung von genomweit größtenteils gleichmäßig verteilten Einzelbasenunterschieden (single nucleotide polymorphisms, SNPs) an einer repräsentativen Stichprobe zeigte eine Assoziation zur Ost-West Verteilung der alternativen Ausprägungsformen weniger SNPs. Weiterführende Sequenzierstudien mit einer erweiterten Stichprobe von Akzessionen bestätigten diese Assoziation in zwei der drei genomischen Regionen. Diese zwei Regionen lagen innerhalb eines Gens und einer unkodierenden Region auf Chromosom 2, ca. 300 000 bp von einander entfernt.
Durch die große Anzahl von 108 erfassten SNPs ließen sich deutlich auch innerhalb dieser genomischen Regionen SNPs nachweisen, die mit der eiszeitlich beeinflussten geographischen Herkunft korreliert waren. Die Mehrheit der SNPs stellten unabhängig voneinander nacheizeitlich entstandene lokale Mutationen dar. Das erklärt, dass bei

genomweiten Stichproben einige wenige SNPs ein geographisches Muster zeigten, das bei der Mehrheit der SNPs nicht zu erkennen war.

Es konnten durch Sequenzvergleiche zwei Consensushaplotypen aus dem Datensatz gebildet werden, die zu jeweils gleicher Frequenz vorkommen und charakteristisch für eine asiatische und eine südwest-europäische Ausgangspopulation sind. Jede Akzession konnte entweder direkt einem Consensushaplotyp zugeordnet werden oder war eine Rekombination zwischen diesen zwei Haplotypengruppen.
Die zwei Gruppen konnten in einer weiteren genomweiten SNP-Analyse innerhalb des Gesamtgenoms wiedergefunden werden. In diesem Zusammenhang wurde auch die Populationsstruktur von *A. thaliana* untersucht.

Die morphologischen und phänologischen Untersuchungen der Akzessionen zeigten eine große Schwankungsbreite der erfassten Merkmale bei zwei Temperaturen (14°C und 22°C). Eine Korrelation zwischen der Länge der phänologischen Phasen und der Temperatur des Herkunftsortes der Akzessionen wurde gefunden. Bei einer höheren Temperatur (22°C) wird die Entwicklungsdauer von Akzessionen aus winterkalten Gebieten stärker beschleunigt als von Akzessionen aus winterwarmen Herkunftsorten. Korrelationen zwischen SNPs und morphologischen Daten wurden in drei Fällen festgestellt.

Durchflußzytometrische Untersuchungen resultierten in signifikanten Genomgrößenunterschieden zwischen den untersuchten Akzessionen mit schwachen Korrelationen zu Längen- und Breitengraden der Herkunftsorte. Es wurden zwei natürlich vorkommende tetraploide Akzessionen gefunden.

Die vorliegende Arbeit erfasst erstmals die Variation innerhalb von *A. thaliana* an einer repräsentativen Stichprobe von Akzessionen des gesamten natürlichen Areals. Der Nachweis von zwei genetisch verschiedenen Haplotypengruppen und die Sammlung einer Reihe neuer Akzessionen mit dem „asiatischen“ Haplotyp legen es nahe, adaptive Unterschiede zwischen den beiden Haplotypengruppen an der Nachkommenschaft von Kreuzungen zwischen den deutlichsten Repräsentanten der beiden Haplotypengruppen in weiteren Studien zu untersuchen.

1. Summary

Arabidopsis thaliana is an important model plant for several biological problems. As a close relative to many agricultural important species, *A. thaliana* plays an important role in transferring results from basic research in a model system to crops. This transfer is facilitated by sequence homologies of some chromosomal regions that transcend genus borders.
Besides the useful variability from induced mutations, a knowledge of natural variability of *A. thaliana* provides different information with a great relevance for the plants in nature. Previous investigations of genetic, morphological and phenotypical variation indicated no correlation with climatic features of the source location of an accession. This might be due to the ongoing postglacial spread of *A. thaliana* in its range which is influenced by an increasing anthropogeneous impact. The survival of plants outside their point of origin in different climatic conditions can be mediated by a wide range of phenotypic plasticity and/or variation in relatively few genes.

The aim of this thesis was the molecular, phenotypic and phenologic characterization of plants from the Eurasian natural range. Correlations between these data were to be investigated. As far as possible, relevant results should be correlated with the geographic coordinates and the climatic conditions at the origin of the accessions.

The investigation of genome wide single nucleotide polymorphisms (SNPs) mostly equally spread in the genome using a representative set of accessions showed an east-west association of the alternative alleles of some SNPs. Sequence studies with a broader set of accessions confirmed this association in two of three sequenced chromosomal regions. These two regions were located in a gene and in a non-coding region on chromosome 2, 300 000 bp apart.
With these data as a guide, further SNPs correlated with the postglacial distribution history could be identified in a genome-wide sample of 108 SNPs. Most SNPs in *A. thaliana* are the result of independent post-glacial mutation events, with the SNPs diagnostic for the glacial refugial populations interspersed. This is why a genome wide survey revealed only a few SNPs with a clear east-west distribution pattern, while most SNPs showed no correlation to either latitude or longitude.

Based on sequence alignments SNPs with equal frequencies for both alleles could be defined in the genome of the two consensus haplotypes, which are characteristic for either an „Asian“ and a „European“ population. Each investigated accession could be assigned to either one of the consensus haplotypes or as a recombination product of the groups.
The two groups were also found in the entire genome in a second genome wide survey. The population structure of *A. thaliana* were also investigated in this context.

Morphological and phenological studies of accessions showed broad limits of variation of the scored features in the two different temperature regimes (14°C and 22°C). A weak correlation to summer and winter warm areas were found. Correlations between genetical and morphological data were found for three SNPs.

Flow cytometry revealed significant intraspecific differences in genome sizes between the investigated diploid accessions. Weak correlations of genome size with latitude and longitude were found. Two naturally occurring tetraploid accessions were found.

This work demonstrates for the first time the intraspecific genetic variation of *A. thaliana* in a sample that is representative for the entire natural Eurasian range of the species. The detection of two distinct haplotype groups and the collection of a fairly large set of new accessions carrying the „Asian“ haplotype provides the basis for an investigation of adaptive traits which should make use of progenies of crosses between the most characteristic representatives of the two haplotype groups.

2. Problemstellung: genetische Variabilität und phänotypische Plastizität

2.1. Definitionen und Hintergründe zur Fragestellung

Genetische und phänotypische Variabilität sind seit vielen Jahren das Thema zahlreicher Untersuchungen. Im Jahr 1911 prägte Johannsen die Begriffe „Genotyp“ und „Phänotyp“.
Genetische Variabilität ist laut Definition die vererbbare Variation innerhalb und zwischen Arten, Varietäten oder Sorten (Aird, 1994). Sie entsteht hauptsächlich durch Rekombination nach Hybridisierung unter der Voraussetzung einer unterschiedlichen Basenabfolge im Genom der Kreuzungspartner. Die Voraussetzungen von unterschiedlichen Nukleotidsequenzen zwischen Genomen bilden genetische Mutationen.
Mutationen können entweder als stille Mutationen („silent mutations“) auftreten, die die translatierte Aminosäure eines Gens und ihre Funktion nicht beeinflussen, oder als verändernde Mutationen („missense mutations“), die eine Änderung der Aminosäureabfolge bewirken (Knippers et al., 1990). Diese veränderten Aminosäureabfolgen lassen zum Teil die Funktion des Proteins unbeeinträchtigt, können aber auch bis zu einem vollständigen Funktionsverlust führen. Zwischen diesen beiden Extrema gibt es alle möglichen Zwischenformen mit mehr oder weniger starker Beeinflussung der Proteinfunktion, die z.T. einen direkten Einfluß auf den Phänotyp haben kann (Knippers et al., 1990).
Die Auswirkungen einer Mutation beschränken sich aber selten auf ein Eiweiß, z.B. auf die Effizienz eines Enzyms. Oftmals, vor allem bei „regulatorisch“ wirksamen Genen, die die Transkription anderer Gene regeln, kommt eine Mutation indirekt, in verschiedenen Geweben und zu spezifischen Zeiten zur Geltung. In der formalen Genetik haben diese Effekte verschiedene Bezeichnungen, z.B. Pleiotropie und Epistasie (Pigliucci, 2001).
Pleiotropie ist die Beeinflussung mehrerer Merkmale durch ein Gen. Natürlich kann Pleiotropie auch durch eine enge Kopplung zwischen Genen, die unterschiedliche Merkmale und damit auch Phänotypen beeinflussen, vorkommen. Epistasie ist die Beeinflussung der Aktion von Genen durch andere Gene, also alle denkbaren Arten dieser Wechselwirkung zwischen verschiedenen Genen einer Eigenschaft (Wricke, 1971). Das kann zu einem komplexen genetischen Hintergrund des betreffenden Merkmals führen. Beide Effekte machen es schwierig, Variation in einem Phänotyp auf Allele eines bestimmten Gens zurückzuführen.
Weiterhin wirken erbliche Effekte, welche die Genexpression beeinflussen ohne die DNA-Sequenz zu verändern, die sogenannten epigenetischen Effekte (Wolffe und Matzke, 1999)

auf den Phänotyp ein. Auch der Faktor Umwelt wird als ein Haupteffekt für die Ausprägung des Phänotyps angesehen (Pigliucci, 2001).
Inwieweit die Reaktionen einer Pflanze auf die sie umgebende Umwelt genetisch determiniert sind, ist oftmals schwer zu bestimmen und daher ein kontrovers diskutiertes Thema (z.B. Battjes und Bachmann, 1994; Lu und Wu, 1986; Dorn et al., 2000; Pollard et al., 2001). In der Züchtung wird das Problem empirisch gelöst, z.B. durch die Prüfung eines Genotyps an mehreren Standorten und über mehrere Jahre. Dabei soll bestimmt werden, ob es wiederholbare Effekte der Umwelten auf den Genotyp gibt und ob diese Effekte möglicherweise vererbt werden.
Die phänotypische Plastizität wird als das Vermögen eines Genotyps definiert, auf bestimmte Umweltbedingungen hin unterschiedliche Phänotypen auszubilden (Pigliucci, 2001). Das fundamentale Untersuchungswerkzeug der phänotypischen Plastizität ist die Idee der Reaktionsnorm, ein Begriff den Wolterek bereits 1909 prägte. Eine Reaktionsnorm stellt die Abhängigkeit der phänotypischen Ausprägung eines Merkmals bei ein und demselben Genotyp von einem Umweltfaktor dar. Sie ist deshalb im strikten Sinn nur bei klonal vermehrtem Material oder bei homogenen Linien feststellbar. Reaktionsnormen werden oft nur bei zwei Werten eines Umweltfaktors (warm/kalt, trocken/feucht, Licht/Schatten, gedüngt/ungedüngt) gemessen und dann vereinfacht als lineare Abhängigkeit dargestellt. Sorgfältige Messungen über einen größeren Bereich des Umweltfaktors zeigen jedoch in der Regel, dass die Abhängigkeit nicht linear ist.

Das Ziel der vorliegenden Arbeit ist es, die genetische Variabilität natürlicher Akzessionen von *A. thaliana* an einer repräsentativen Stichprobe des eurasischen Areals anhand von gleichmässig über das ganze Genom verteilten genetischen Markern zu erfassen. Durch Assoziationen zwischen Markerallelen und geographischen sowie morphologischen und phänologischen Eigenschaften sollen Rückschlüsse über Adaption und Selektion am natürlichen Standort gezogen werden. Weiterhin sollen Aussagen über die Populationsstruktur und die postglaziale Verbreitung von *A. thaliana* anhand der heutigen Verbreitung von genetischen Polymorphismen getroffen werden.

A. thaliana ist aufgrund ihrer Eigenschaften in der modernen Forschung ein ideales Objekt zum Studium der genetischen Variabilität und des Einflusses der Umwelt auf den Genotyp (Alonso-Blanco und Kornneef, 2000). Die Tatsache, dass natürliche Populationen von *A. thaliana* ideal für genetische und physiologische Studien geeignet sind, wurde schon früh

entdeckt (Laibach, 1940). *Arabidopsis* ist relativ nahe mit einer großen Anzahl landwirtschaftlich relevanter Arten verwandt. Dazu gehören *Brassica oleracea* L. (Broccoli, Rosenkohl, Kopfkohl, Grünkohl, Kohlrabi), *B. juncea* (L.) (Sareptasenf), *B. nigra* (L.) W.D.J. Koch (Schwarzer Senf), *B. napus* L. var. *napobrassica* (L.) Rchb. (Steckrübe), *B. napus* var. *napus* (Raps), *B. rapa* L. (Rübsen, Chinesensenf, Chinakohl); andere, lokal bedeutsame Kulturarten sind z.B. *Raphanus sativus* (Radieschen), *Sinapis alba* (Weißer Senf) und *Eruca vesicaria* (L.) Cavan. var. *sativa* (Mill.) Thell. (Rukola).

Ergebnisse, die an *A. thaliana* gefunden werden, können aufgrund von Sequenzhomologien innerhalb der DNA einfacher auf eng verwandte Arten als auf phylogenetisch weit entfernte Arten übertragen werden. Kuittinen et al. (2002) fanden interspezifische Sequenzhomologien bei der Untersuchung von 22 PCR-Primerpaaren für ökologisch relevante Gene wie z.B. für den Zeitpunkt der Reproduktion, den Sekundärmetabolismus und die Abwehr gegen Pathogene in *A. thaliana* und vier Verwandten, u.a. *B. oleracea*. Ein direkter Vergleich zwischen den Genomen von *B. oleracea* und *A. thaliana* ergab eine deutliche Kolinearität in einer Region von 2,1 Mbp (Lukens et al., 2003). Das entspricht 1,7% des Gesamtgenoms von *A. thaliana* bei einer geschätzten Genomgröße von 125 Mbp (*Arabidopsis* Genome Initiative, 2000).

Koch et al. (2001a) sequenzierten die konservierten Promotorregionen der Gene *Chs* (Chalconsynthase) und *Apetala3* innerhalb der Brassicaceae bei 22 Arten, darunter auch *A. thaliana*, *B. oleracea* und *Sinapis alba*. Sie fanden in ihrer Studie, dass wichtige regulatorische Sequenzelemente (z.B. die TATA-Box) innerhalb der Brassicaceae konserviert sind. Weiterhin fanden sie bisher unbekannte und hoch konservierte Sequenzmotive in den Promotorregionen, die aufgrund ihrer hohen Konserviertheit vermutlich regulatorische Elemente darstellen. Diese Basenabfolgen waren noch stärker konserviert als die bisher durch funktionelle Studien bekannten Motive.

Die Unterscheidung zwischen genetischer Variabilität und phänotypischer Plastizität bei der Adaptation einer Gruppe von Genotypen an ihre Umwelt ist auch für landwirtschaftlich relevante Fragestellungen von Bedeutung. Untersuchungen über die Reaktionsbreite von Genotypen und ihrer Widerstandsfähigkeit gegenüber Umweltschwankungen können zur Aufklärung des Anbauverhaltens, der Standortausnutzung sowie der Ertragsbildung der angebauten Sorten (Genotypen) beitragen.

Dabei ist ein einheitliches Maß an Plastizität innerhalb eines Genotyps (einer Sorte) erwünscht. Nur so können die Anforderungen des Saatgutverkehrsgesetzes an die

Homogenität einer Sorte gewährleistet werden (SaatVerkG, §32). Eine hohe Plastizität des Genotypes gegenüber schwankenden Umweltbedingungen kann dagegen positiv sein und Ertragsstabilität sichern.
In erster Näherung wird phänotypische Plastizität als „nicht-genetische“ Variablität den phänotypischen Unterschieden gegenüber gestellt, die auf genetischer Variabilität beruhen. Allerdings ist die Plastizität als Merkmal der Reaktionsnorm selbst genetisch bedingt und es ist durchaus möglich, genetische Variabilität für die Plastizität (die Anpassungs-fähigkeit) zwischen Genotypen einer Art zu finden.
Die Identifizierung von Plastizitätsgenen, die Schlichting und Pigliucci (1993) als regulatorische Gene definieren, welche die phänotypische Expression kontrollieren und unabhängig von Normwerten, d.h. Merkmalsmittelwerten einer repräsentativen Stichprobe sind, kann wichtig für die Erklärung von agronomischen Eigenschaften wie der Resistenz oder der Höhe des Ertrages bei bestimmten Umweltbedingungen sein.

Die Definition von Plastizitätsgenen ist jedoch umstritten, da nach Via (1993) Plastizität nicht das Ziel, sondern ein Nebenprodukt der natürlichen Selektion ist. Sie postuliert, dass Plastizität nicht unabhängig von Merkmalsmittelwert gesehen werden kann und es somit keine eigenständigen Plastizitätsgene gibt, sondern dass die in einigen Studien beobachtete Plastizität mithilfe von Genen, die direkt für die Merkmalsausprägung zuständig sind, beschrieben werden kann. Somit wird direkt auf das Merkmal und nur indirekt auf Plastizität selektiert. Andererseits hatte Wright bereits 1931 festgestellt, dass Plastizität wahrscheinlich das Hauptziel der Selektion ist, indem sie durch individuelle lokale Anpassung den Einfluß der Selektion auf den Genotyp mildert. Somit wird durch die phänotypische Plastizität der direkte Selektionsdruck auf den Genotyp verringert.
Wird die genetische Variation eines Merkmals direkt exprimiert und damit im Phänotyp sichbar (=„hard-wired“), kann die sichtbare phänotypische Variabilität mit der genetischen Variation gleichgesetzt werden. Die meisten Merkmalsausprägungen scheinen aber zumindest einen gewissen Grad an Plastizität zu haben und somit die direkte Selektion auf den Genotyp zu erschweren.

2.2. Single nucleotide polymorphism (SNP) - der (fast) ideale molekulare Marker?

Als Einzelnukleotidpolymorphismen (SNPs) werden spezifische Positionen im Genom bezeichnet, an denen innerhalb einer Art oder Population Punktmutationen, entweder Einzelbasen oder Insertionen/Deletionen (InDels) mit einer Länge von bis zu 20 bp, auftreten (Wang et al., 1998). Meist sind diese Mutationen biallelisch, d.h. in nur zwei Nukleotidzuständen ausgeprägt.

Im Folgenden wird gelegentlich über „Allele“ eines Basenpaares gesprochen. Da mehrere SNPs in einem Gen auftreten und unabhängig voneinander rekombinieren können, können auch bei „biallelischen“ SNPs mehr als zwei Allele von einem Gen vorliegen. Im Vergleich zu anderen molekularen Markern wie z.B. Mikrosatelliten sind sie langsam evolvierende, relativ stabile Marker.

SNPs gelten als die am weitest verbreiteten Polymorphismen im Genom. Beim Menschen schätzt man, dass sie ca. 90% der genetischen Variation im Genom ausmachen (Collins et al., 1998). Dabei können SNPs in kodierenden und nicht kodierenden Bereichen liegen und ihre Allele im kodierenden Bereich können somit Genprodukte und den Phänotyp eines Individuums direkt beeinflussen. SNPs werden in verschiedenen Arten im Durchschnitt alle 200-500 bp in nicht kodierenden und alle 500-1000 bp in kodierenden Regionen erwartet (Brumfield et al., 2003).

Bisher sind in verschiedensten Organismen, z.B. Mensch (*Homo sapiens*) (Sachidanandam, 2001; Venter et al., 2001), Maus (*Mus musculus*) (Lindblad-Toh et al., 2000), *Caenorhabditis elegans* (Wicks et al., 2001), *Drosophila melanogaster* (Hoskins et al., 2001) und *A. thaliana* (Cho et al., 1999; Jander et al., 2002; Schmid et al., 2003), eine Vielzahl von Einzelbasenunterschieden erfasst. Auf Grund ihrer Häufigkeit und Auftretens an vielen Positionen des Genoms sind sie ideal für wissenschaftliche Fragestellungen, z.B. in der Phylogeographie und für Kopplungs-, Assoziations- und Populationsstudien. Sie finden aber auch außerhalb der Forschung in der Diagnose, Pharmakogenetik und Forensik Anwendung.

SNPs können derzeit mit verschiedensten Methoden analysiert werden. Die jeweilige Methodenwahl bzw. -kombination hängt dabei von der Fragestellung des Projektes sowie den zur Verfügung stehenden Kapazitäten ab. SNP-Erfassungsmethoden sind sehr unterschiedlich und können grob in je zwei Arbeitsschritte unterteilt werden (Abb. 1). Dabei dient der erste Schritt zur Erfassung der SNP-Position, der zweite analysiert das jeweilige Allel. Die Erfassungsmethoden Hybridisierung, Primerverlängerung (PCR-

basiert), Oligonukleotid-Ligation (Ligase-basiert) und Nuklease-Restriktion (Restriktionsenzym-basiert) können mit den Analysemethoden Gelauftrennung (Agarose oder Polyacrylamid), Arrays (Glas oder Filter), Massenspektrometrie und Mikrotiterplattenlaser (Erfassung floureszenzmarkierter Proben) in fast jeder Weise kombiniert werden.

Somit müssen Vor- oder Nachteile in den Bereichen Kosten, Arbeitszeitaufwand, Automatisierung, Genauigkeit und Zuverlässigkeit für jede Methodenkombination jeweils einzeln bewertet werden. Generell gilt die DNA-Sequenzierung als Standard, an dem alle anderen Methoden vor allem in Bezug auf Genauigkeit und Zuverlässigkeit gemessen werden (Gut, 2001; Kwok, 2001; Syvänen, 2001). Die Nachteile dieser Methode für die SNP-Erfassung liegen in dem vergleichbar hohen Zeitaufwand für die Probenvorbereitung, möglichen Basenfehlpaarungen im Primerbereich (z.B. innerhalb von Genfamilien) und Unklarheiten im anschließenden Alignment.

In der vorliegenden Arbeit wurde ein Großteil der SNPs mithilfe der Sequenzierung erfasst. Weiterhin wurden cleaved amplified polymorphic sequences (CAPSs), die auf der Restriktion eines PCR-Amplifikates mit anschließender Gelauftrennung beruhen, verwendet. Diese Methode ist zuverlässig, kostengünstig und stimmte mit den Daten aus der Sequenzierung vollständig überein. Der hohe Probendurchsatz der CAPSs ist ideal für eine schnelle Erfassung von Polymorphismen. Nachteile liegen in der möglicherweise fehlenden Restriktionsschnittstelle um den SNP und der schlechten Erfassung von Fragmentgrößenunterschieden von unter 20 bp auf Agarosegelen.

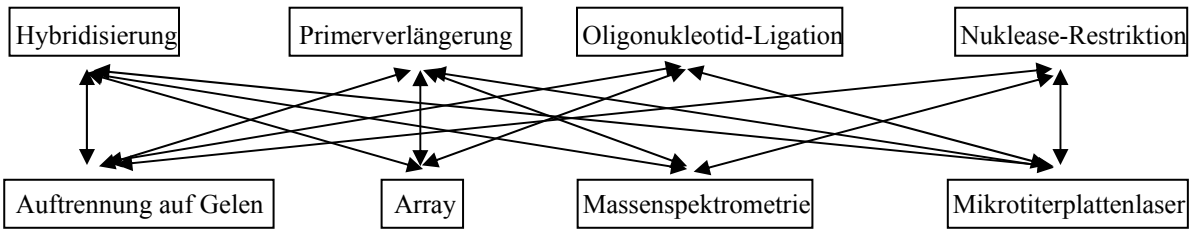


Abb. 1. SNP-Erfassungsmethoden aufgeteilt in Probenvorbereitung und Analyse nach Gut (2001).

Eine nicht zu unterschätzende und eventuell die größte Gefahr bei der Interpretation von SNPs liegt aber nicht im technischen Bereich, sondern in der Beeinflussung der Aussage durch die Methode, mit der die herangezogenen SNPs ausgewählt werden. Dabei können wiederum zwei Strategien (Brumfield et al., 2003) verfolgt werden: zum einen wird jede variable Position als SNP angesehen, zum anderen kann eine Mindestfrequenz innerhalb einer vorgegebenen Population festgesetzt werden, unter der variable Positionen nicht

berücksichtigt werden. Dadurch können Polymorphismen, die auf Sequenzierfehlern beruhen, ausgeschlossen werden. Dieses Vorgehen kann aber auch zum Verlust wertvoller Daten führen.

Eine gängige Strategie zur SNP-Detektion ist die Identifizierung von SNPs (meist durch Sequenzierung) an einem kleinen Testset (z.B. Törjék et al., 2003; Schmid et al., 2003) und die anschließende Untersuchung ausgewählter Polymorphismen am gesamten Datensatz. Das kann erheblich Zeit und Geld einsparen, da die identifizierten SNPs mit Nicht-sequenziermethoden (z.B. CAPSs, Massenspektrometrie oder Mikroarrays) detektiert werden können. Der entscheidende Nachteil liegt in dem Verlust von Variation und damit Information, bedingt durch die Auswahlkriterien des Testsets, das wegen der geringen Anzahl von Individuen oder einer Auswahl, die nicht repräsentativ für das spätere Untersuchungsmaterial ist, damit gewonnene Resultate beeinflussen kann (=ascertainment bias). Das betrifft sowohl informative Polymorphismen, die nicht im Testset gefunden wurden, als auch Assoziationen zwischen Polymorphismen, die innerhalb des Testsets, nicht aber für das spätere Untersuchungsmaterial signifikant sind. Wakeley et al. (2001) erachten SNPs nur dann als nützlich, wenn das Ausmaß des „ascertainment bias“ innerhalb der Studie abzuschätzen ist.

In einem Testset gefundene, in der Gesamtpopulation aber nur in geringer Frequenz vorhandene Polymorphismen, können eine falsche genetische Separation vorspiegeln. Entscheidende Polymorphismen können auf Grund des „ascertainment bias“ für nachfolgende Studien verloren gehen. Somit können sich falsche Schlußfolgerungen über genetische Unterschiede zwischen Organismen ergeben.

Nicht nur die Auswahl der untersuchten genetischen Polymorphismen, sondern auch die Auswahl der Individuen, die das Testset bilden, kann das Ergebnis einer Untersuchung stark beeinflussen. Auf die Zusammenstellung des Test- und Gesamtsets wird im Abschnitt 3.2. (Bisherige Beschränkungen der Stichprobenauswahl bei *Arabidopsis thaliana*) deshalb näher eingegangen.

3. *Arabidopsis thaliana* als Modell für Variabilität und Plastizität

3.1. Das Untersuchungsobjekt

A. thaliana (L.) Heynh. (Brassicaceae) (Al-Shebaz et al., 1999) wächst bevorzugt auf kargen Sandstandorten, z.B. in Kiefernwäldern oder Steppen, aber auch auf verwitterten Granitböden und anthropogen gestörten Flächen. Ihr natürliches (autochthones) Verbreitungsgebiet liegt in Eurasien (Abb. 2). In den letzten Jahren hat sich ihr Areal durch anthropogene Einflüsse bis nach Australien, Neuseeland, Südamerika und Afrika südlich der Sahara ausgedehnt (Hoffmann, 2002). Auch nach Nordamerika und Ostasien scheint *A. thaliana* nur mithilfe des Menschen gekommen zu sein. Akzessionen aus diesen Gebieten gruppieren in molekularbiologischen Studien eindeutig mit europäischen Akzessionen (Bergelson et al., 1998; Breyne et al., 1999; Miyashita et al., 1999; Vander Zwaan et al., 2000).

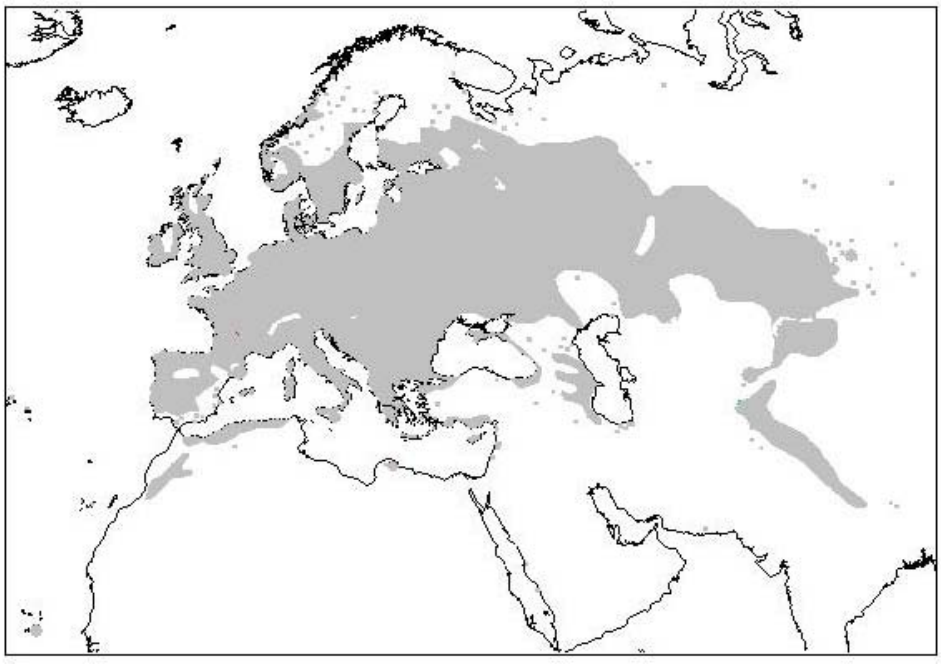


Abb. 2. Verbreitungsgebiet von *A. thaliana* nach Hoffmann et al. (2003b). Die grau schattierte Fläche repräsentiert das natürliche Areal.

Das Ursprungszentrum der Art vermutet Redéi (1969) in Europa, Berger (1965) in den Gebirgen des Westhimalaja. Auch Price et al. (1994) schlagen auf Grund von vergleichenden Untersuchungen mit verwandten Arten einen mittelasiatischen Ursprung von *A. thaliana* vor. Der letzte gemeinsame Vorfahre zwischen den Gattung *Arabidopsis*

und *Brassica* wird auf einen Zeitraum vor 20 Millionen Jahren geschätzt (Koch et al., 2001b).

Derzeit wird *A. thaliana* zur Bearbeitung verschiedenster biologischer Fragestellungen verwendet. Eigenschaften wie das kleine Genom (~130 Mbp) mit einem relativ geringen Anteil repetitiver DNA, eine geringe Chromosomenanzahl (n=5), ein schneller Reproduktionszyklus (~6 Wochen), einfache Kultivierung, minimale Platz- und Pflegeansprüche, eine hohe Fruchtbarkeit als spontaner Selbstbefruchter sowie eine Vielzahl an bekannten Markern machen sie dabei zur idealen Modellpflanze (Meyerowitz, 1989).

Um die evolutiven Prozesse und Interaktionen von Pflanzen in ihrer Umgebung verstehen zu können, ist es wichtig, mit natürlichen Akzessionen zu arbeiten. Dabei zeigt gerade *A. thaliana* eine bemerkenswerte Variabilität in vielen Merkmalen, z.B. Resistenzen gegen Insekten und Pilze, Toleranzen gegenüber Schwermetallen und Salz, winter- gegen sommeranuelle Lebensweise, aber auch in morphologischen Merkmalen wie z.B. Blattstruktur, Fruchtdiversität, Samenproduktion, Haarmorphologie (Koch et al., 2003).

Auch die Variabilität der Morphologie der ganzen Pflanze ist bemerkenswert (Abb. 3).

Abb. 3. *A. thaliana* nach 46 Tagen Wachstum bei 18°C (inklusive 7 Tage Vernalisation bei 4°C; eigene Anzucht). Akzessionen von links nach rechts: *Tadj* (Tadjikistan), *Cvi-0* (Cap Verden), *Pa-1* (Palermo, Italien).

A. thaliana hat eine größere klimatische Amplitude als andere Brassicaceae (Hoffmann, 2002). Durch ihre einjährige Lebensweise und ihre schlechte Konkurrenzfähigkeit unterliegt sie vermehrtem Selektionsdruck. Das macht sie auch für Studien zur natürlichen Variabilität und Plastizität zu einem idealen Untersuchungsobjekt.

3.2. Bisherige Beschränkungen der Stichprobenauswahl bei *Arabidopsis thaliana*

Derzeit sind ca. 750 natürliche Aufsammlungen (http://www.arabidopsis.org/info/about arabidopsis.jsp) in 3 Stock Centren erhältlich (ABRC: *Arabidopsis* Biological Resource Center an der Ohio State University, USA; NASC: Nottingham *Arabidopsis* Stock Centre an der Nottingham University, GB; SASSC: Sendai *Arabidopsis* Seed Stock Center an der Miyagi University of Education, Japan). Die eingelagerten Akzessionen stammen zum größten Teil aus Europa und Japan. Aus Mittelasien, dem postulierten Diversitätszentrum für *A. thaliana* (Price et al., 1994), stammen nur Aufsammlungen von 12 Standorten.

Bisherige Studien zur molekularen Variabilität und phänotypischen Plastizität (Übersicht in Pigliucci, 2001 und Hoffmann et al., eingereicht) wurden fast ausschließlich mit Akzessionen aus den Stock Centren durchgeführt. Asiatisches Material wurde nur aus den Stock Centren genutzt. Somit blieb ein großer Anteil der natürlichen Variabilität aufgrund der Stichprobenauswahl bei vorherigen Untersuchungen unberücksichtigt.

Diese einseitige Stichprobenauswahl der Akzessionen erklärt die bisher fast vollständige Abwesenheit der Korrelation von genetischer Variabilität mit geographischem Ursprung in Studien zur Erfassung der intraspezifischen genetischen Variabilität von *A. thaliana* (z.B. King et al., 1993; Ullrich et al., 1997; Bergelson et al., 1998 mit restriction fragment length polymorphisms (RFLPs); Breyne et al., 1999; Miyashita et al., 1999; Erschadi et al., 2000 mit amplified fragment length polymophisms (AFLPs); Innan et al., 1997; Loridon et al., 1998 mit Mikrosatelliten). Auch in Sequenzstudien einzelner Gene (Übersicht in Aguadé, 2001) oder ganzer Chromosomenregionen (Haubold et al., 2002; Nordborg und Tavaré, 2002) wurde keine Korrelation gefunden.

Sharbel et al. (2000) hatten ebenfalls diese Beschränkung in ihrem Probenmaterial. Als molekulare Marker benutzten sie AFLPs, die zufällig über das gesamte Genom streuen und fanden als Erste Anzeichen für signifikante genetische Isolation durch geographische Distanz zwischen 94 europäischen und nur 9 asiatischen Akzessionen. Dies zeigt, dass durch die Unabhängigkeit der AFLP-Marker über den gesamtem Datensatz die Suche nach spezifischen Markern für asiatische Akzessionen nicht von vorn herein erschwert wurde und somit eine Trennung zwischen europäischem und asiatischem Material gefunden werden konnte.

Zwei umfangreiche SNP-Datenbanken, die Cereon-Datenbank (Jander et al., 2002; http://www.arabidopsis.org/Cereon/index.html) und die MASC-Datenbank (Schmid et al., 2003; http://www.mipz-koeln.mpg.de/masc/), sind im Internet öffentlich zugänglich. Die

Polymorphismen der Cereon-Datenbank beruhen auf Sequenzvergleichen zwischen nur zwei Akzessionen, *Col* und L*er*. Diese Akzessionen werden entweder als single seed descent oder Liniengemische in den Stock Centren vermehrt und abgegeben. L*er* ist jedoch keine natürliche Akzession, sondern aus Mutation durch Röntgenstrahlen aus der natürlichen Aufsammlung *La-0* entstanden. Die Akzession *Col* ist aus *La-0* durch natürliche Auslese hervorgegangen (Nottingham *Arabidopsis* Stock Centre; http://nasc.nott.ac.uk/). Somit sind die detektierten Polymorphismen nicht repräsentativ für die natürliche Sequenzvariabilität.
In der MASC-Datenbank wurde ein Großteil der SNPs durch Sequenzvergleiche von Individuen aus bis zu zwölf ausschließlich europäischen Akzessionen detektiert. Es sind aber auch Daten über weitere Polymorphismen durch Sequenzvergleiche zwischen den zwei Akzessionen *Col* und *C24* (Törjék et al., 2003) enthalten. Bei der Akzession *C24* wird angenommen, sie sei natürlich, d.h. nicht durch Mutation erzeugt (Nottingham *Arabidopsis* Stock Centre; http://nasc.nott.ac.uk/), es gibt aber auch Hinweise, dass sie identisch mit der Akzession *Co-0* (Coimbra, Portugal) ist (M. Koornneef, pers. Komm.).

Auch phänotypische Untersuchungen wurden größtenteils mit europäischen Akzessionen durchgeführt (z.B. Westerman und Lawrence 1970; Pigliucci et al. 1999; Dorn et al. 2000; Pigliucci und Marlow 2001; Pollard et al. 2001). Dadurch blieben Akzessionen aus z.B. kontinentalen Gebieten oder großen Höhenlagen Asiens völlig unberücksichtigt. Das Gesamtspektrum der phänotypischen Variabilität ist somit wahrscheinlich nicht repräsentativ abgedeckt. Korrelationen zwischen phänotypischen Merkmalen und Längen- bzw. Breitengraden wurden bisher deshalb nur im europäischen Teil des Areals gefunden (Li et al., 1989, Stenoien et al., 2002). Dabei spielte die Auswahl der Akzessionen vermutlich eine große Rolle. Aus eigenen Untersuchungen mit Akzessionen aus Europa und Asien (Hoffmann et al., eingereicht) geht hervor, dass sowohl morphologische als auch phänologische Merkmale mit den monatlichen Mittelwerten von Temperatur und Niederschlag des Herkunftsortes korreliert sind. Demnach scheint es, als ob auch bei phänotypischen Untersuchungen die bisherige Akzessionsauswahl nicht repräsentativ für *A. thaliana* war.

In der vorliegenden Arbeit wurde versucht, den systematischen Fehler bei der Akzessionsauswahl durch neue Aufsammlung in Usbekistan und Sibirien zu minimieren.

3.3. Populationsstruktur von *Arabidopsis thaliana*

A. thaliana ist ein Selbstbefruchter mit einer natürlichen Auskreuzungsrate von unter 1% (Abbott und Gomes, 1989). Die genetische Variation innerhalb von Akzessionen ist viel geringer als zwischen Akzessionen (Kuittinen et al., 1997; Bergelson, 1998; Kuittinen, et al., 2000). Somit scheint die in vielen Studien gängige Stichprobe von einer Pflanze pro Akzession angemessen zu sein. Das hat weiterhin den Vorteil, dass ein größeres geographisches Areal abgedeckt werden kann. Dieser Ansatz schließt jedoch die Erfassung von eventuellen Variationen innerhalb einer Akzession aus. Weiterhin ist die Abgrenzung von geographisch dicht beieinander vorkommenden Akzessionen problematisch, da sie gelegentlich genetisch unterschiedlicher als geographisch weit voneinander entfernte Akzessionen sind (z.B. *Lisse-1* und *Lisse-2* im Gen *PISTILLATA*, Purugganan und Suddith, 1999). Wahrscheinlich können durch die hohe Selbstungsrate von *A. thaliana* die Pflanzen, die unabhängig voneinander die gleiche Lokalität besiedelt haben, genetisch unterschiedlich bleiben.

In allen bisherigen Sequenzstudien wurde ein Überschuß an niedrigfrequenten Allelen bzw. an Singletons (einmalig vorkommenden SNPs) gefunden (Überblick in Aguadé, 2001).

Das deutet nach der Theorie der neutralen Mutationen (Tajima, 1989) auf eine kürzlich erfolgte Arealausdehnung hin, die wahrscheinlich nach der letzten Eiszeit erfolgt ist (Aguadé, 2001). Dabei geht die Berechnung von Tajima´s D, dem aus den Daten resultierenden Wert zur Annahme oder Ablehnung der neutralen Mutation, von den Voraussetzungen einer zufälligen Stichprobe, keiner Rekombination und der Erfüllung des Hardy-Weinberg-Gleichgewichts für die Population aus. Diese Voraussetzungen sind in diesen Untersuchungen nicht vollständig erfüllt, z.B. das Hardy-Weinberg-Gleichgewicht tritt theoretisch nur bei Fremdbefruchtern auf. Ein negatives signifikantes Tajima´s D wird dabei bei wenigen genetischen Polymorphismen errechnet und wird als kürzlich erfolgte Selektion durch z.B. letale Mutationen oder durch Flaschenhalseffekte interpretiert. Ein signifikant positives Tajima´s D wird bei vielen DNA-Polymorphismen wie im Falle von *A. thaliana* errechnet und als kürzliche Ausdehnung der Population, z.B. bei Arealerweiterung, gewertet (Tajima, 1989).

Bei vielen Sequenzstudien wurde ein genomweites Auftreten von Dimorphismen (nur zwei Allele pro Locus) und zwei daraus resultierende Consensushaplotypengruppen beobachtet (Innan et al., 1996: *Adh*; Kawabe und Miyshita, 1999: *ChiB*; Stahl et al., 1999: *Rpm1*;

Aguadé, 2001: *FAH1*, *F3H*; Hauser et al., 2001: *GL1*). Durch die unterschiedliche Akzessionsauswahl in den verschiedenen Studien konnten diese Gruppen jedoch nicht klar abgegrenzt werden. Die höchste Nukleotiddiversität wurde zwischen den zwei jeweiligen Gruppen gefunden, mit geringer Variation innerhalb der Haplotypen. Weiterhin konnten Rekombinanten in den Regionen *Adh*, *ChiA*, *Rpm1*, *FAH1*, *F3H* und *GL1* zwischen den zwei Gruppen detektiert werden. Dass Rekombination innerhalb von *A. thaliana* in einem kleinen genomischen Abschnitt auftreten kann, zeigten Hanfstingl et al. (1994). Sie fanden Rekombination in der *Adh*-Region im Abstand von 350 bp. Durch die große Anzahl an Individuen von *A. thaliana* scheint es möglich, dass sich einige der zum geringen Maße entstehenden Rekombinanten teilweise innerhalb von Populationen halten können.

3.4. Intraspezifische Variabilität von Genomgrößen

Die DNA-Gehalte der haploiden Genome können innerhalb der Angiospermen um bis zu 3 Potenzen variieren (Bennett und Leitch, 1995). Das Auftreten von Genomgrößenunterschieden unterhalb der Artebene wird jedoch kontrovers diskutiert (z.B. Greilhuber, 1998). Unterschiedliche intraspezifische DNA-Gehalte können durch Chromosomenaberrationen hervorgerufen werden. Diese Mutationen können entweder zum Verlust von Nukleotiden (Deletion, Chromosomenverschmelzung oder Abbrechen von Chromosomenteilen), zur Vervielfältigung von Nukleotiden (Insertionen, Duplikationen) oder zur Umstrukturierung (Chromosomenbrüche) führen (Stebbins, 1971).

Greilhuber (1998) führte die Mitteilungen über innerartliche Genomgrößenunterschiede in vielen Studien auf taxonomische Fehlbestimmungen (z.B. *Scilla* ssp.) oder technische Fehler zurück (Anfärbung bei verschiedenen Temperaturen, gestörte Anfärbung durch polyphenolhaltige Inhaltsstoffe wie z.B. Tannin vor allem bei Feulgenfärbung, externe Standards bei Messungen). Weitere Fehlerquellen können aber auch durch die Benutzung unterschiedlicher Messgeräte (z.B. Flowcytometer mit Laser oder Lampe) entstehen. Dolezel et al. (1998) fanden in einem Vergleich zwischen 4 Laboratorien Abweichungen von 42% bei der Messung der Genomgröße von *A. thaliana*. Dabei wurde in allen Laboratorien mit einem einheitlichen Ausgangssstandardwert von *Allium cepa* mit 33.5 pg begonnen.

Viele Studien versuchten, den exakten Gehalt der nuklearen DNA von *A. thaliana* zu bestimmen, weil es sich dabei um einen der kleinsten Werte höherer Pflanzen handelt. Die

angegebenen Werte für den haploiden Chromosomensatz (1C-Wert) haben eine beachtliche Schwankungsbreite zwischen 0,051 pg (Francis et al., 1990) und 0,215 pg (Dolezel et al., 1998). Die *Arabidopsis* Genome Initiative (2000) schätzte nach der fast vollständigen Sequenzierung einen 1C-Wert von 0,128 pg. Dieser Wert wurde von Bennett et al. (2003) nach vergleichenden Messungen mit dem vollständig sequenzierten Genom von *Caenorhabditis elegans* nach oben korrigiert (~0,16 pg). Die meisten Studien benutzten für ihre Messungen nur die Akzession *Col*, innerartliche Genomgrößenmessungen wurden bisher noch nicht mit einer repräsentativen Stichprobe von Akzessionen durchgeführt.

3.5. Ergebnisse

Das Ziel der vorliegenden Arbeit war es, die genetische Variabilität natürlicher Aufsammlungen von *A. thaliana* in ihrem gesamten eurasischen Areal anhand von gleichmässig über das ganze Genom verteilten Markern zu erfassen und Assoziationen zwischen Markerallelen und geographischen sowie morphologischen und phänologischen Eigenschaften zu finden. Dazu wurde die genetische Variabilität auf Nukleotidebene mit Einzelnukleotidpolymorphismen (single nucleotide polymorphisms, SNPs) analysiert. Morphologische und phänologische Eigenschaften der Akzessionen wurden in Temperaturversuchen bei zwei unterschiedlichen Temperaturen erfasst. Genomgrößenunterschiede wurden mithilfe der Durchflußzytometrie gemessen.
Die daraus resultierenden Ergebnisse finden sich wie folgt in den vier Artikeln wieder:

In Artikel 1 (Schmuths et al., 2004b) wurde mithilfe einer genomweiten CAPS (cleaved amplified polymorphic sequence)-Studie in 49 weltweiten Akzessionen erstmals ein Hinweis auf eine Korrelation zwischen geographischer Herkunft und genetischer Variabilität mithilfe von Markern mit bekannter Lage im Genom (nicht anonym) bei *A. thaliana* gefunden. Aufgrund dieses Hinweises wurden die flankierenden Regionen um diese Punktmutationen sequenziert. Diese zwei sequenzierten Regionen befinden sich im Abstand von ~300 kbp auf Chromosom 2. Die sich ergebene geographische Trennung zwischen Ost und West findet sich auch im Gesamtgenom wieder (Artikel 4, Schmid et al., eingereicht). Die Abdeckung des gesamten eurasischen Areals durch zusätzliche

gesammelte sibirische Akzessionen und deren Sequenzierung reduzieren tendenzielle Einflüsse bei der Akzessionsauswahl auf ein Minimum.

In Artikel 2 (Schmuths et al., 2004a) wurden Genomgrößen von 21 eurasischen Akzessionen mithilfe der Durchflußzytometrie gemessen. Es wurden signifikante intraspezifische Unterschiede zwischen den 19 diploiden Akzessionen gefunden. Zwei der untersuchten Akzessionen waren tetraploid.

In Artikel 3 (Hoffmann et al., eingereicht) wurden die morphologischen und phänologischen Unterschiede innerhalb von *A. thaliana* in zwei unterschiedlichen Temperaturregimen (14°C und 22°C) untersucht. Zahlreiche Korrelationen zwischen morphologischen, phänologischen, molekularen und geographischen Merkmalen wurden berechnet.

Pleistocene Refugien und die postglaziale Expansion von *A. thaliana* wurden in Artikel 4 molekular untersucht. Dazu wurden 335 eurasische Akzessionen mit einem genomweiten Set von 115 SNPs mithilfe von Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight (MALDI-ToF, Haff and Smirnov, 1997) untersucht. Da die SNPs zwischen nur zwei Akzessionen detektiert wurden, zeigt diese Studie deutlich, wie sich der systematische Ansatz der SNP-Auswahl auf das Versuchsergebnis auswirkt.

Geographic Distribution and Recombination of Genomic Fragments on the Short Arm of Chromosome 2 of *Arabidopsis thaliana*

H. Schmuths, M. H. Hoffmann, and K. Bachmann. 2004. Plant Biology. 6: 128-139.

Abstract

Range expansion from pleistocene refugia and anthropogenic influences contribute to the present distribution pattern of *Arabidopsis thaliana*. We scored a genome-wide set of CAPSs and found two markers with an East-West geographic distribution across the Eurasian range of the species. Regions around the two SNPs were sequenced in 98 accessions including newly collected plants from Middle Asia and Western Siberia. These regions correspond to a gene (~1500 bp) and a non-coding region (~500 bp) 300 kbp apart on chromosome 2. Nucleotide diversities, π, of the two sequenced fragments were 0,0032 and 0,0130. The haplotypes of both sequences belonged to either of two groups, a rather uniform „Asian“ and a more variable „European“ haplotype group on the basis of non-disjunct clusters of SNPs. Recombination between „Asian“ and „European“ haplotypes occurs where they meet. Especially in the „European“ haplotype, many rare SNP variants representing independent mutations are scattered among the shared haplotype-specific SNPs. This agrees with previous suggestions of two large haplotype groups in *A. thaliana* and the postglacial colonisation of central Europe from the East and the West. A clear correlation between climatic factors and the haplotype distribution may reflect the dispersal history rather than local climate adaptation. The pattern of SNP variation within the contiguous sequences explains why only a minority of SNPs selected across the genome show evidence of this geographic pattern.

Keywords: sequence polymorphism, *Arabidopsis thaliana*, natural populations, biogeography, pattern, recombination

Introduction

„Phylogeography“, the comparison of the geographic distribution of molecular markers with their phylogenetic relationships, is widely used to reconstruct species histories (Avise et al., 2000). This approach has revealed various routes for the recolonisation of central and Northern Europe after the ice ages in a variety of species (Hewitt et al., 1999; Comps et al., 2001; Stehlik et al., 2002; Wares, 2002). Recent distribution by humans impedes the easy detection of postglacial recolonisation patterns, especially in a model research plant like *Arabidopsis thaliana*. Up to now, geographic patterns in the distribution of molecular markers in *A. thaliana* were only found by Sharbel et al. (2000) using AFLPs, and in a meta-data analysis by Hoffmann et al. (2003b) using all previously published DNA sequences. Other studies with anonymous markers (e.g. King et al., 1993; Ullrich et al., 1997; and Bergelson et al., 1998 with RFLPs; Breyne et al., 1999; Miyashita et al., 1999; and Erschadi et al., 2000 with AFLPs; and Innan et al., 1997; and Loridon et al., 1998 with microsatellites) and sequence data from single genes (summary in Hoffmann et al., 2003b) revealed no clear geographic patterns. Recent studies were carried out with more accessions on regions spanning 400 kbp around the *FRI* locus on chromosome 4 (Hagenblad and Nordborg, 2002) and an analysis of recombination in a region of 170 kbp on chromosome 5 (Haubold et al., 2002).

The previous studies have resulted in a better understanding of the intraspecific evolutionary history of *A. thaliana* (reviewed in Haubold et al., 2002; Kuittinen et al., 2002). Three observations are supported by these studies: first, an excess of rare polymorphisms could be observed in the DNA sequences (Purugganan and Suddith, 1998, 1999). Second, most polymorphisms in *A. thaliana* are alternatives between only two nucleotides (e.g. Aguadé, 2001). Third, comparisons of sequences among accessions of *A. thaliana* show a surprisingly large number of recombination events for a predominantly selfing species (reviewed in Kuittinen and Aguadé, 2000; Haubold et al., 2002). Estimates of outcrossing rates from molecular data vary between 0,3 and 1.0 % (Abbott and Gomes, 1989; Bergelson et al., 1998). Kuittinen et al. (2002) also observed patterns characteristic for inbreeding species. Recombination involving small (350 bp; Hanfstingl et al., 1994) or large (150 kbp; Nordborg and Tavaré, 2002) segments was demonstrated. The results in *A. thaliana* and similar ones in another selfing plant species, *Hordeum spontaneum,* (Lin et al., 2002) led to the conclusion that an appreciable number of recombination events can accumulate in highly selfing taxa.

Recently, we scored 67 cleaved amplified polymorphic sequences (CAPSs) in 49 accessions all over the nuclear genome and looked for correlations of individual markers with morphological and phenological features. The morphological and phenological correlations are presented in Hoffmann et al. (in review). During this analysis, we found that three of the 67 markers showed a striking east-west geographical separation. This raised the question if this observation was a statistical artifact, and if not, why there was no significant geographical signal in the majority of the randomly chosen single nucleotide polymorphisms (SNPs). We investigated this by sequencing longer contiguous regions around the CAPS markers and found that the geographic pattern for one of the three was not paralleled by linked markers. However, a large number of SNPs in a gene (~1500 bp) and a non-coding region (~500 bp), 300 kbp upstream from this gene on chromosome 2, that were sequenced in 98 accessions, including newly collected accessions mainly from Uzbekistan and Siberia, the Eastern margin of the range of *A. thaliana* (Hoffmann, 2002), revealed two versions of the same evolutionary and geographic history of the accessions that differed only in the greater amount of mutation and recombination in the non-coding region as compared to the coding sequence. Our results support the conclusion by Sharbel et al. (2000) that the (weak) geographic signal in the intraspecific genetic variation of *A. thaliana* reflects the postglacial encounter between two genetically differentiated subgroups, one extending out of Asia, one from Southern Europe. We predict that comparative sequencing sufficiently long contiguous stretches of sequence anywhere in the genome of *A. thaliana* will identify clusters of non-disjunct „old“ polymorphisms among recent independent mutational events and add further details to the post-glacial evolutionary history of the species.

Material and Methods

Plant material

98 accessions were included in this study, of which 57 were obtained from the Nottingham *Arabidopsis* Stock Centre and 41 were collected in the wild (Table 1). Genomic DNA of single plants was extracted from dried leaf material using the DNeasy extraction kit (QIAGEN, Hilden) according to the manufacturer´s instructions.

Table 1. List of accessions collected in the wild.

accession	location
Sie	Siena, Italy
Par	Ville de Paris, France
Gat	Gatersleben, Germany
Bad	Badetz, Germany
Wt	Wittenberg, Germany
Han	Handorf, Germany
Gie	Gievenbeck, Germany
Dül	Dülmen, Germany
Sij-1, -2, -4	Sijak, Uzbekistan*
Kly-1, -2, -3, -4, -6	Kolyvan, Russia*
K-oz-1, -2, -3	Kolyvanskoe ozero, Russia*
Leb-1, -2, -3, -4	Lebjaschje, Russia*
Nov-1, -2, -3	Novojegorjevskoje, Russia*
Nos	Novosovjetski, Russia
Rak-1, -2, -3	Rakity, Russia*
Bas-1, -2, -3	Bastan, Russia*
Sev	Sewerka, Russia
Cha-1, -2	Chabary, Russia*
Pan	Pankruschicha, Russia
Mas	Masljacha, Russia
Bij	Bijsk, Russia
Mal	Puerto de Soller, Majorca
Dar	Darjali, Georgia

* Accessions with identical names but different numbers (e.g. Rak-1, -2, -3) are geographically separated by at least 10 km.

CAPSs and DNA sequencing

67 single nucleotide polymorphisms (SNPs) from all across the nuclear genome were scored as CAPSs. K. J. Schmid, MPI Jena, Germany, kindly supplied 56 primer pairs (sequences and SNP positions available in MASC data base (MASC02639, MASC02712, MASC02716, MASC02734, MASC02790, MASC02895, MASC02948, MASC02949, MASC03021, MASC03091, MASC03109, MASC03114, MASC03158, MASC03218, MASC03221, MASC03332, MASC03336, MASC03341, MASC03400, MASC03465, MASC03522, MASC03603, MASC03644, MASC03765, MASC03799, MASC04010, MASC04053, MASC04054, MASC04181, MASC04209, MASC04217, MASC04308, MASC04366, MASC04401, MASC04538, MASC04655, MASC04725, MASC04825, MASC05070, MASC05074, MASC05116, MASC05139, MASC05302, MASC05383, MASC05385, MASC05386, MASC05398, MASC05481, MASC05622, MASC05800, MASC05813, MASC05828, MASC06045, MASC06057, MASC06108, MASC06205) (Schmid et al., 2003)). Further primer sequences and protocols are available from the author upon request.

The geographic distributions of the alternate alleles were examined using the Geographical Information System Arc/Info (ESRI, 1992). Three stretches of ~500 bp around CAPSs with an apparently non-random, essentially East-West geographical distribution were sequenced in two European accessions with the „Western" CAPS allele and two Asian accessions with the „Eastern" allele. The geographic signal in one CAPS fragment (amplified with MASC05074) was limited to the single nucleotide within the sequence and probably a statistical artifact. This fragment was not examined further. The geographic patterns of the remaining two CAPSs, amplified with primers MASC05800 and MASC06057, were paralleled by additional SNPs in the surrounding sequence suggesting patches of non-disjunction. MASC05800 and the surrounding regions amplified with primer pairs At2g06530-1 and At2g06530-2 were sequenced in 98 accessions. The MASC06057-region could be sequenced only in 95 of these, since accessions *Stw-0*, *Ws* and *Oy-0* yielded no amplicon.

We sequenced the gene At2g06530 surrounding fragment MASC05800 based on information available at that time (MAtDB, from MIPS, http://mips.gsf.de). However, in the course of this study a new interpretation of the gene sequence was published, so that the region sequenced by us misses the first exon and parts of the first intron.

PCR products were purified with the QIAquick PCR purification kit (QIAGEN, Hilden) and used as templates for cycle sequencing with Rhodamine-Mix Kit by ABI (Foster City, CA). Both strands were sequenced. In order to obtain some information about the 300 kbp between the two sequenced sites, the list of the polymorphic sites between the accessions *Col* and L*er* (Cereon database, Jander et al., 2002) was examined. Four CAPSs were selected two of which are placed in a distance of ~400 bp of each other. Additional primers for these two regions (C-t9f8, CAPSs located in At2g06840, a putative retroelement polyprotein gene, and C-t4e14, CAPSs located in a non-coding region) were designed based upon the published sequence of the accession *Col* using the computer programme PRIMER3 (http://www.genome.wi.mit.edu/genome_software/other/primer3). Amplicons were cleaved with the enzymes *Taq*I/*Alu*I and *Taq*I/*Mae*I (MBI-Fermentas, St. Leon Roth; Roche, Mannheim), respectively, following the manufacturer´s instructions.

Analysis

Sequences were edited using „Chromas 1.62“ (http://www.technelysium.com.au/chromas.html) and aligned with Bioedit (Hall, 1999). Gaps were treated as fifth character, InDels were treated as single events. The average number of nucleotide differences per site, π (Nei, 1987), and the minimum number of recombination events, R_m (Hudson and Kaplan, 1985), was calculated. To determine if the sample was in mutation-drift equilibrium Tajima´s test of neutral evolution for nucleotide sequences (Tajima, 1989) was calculated. These analysis were performed using DnaSP (Rozas and Rozas, 1999). Minimum spanning trees were estimated with TCS (Clement et al., 2000) which uses a parsimony algorithm as defined in Templeton et al. (1992).

Mantel tests were computed with the normalized Mantel statistic R_M with 999 permutations using R PACKAGE 4.0 (Casgrain and Legendre, 1999). Accessions from America, Japan, the Canary Islands and Cape Verde Islands were excluded from the Mantel test with distance classes. The „classical“ overall correlation between a Euclidean genetic distance matrix and the geographic distance matrix (Mantel, 1967) and the correlation of the genetic distance matrix between all pairs of individuals and specifically adapted geographic model matrices (distance class matrices, Oden and Sokal, 1986) were computed. We chose 30 distance classes in steps of 235 km, i.e. distance class 1 comprises accessions up to 235 km apart. Corresponding R_M-values were computed individually and plotted against distance classes. Mantel correlograms calculated with distance classes can reveal general patterns within the data but cannot indicate how specific accessions are correlated with each other.

RecPars (Hein, 1990, 1993) and RecMin (Myers and Griffiths, 2003) were used as tests for recombination. RecPars performs a parsimony analysis (Hein, 1990, 1993) of a set of DNA sequences. It tries to find the best phylogeny for different segments of the sequences and thereby to postulate a recombination event between these segments. Since RecPars becomes computationally limited with a large number of sequences because the method must consider a large number of possible evolutionary trees at each locus, we only used one sequence per haplotype. Substitution costs were set to 100 for each substitution or InDel. We used two rounds of parses, which means that the sequences were scanned first forward and then backward.

RecMin includes two new statistics that compute two lower bounds on the number of recombinations, R_h and R_s, that offer an improvement over the minimum number of recombination R_m (Hudson and Kaplan, 1985). Thus, for any given data set is $R_s \geq R_h \geq R_m$

(Myers and Griffith, 2003). R_h, based on bounding the number of recombination events by calculating the difference between the number of observed types in the sample and the number of segregating sites, was calculated with the maximal subset size 6, width 15 and unknown ancestral types for the two datasets. R_s bounds the number of recombinations by approximating the history of the data using a simplified version of recombination events, in such way that any true history for the data has more recombination events than one of these approximate histories. It was incomputable due to the large size of the dataset (Myers and Griffiths, 2003).

Since the variation among the sequences consists of a complex pattern of recombination and mutation within the recombining blocks, split decomposition (Bandelt and Dress, 1992) was calculated using SplitsTree, version2 (Huson, 1998). In contrast to maximum parsimony and maximum likelihood, this method is transformation based and does not optimize the parameters for estimating phylogenetic trees. Split decomposition determines the maximum distance between accessions by "canonical decomposition", i.e. a fragmentation of the data set. This results in a sum of "weakly compatible splits", i.e. groups with nearly no similar elements (Huson, 1998). Conflicting relationships suggesting recombination are not resolved but represented by a net-like structure.

The splits graph was obtained by applying the Hamming distance, which determines the number of characters that need to change from one state to another, and split decomposition. Characters with a frequency of less than 5% and constant characters were excluded from the analysis (Lockhart et al., 1996; Huson, 1998) resulting in clearer splits and preventing an overestimation of singletons and low frequency SNPs. Triangle inequalities in the tree were forced. This approach checks for a given distance whether the triangle inequalities hold; if they do not, then the „Force Triangle Inequalities" can be used to force them by adding an appropriate offset to all distances (Huson, 1998).

Climate data were kindly provided by W. Cramer (CLIMATE database version 2.1, Potsdam Institute of Climate Impact Research, Germany). The data of the monthly means of temperature and precipitation have a longitude/latitude resolution of 0,5x0,5 degree. Since there was no normal distribution of the temperature and precipitation data (Kolmogorov-Smirnov test), the mean of these data could not be used for further parametric statistical tests. Sachs (1997) proposes the use of the median in non-parametric tests for data with asymmetrical distribution. Sequences were encoded in a 0-1-matrix and divided on the median of temperature and precipitation into two groups for each month, respectively. Sequences of accessions identical to the median were not grouped but

excluded from the calculations. Significant differences between the two groups were estimated with the Mann-Whitney-U-Test. The Kolmogorov-Smirnov test for normality and the Mann-Whitney-U-Test were computed with SPSS 10.0.
Geographic visualisation was performed using the Geographical Information System Arc/Info (ESRI, 1992).

Results

Genome wide CAPSs marker survey

All 67 genome wide CAPSs are alternatives between only two nucleotides throughout the sample. Frequencies for the rarer of the two nucleotides are shown in Fig. 1. A small group of 5 CAPS (7,5%) has a frequency of less than 0,1. The frequency of the majority of the SNPs (80,6%) ranges between 0,2 and 0,4. About 12% of the substitutions have frequencies higher than 0,4. A more detailed analysis of the data is presented in Schmuths et al. (in prep.).

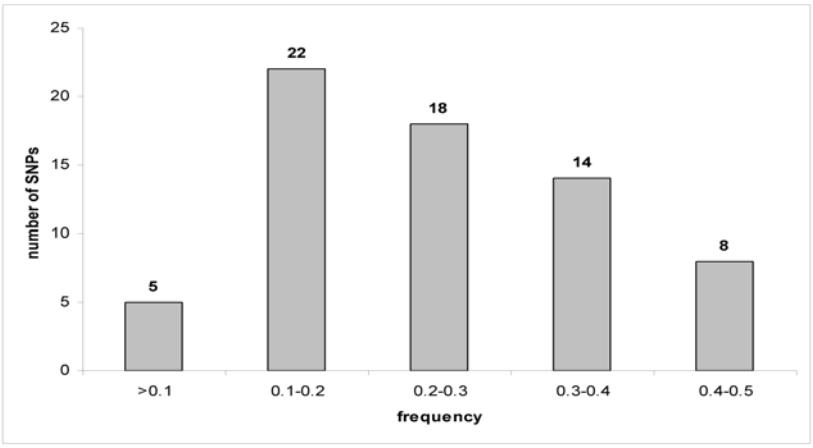


Fig. 1. Allele distribution of the 67 CAPS investigated in 49 accessions (for details refer to the text).

We analyzed the data set of 49 accessions for non-random patterns of geographic distribution of their alternative alleles across the Eurasian sampling area. Three such CAPSs were detected by a Kolmogorov-Smirnov test: MASC06057 (p=0,0001), MASC05800 (p=0,0002), and MASC05074 (p=0,019). The remaining 64 CAPSs with P-values above 0,039 were not further examined. To elucidate whether these significances below 0,02 were true for the entire genomic region or mutational or statistical artefacts, sequences (~500 bp) surrounding these CAPSs were determined in a subset of two European and two Asian accessions. In two of the sequenced regions, SNPs near the

selected one showed a correlated geographical pattern. The geographic signal in the third region (amplified with MASC05074) was limited to the selected CAPS marker even though the number of SNPs among the four accessions in the third sequenced region was similar to the two other fragments. This region was not studied further. The two regions with an extended geographical signal are located in a distance of 300 kbp on chromosome 2. One fragment (amplified with MASC05800) is located in an expressed gene with unknown function, At2g06530, whereas the second fragment (amplified with MASC06057) is situated in a non-coding region.

Sequence analysis

We scored 102 SNPs, 2 InDels and 4 CAPSs in 98 accessions revealing 53 different haplotypes (Fig. 2, EMBL accession numbers AJ628449-AJ628641). The nucleotide diversities, π, of the sequenced ~1500 bp fragment in At2g06530 (in the following named fragment 1) and the ~500 bp region amplified with MASC06057 (in the following named fragment 2) are 0,0032 and 0,0130, respectively. When the analysis was limited to third positions in the coding region and non-coding positions of fragment 1 the nucleotide diversity increased to π_{Silent}=0,0041.

Computations of neutral evolution of the sequences (Tajima, 1989) resulted in a negative and not significant Tajima´s D; fragment 1: D=-1,21 (p>0,1), fragment 2: D=-1,31 (p>0,1). This indicates neutral evolving sequences with an excess of rare polymorphisms due to bottlenecks or population fusion (e.g. Haubold et al., 2002).

We calculated recombination parameters with the minimum number of recombination events, R_m (Hudson and Kaplan, 1985), with RecPars (Hein, 1990, 1993) and with R_h (Myers and Griffiths, 2003). The possibility of recombination in fragment 1 was rejected by all methods (R_m, RecPars=0 and R_h). In fragment 2 recombination events were suggested by all methods. R_m estimated 7 recombination events, RecPars resulted in two recombination phylogenies (tree 1: cost=1000 and recomb=1, tree 2: cost=2300 and recomb=1). The number of recombinations, R_h, was 9. One recombination event each was suggested between positions 2882636 and 2882651, 22882653 and 2882662, 2882735 and 2882737 and 2882743, 2882748 and 2882751 and 2882791, 2882818 and 2882827. Between positions 2882907 and 2882942 two recombination events were estimated.

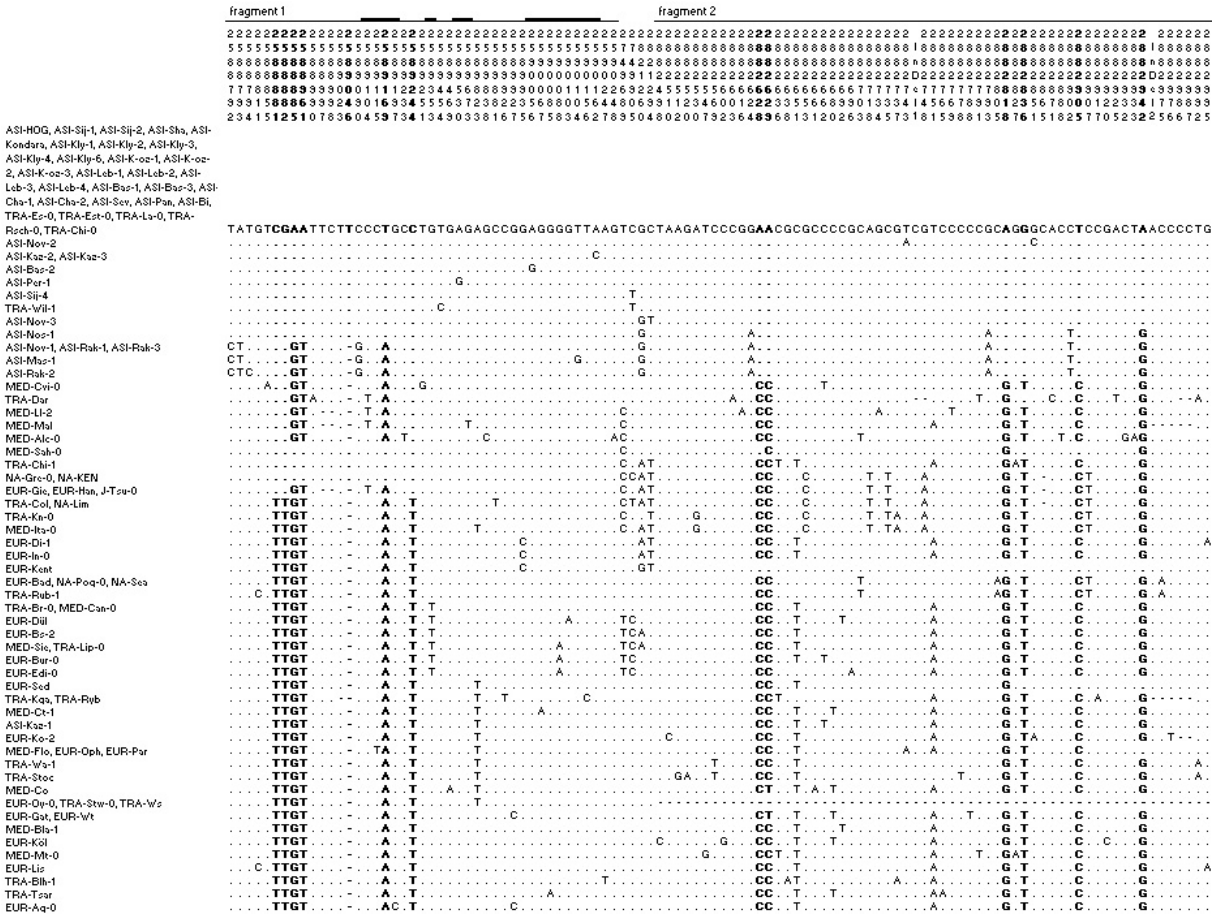


Fig. 2. 108 polymorphic sites investigated. Positions on chromosome 2 are indicated by above numbers. Alleles similar to the first sequence are indicated by dots (.), deletions are indicated by dashes (-). Bolted numbers indicate the „Asian" and the „European" consensus sequences. Positions between 2749689 – 2821294 were detected with CAPSs (see CAPSs and DNA sequencing). Alleles similar to the first sequence indicate identical substitutions but smaler (2749689 and ..905) or not cleavable amplicons (2882969 and ..986). InDel1 is 4 bp (CTGA) and InDel2 is 17 bp (AGAAATCCCTAAAAGGT). Used prefix stands for: ASI = Asian accessions (eastwards of 50° longitude), EUR = European accessions (westwards of 15° longitude), TRA = transition accessions (located between European and Asian accessions), MED = Mediterranean accessions (southwards of the Alps and Pyrenees), NA = North American accessions, J = Japanese accessions (regions were choosen mainly following suggestions of Hewitt (1999)).

Genetic and geographic relationships

Calculating separate minimum-spanning trees with the SNPs sets of the two 300 kbp distant regions revealed groups of similar accessions corresponding to geographic areas of origin (Fig. 3). These trees obviously do not reflect the phylogeny (mutational history) of the sequences (Templeton et al., 1992) but reflect both phylogenetic information and the recombination history of the sequences.

Similar haplotype patterns in both fragments separate the accessions into three main geographic groups (for explanation see Fig. 2): (1) The large „Asian group", comprising Middle Asian, Eastern European and three American accessions. A Spanish accession

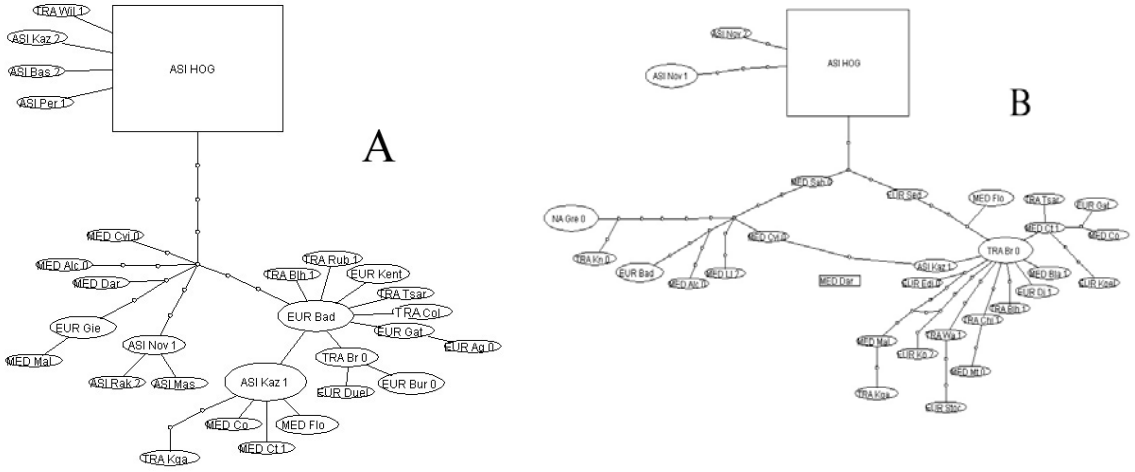


Fig. 3. Minimum-spanning-trees calculated with a parsimonious algorithm. **A**: fragment 1: 95% connecting limit is 16 steps **B**: fragment 2: 95% connecting limit is 9 steps. For used prefix see Fig. 2. Accessions not listed below present single haplotypes.
A: ASI-HOG = ASI-HOG, ASI-Sij-1, ASI-Sij-2, ASI-Sij-4, ASI-Sha, ASI-Kondara, ASI-Kly-1, ASI-Kly-2, ASI-Kly-3, ASI-Kly-4, ASI-Kly-6, ASI-K-oz-1, ASI-K-oz-2, ASI-K-oz-3, ASI-Leb-1, ASI-Leb-2, ASI-Leb-3, ASI-Leb-4, ASI-Nov-2, ASI-Nov-3, ASI-Nos, ASI-Bas-1, ASI-Bas-3, ASI-Cha-1, ASI-Cha-2, ASI-Sev, ASI-Pan, ASI-La-0, ASI-Bij, TRA-Chi-0, TRA-Chi-1, TRA-Es-0, TRA-Est-0, TRA-Rsch-0, NA-Gre-0, NA-KEN, MED-Sah-0; **ASI-Kaz-2** = ASI-Kaz-2, ASI-Kaz-3; **EUR-Gie** = EUR-Gie, EUR-Han, J-Tsu-0, MED-Ll-2; **ASI-Nov-1** = ASI-Nov-1, ASI-Rak-1, ASI-Rak-3; **EUR-Bad** = EUR-Bad, EUR-Koel, MED-Bla-1, MED-Mt-0, TRA-Kn-0, NA-Pog-0, NA-Sea; **EUR-Kent** = EUR-Kent, EUR-Di-1, EUR-In-0; **TRA-Col** = TRA-Col, NA-Lim; **TRA-Rub-1** = TRA-Rub-1, EUR-Lis; **EUR-Gat** = EUR-Gat, EUR-Wt; **TRA-Br-0** = TRA-Br-0, EUR-Bs-2, MED-Can-0; **EUR-Bur-0** = EUR-Bur-0, EUR-Edi-0, MED-Sie, TRA-Lip-0; **ASI-Kaz-1** = ASI-Kaz-1, EUR-Oy-0, EUR-Sed, EUR-Stoc, EUR-Ko-2, TRA-Stw-0, TRA-Wa-1, TRA-Ws, MED-Ita-0; **MED-Flo** = MED-Flo, EUR-Oph, EUR-Par; **TRA-Kga** = TRA-Kga, TRA-Ryb
B: ASI-HOG = ASI-HOG, ASI-Kaz-2, ASI-Kaz-3, ASI-Sij-1, ASI-Sij-2, ASI-Sij-4, ASI-Sha, ASI-Kondara, ASI-Kly-1, ASI-Kly-2, ASI-Kly-3, ASI-Kly-4, ASI-Kly-6, ASI-K-oz-1, ASI-K-oz-2, ASI-K-oz-3, ASI-Leb-1, ASI-Leb-2, ASI-Leb-3, ASI-Leb-4, ASI-Nov-3, ASI-Bas-1, ASI-Bas-2, ASI-Bas-3, ASI-Cha-1, ASI-Cha-2, ASI-Sev, ASI-Pan, ASI-La-0, ASI-Bij, ASI-Per-1, TRA-Chi-0, TRA-Es-0, TRA-Est-0, TRA-Rsch-0, TRA-Wil-1, EUR-Kent; **ASI-Nov-1** = ASI-Nov-1, ASI-Nos, ASI-Mas, ASI-Rak-1, ASI-Rak-2, ASI-Rak-3; **NA-Gre-0** = NA-Gre-0, NA-KEN, NA-Col, NA-Lim, EUR-Gie, EUR-Han, J-Tsu-0; **TRA-Kn-0** = TRA-Kn-0, MED-Ita-0; **EUR-Bad** = EUR-Bad, TRA-Rub-1, NA-Pog-0, NA-Sea; **TRA-Br-0** = TRA-Br-0, TRA-Lip-0, EUR-In-0, EUR-Ag-0, EUR-Bs-2, MED-Can-0, MED-Sie; **MED-Flo** = MED-Flo, EUR-Oph, EUR-Par; **EUR-Gat** = EUR-Gat, EUR-Wt; **MED-Bla-1** = MED-Bla-1, EUR-Duel; **EUR-Di-1** = EUR-Di-1, EUR-Lis; **TRA-Kga** = TRA-Kga, TRA-Ryb; **ASI-Kaz-1** = ASI-Kaz-1, EUR-Bur-0.

(*Sah-0*) was associated with this group mainly on the basis of fragment 1. (2) The „European group“ comprises accessions from all across Europe and includes accessions from Libya and one from Kazakhstan. (3) The „Mediterranean group“ is the most heterogeneous group mainly encompassing accessions from the Mediterranean region, the Republic of Georgia, some Siberian accessions (only in fragment 1) and two American accessions. Sequence variation in fragment 1 resulted in a more structured tree than that of fragment 2 which points to less recombination among the „European“ accessions in the gene than in the non-coding region since in both fragments the „Asian“ accessions are rather uniform.

The trees show one and twelve homoplasies, Fig. 3A, B respectively. This high number of homoplasies may be due to the calculation with the parsimony algorithm used to calculate the trees in the presence of recombination. Due to the high number of detectable recombinations in fragment 2, this tree reconstruction results in several possibilities of minimal „mutation“ steps, indicated by the net-like structure (Fig. 3). The accession *Dar* contains so many conflicting signals and a large amount of singletons in fragment 2, that no integration with 95% probability is possible, not even in a net structure. SNPs with roughly equal frequencies („intermediate polymorphisms“) of the alternative nucleotides might indicate some older polymorphisms on the genealogy or some form of selection (Konnert and Bergmann, 1995; Akashi, 1999). This is supported by their association in linked clusters. There are shared intermediate SNPs among the „European“ and the „Mediterranean“ groups and *Dar*, while the „European“, the „Mediterranean“ accessions, and *Dar* differ in their low-frequency alleles.

The geographic structure in fragment 1 (Fig. 3A) is mainly based on a 360 bp-region, comprising parts of the first intron and second exon of the gene (positions P258881 – P2589244). Interestingly, the „intermediate polymorphisms“ in this region are synonymous mutations (data not shown). In the exons of fragment 1 nine amino acid changes among alleles were revealed (data not shown). All of these substitutions occur with a very low frequency in the tested set (one to four accessions per substitution). Six of the nine amino acid changes are located in the last exon.

Since the haplotypes of the various accessions differ by mutation as well as by recombination, we applied split decomposition (Bandelt and Dress, 1992) for an analysis and visualization of reticulate patterns resulting from recombination. The splits tree network (Fig. 4) unfolds between the rather uniform „Asian“ and the more variable „European“ group of accessions. It has a poor fit of 33.7%. A bootstrap support could not be estimated due to the small number (36) of informative characters.

Accessions with conflicting signals form several independent rectangles of which each suggests a series of sequential recombinations, even if the sequential order of these cannot be resolved. These are:

i) Asia – *Nos* – *Rak1-3*, *Mas1*, *Nov1* – X1 – Asia

ii) Asia – *Nov3* – X2 – *Kent* –X3 – Europe

iii) Asia – *Sah-0*, *Chi-1*, *Gre-0*, *KEN* – *Gie*, *Han*, *Tsu*, *Ll-2*, *Mal*, *Dar*, *Alc*, *Cvi* – Europe

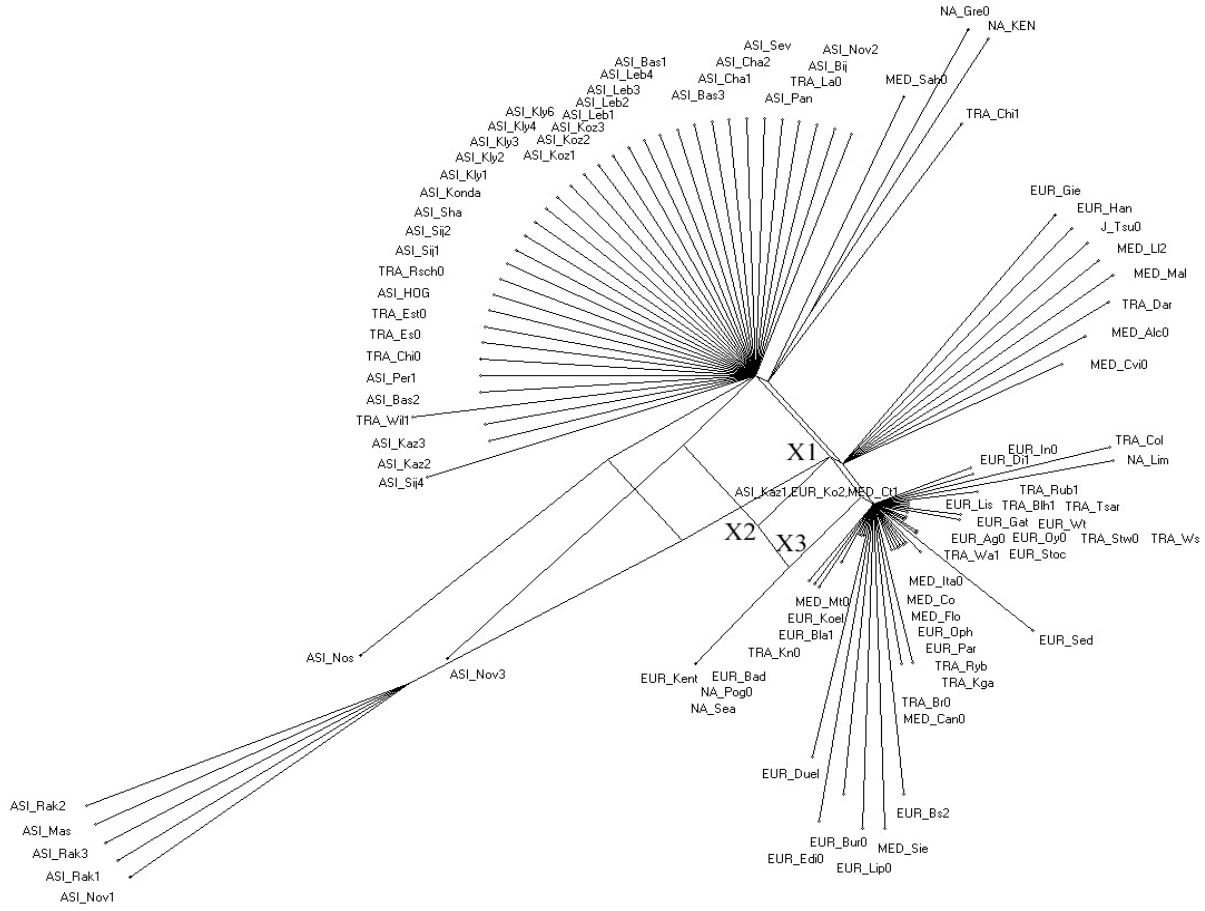


Fig. 4. Splits tree estimated with substitutions with a frequency of more than 5% for the less presented allele. Fit of the tree is 33,7%. X1, X2 and X3 are cross points between rectangles. For used prefix see Fig. 2.

The *Nos* haplotype apparently consists mostly of sequence of the „Asian“ type and a segment which is typical for a small uniform „Siberian“ group of accessions (*Rak1*, *Rak2*, *Rak3*, *Mas1*, *Nov1*). This suggests that *Nos* is the result of a hybridisation event between the two groups. *Nov3* also consists mainly of the „Asian“ type and shares a genomic region (smaller than the one in *Nos*) with the small „Siberian“ group (Fig. 4). The *Kent* haplotype is nearly half „European“ and half „Asian“ sequence, with one substitution shared with the „Siberian“ group which might be a parallel evolution event. The two subgroups iii contain substitution patterns from both „Asia“ and „Europe“: *Chi-1*, *Sah-0*, *Gre-0* and *KEN* are recombinants between the two sequenced fragments, whereas *Gre-0* and *KEN* contain additional SNPs only occurring in the „European“ sequence type. A shared substitution pattern in the 360 bp-region (indicated in Fig. 7) showing a strong geographical signal explains the intermediate position of the second mainly „Mediterranean“ subgroup. The „Siberian“ subgroup (*Rak1*, *Rak2*, *Rak3*, *Mas1*, *Nov1*) also contains this crossover. Although the accessions *Ll-2* and *Sah-0* come from the Iberian Peninsula they are closely

related to the „Asian“ group. This corresponds to the position of these two accessions in the AFLP analysis by Sharbel et al. (2000).

The geographic information in the splits tree diagram is illustrated in Fig. 5, where the percentage of diagnostic consensus „Asian“ and „European“ SNP markers (indicated in bold numbers in Fig. 2) is indicated for each accession by the black and the grey sectors.

The overall Mantel correlation of the genetic and geographical distances was 0,46 (*P*=0,001), indicating that genetic distance increased with increasing geographical distance among accessions.

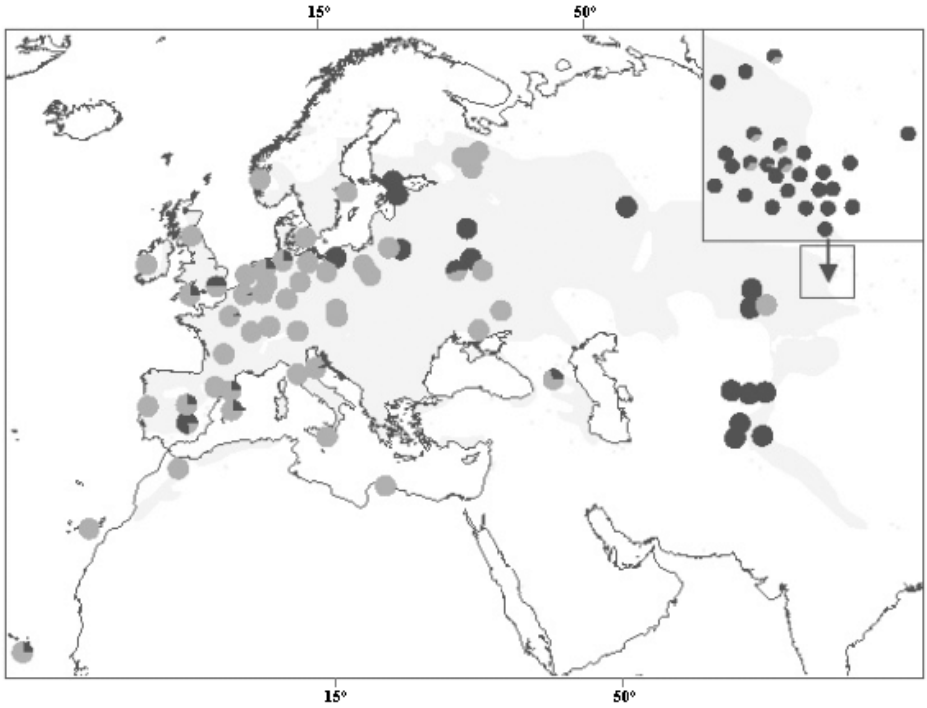


Fig. 5. Map of the investigated accessions. Light grey shaded background area represents the natural range of *A. thaliana*. Grey parts represent „European“-specific portion and black parts represent „Asian“-specific percentage of the consensus sequence marked in Fig. 2.

Mantel tests with distance classes showed a decrease from significantly positive R_M-values in distance classes 1 and 2 (accessions up to 470 km apart) to significantly negative values in distance classes 20 to 25 (4465 – 5875 km) and 29 to 30 (6580 – 7050 km). Significantly negative values indicate very distinct genetic pattern in the defined geographic region. In distance class 26 to 28 (5875 – 6580 km) R_M-values were also negative but not significant (Fig. 6). Distance classes 7 and 8 (1645 – 1880 km) showed again significant positive values (R_M=0,08 and 0,15, both with *P*=0,001). These classes

comprise mainly accessions from Siberia, Uzbekistan, Tajikistan and Middle Russia which are characterized by similar haplotypes occurring over a wide North-South range.

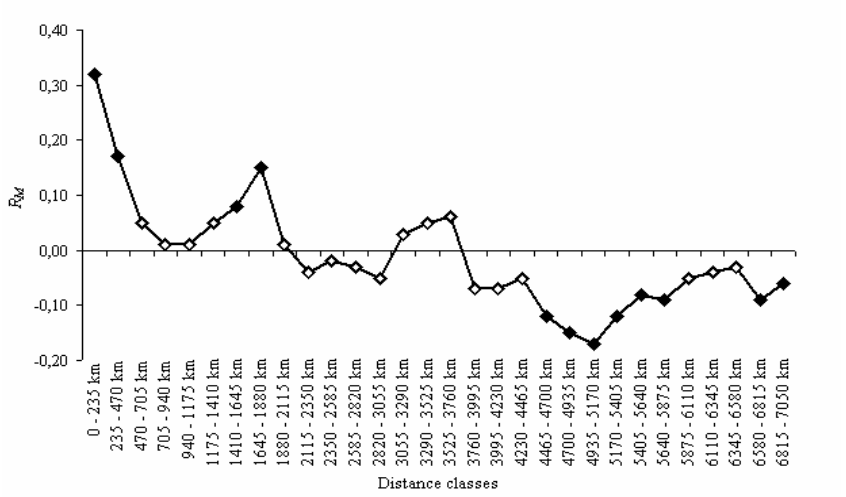


Fig. 6. Correlogramm of Mantel R_M-values per distance class of 90 *A. thaliana* accessions (for details refer to the text). Filled diamonds indicate R_M-values significantly different from zero at P=0,05 after regular Bonferroni correction.

Association of the SNP data to temperature and precipitation

The distribution of *A. thaliana* spans a wide range of climatic regions. Western and Middle Europe as well as Japan and the American coasts are characterised by an oceanic climate, i.e. low annual temperature amplitude and moist conditions. Eastern Europe, Middle Asia and Siberia have continental climate (large temperature amplitude with partly pronounced drought period). Moderate temperatures throughout the year and precipitation occurring mainly in winter characterize the Mediterranean region.

Climate data from the sites of origin of the various accessions were not normally distributed (Komolgorov-Smirnov test; data not shown). Therefore, the non-parametric Mann-Whitney-U-Test was applied to test whether a SNP distribution is associated with temperature and/or precipitation at the sites of origin. The two climate indicators used in this test were separated by the median of temperature and precipitation for each SNP, respectively. Thus, false correlations due to less frequent or unique alleles could be eliminated. The two groups were tested for significant differences with the Mann-Whitney-U-Test.

Significance levels for April conditions are indicated by asterisks in Fig. 7. Significant associations with temperature and precipitation data at a level of P=0,001 were only

obtained with SNPs that have an intermediate frequency (0,28 to 0,48 for the less frequent alleles). This excludes the possibility of spurious associations due to singletons or less frequent substitutions.

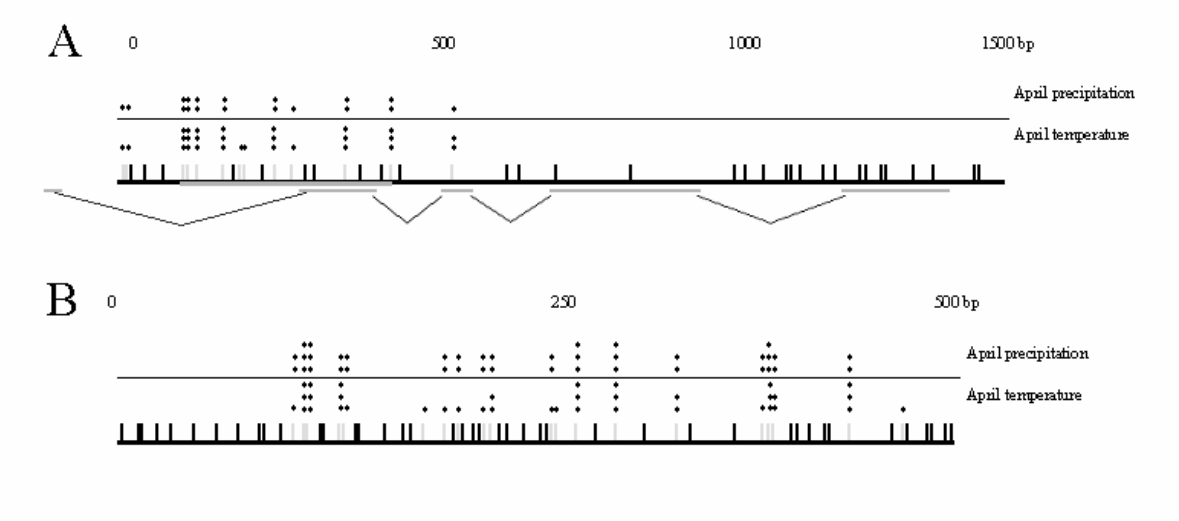


Fig. 7. Exon-Intron spreading of **A**: gene At2g06530 and **B**: MASC06057-region. SNPs on chromosome 2 are indicated by bars. Asterisks show the level of significance of the Mann-Whitney-U-Test (* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$) with precipitation and temperature. The grey region in **A** indicates a shared crossover in the 360 bp-region, mentioned in the text.

Similar highly significant associations as shown in Fig. 7 were also obtained with temperature in the months from September to March and for precipitation in the months September, January and February (data not shown). Temperature and precipitation data for the other months showed either weak or no significance.

Discussion

Several sequencing studies detected biallelic variation between nucleotides and a surplus of rare alleles in *A. thaliana* (e.g. Hanfstingl et al., 1994, and Innan et al., 1996: *Adh*; Kawabe et al., 1997: *ChiA*; Caicedo et al., 1999: *Rps2*; Kawabe and Miyshita, 1999: *ChiB*; Stahl et al., 1999: *Rpm1*; Purugganan and Suddith, 1998: *CAL*, 1999: *AP3*, *PI*; Aguadé, 2001: *FAH1*, *F3H*; Hauser et al., 2001: *GL1*). The frequency distribution of our 67 genome-wide SNP set differs from these studies. The lack of a surplus of rare alleles is due to a conscious bias in the selection of the database. 56 of the 67 investigated substitutions were detected between the accessions *Col* and *C24* and were selected for SNPs that showed intermediate frequencies in a set of up to 8 accessions (MASC database; Schmid et al., 2003).

Previous investigations of the structuring of genetic markers in *A. thaliana* have not detected clear geographic patterns (e.g. King et al., 1993; Ullrich et al., 1997; and Bergelson et al., 1998 with RFLPs; Breyne et al., 1999; Miyashita et al., 1999; and Erschadi et al., 2000 with AFLPs; and Innan et al., 1997; and Loridon et al., 1998 with microsatellites and sequence data from single genes (overview in Hoffmann et al., 2003b). Sharbel et al. (2000) using AFLPs from all across the genome have shown a general increase of genetic distance with geographic distance and a noticeable difference between accessions from the Asiatic part of the range and the rest. This study is remarkable because a fairly large set of range-wide accessions was investigated with randomly distributed genome-wide genetic markers.

Our approach differs from the earlier studies since we have specifically investigated two single-base polymorphisms from a genome-wide sample of 67 that individually showed a distinct geographical distribution pattern, i.e. those that were responsible for a weak geographical signal in the overall dataset (data not shown). Of the three polymorphisms found, the geographical correlation of one was restricted to this site with no correlated variation in linked polymorphisms in the 500 bp surrounding it. In the two remaining cases, a large number of nucleotide polymorphisms near the selected CAPSs was eventually found in a sample of 98 accessions, 43 surrounding MASC05800 and 61 surrounding MASC06057. This dense pattern of markers has provided detailed insight in the recent evolutionary history of the two regions on chromosome 2. Two observations are especially significant. First, the geographical pattern found in the two selected CAPSs is paralleled by that of most polymorphisms with allele frequencies of about half and half in the surrounding sequence. In these cases, one of the alleles is unique or predominant in „Asian“ accessions, the other one in „European“ accessions. Second, scattered among these, SNPs with a more uneven distribution are found, where the rare allele is usually associated with the „European“ alleles of the more frequent ones (Fig. 2). This amount of rare alleles is similar to former sequence studies (e.g. Aguadé, 2001 and references therein). The SNPs with intermediate frequency and an Asian vs. European distribution probably represent two „ancient“ haplotypes that have formed in geographical isolation. There is a striking difference in the variation within each of the two haplotype groups. One invariant version of the „Asian“ haplotype is present in 29 of the accessions from Uzbekistan, Tajikistan, Russia and Poland, while each of the „European“ accessions from France, Germany, the Netherlands, Belgium, and Sweden has some additional single-base mutations or InDels so that no more than 3 of the „European“ accessions have completely

identical sequences for the examined chromosomal region. This suggests that the „Asian" haplotype has gone through a recent population bottleneck, while the „European" haplotypes have accrued many independent mutations.

While most of the sequences can be assigned to one of the two basic haplotype groups, split decomposition (Bandelt and Dress, 1992) revealed recombinants where one of the two sequenced fragments represented the „Asian", the other the „European" haplotype (e.g. *Sah-0*, *Chi-1* and, reciprocally, *Kent* in Fig. 4). Also two other groups were placed by split decomposition in the network between „Asian" and „European" accessions: *Cvi-0*, *Dar*, *Ll-2*, *Mal*, *Alc-0*, *Gie*, *Han*, *Tsu-0* and *Nov-1*, *Rak-1*, *Rak-2*, *Rak-3*, *Mas-1*. Both groups have a 360 bp region of linked substitutions in the first fragment (positions P258881 – P2589244; Fig. 2). This involves the „intermediate polymorphisms", i.e. those with about equal frequencies in the sample, while rare polymorphisms form subgroups within linked blocks.

This interpretation agrees with the results by Hanfstingl et al. (1994) who found recombination about once every 350 bp in *A. thaliana*. Strikingly, R_m, a four gamete test approach (Hudson and Kaplan, 1985), RecPars (Hein, 1990, 1993) and R_h (Myers and Griffiths, 2003) did not confirm recombination in between fragment 1 and suggested recombination in between fragment 2.

It is known, that R_m misses most of the recombination events in the sample history (Hudson and Kaplan, 1985), particularly where the mutation data are limited or the rate of recombination is high. This might be the case in our data. R_h (Myers and Griffith, 2003) has weaknesses because it uses the same assumptions as R_m. Furthermore it is not always possible to construct the needed data history. On the other hand, split decomposition (Bandelt and Dress, 1992) detected conflicting signals, probably due to recombination.

Statistical approaches for the detection of recombination are preferentially designed to detect rare recombination events (Maynard Smith, 1999). Posada (2002) compared 14 methods for detecting recombination in sequence data, and suggested that „Conclusions here depend on a good knowledge of which data sets are recombinant and which ones not". Posada (2002) and Posada et al. (2002) suggest to employ various approaches side by side.

Undoubtedly, the haplotype trees in Fig. 3 represent recombination between and within the „Asian" and „European" haplotype groups in addition to mutation within each group. The large number of hypothetical haplotypes that has to be postulated to connect the „Asian" and the „European" group in Fig. 3A is more likely to be an artifact of recombination among blocks of SNPs after the two groups have rejoined than to indicate the mutational

history leading to the two groups. The even more complex network in Fig. 3B (noteworthy, the non-coding sequence) also indicates recombination, which was estimated with the recombination parameters.

A similar pattern of recombination between two basic haplotypes in *A. thaliana* has been observed with sequencing studies by Kawabe et al. (1997) for *ChiA* on chromosome 5, by Kuittinen and Aguadé (2000) for *CHI* on chromosome 3, and by Aguadé (2001) for *FAH1* and *F3H* on chromosome 4 and 3. These authors suggested that the two recombining haplotypes have recently rejoined and recombined after a period of isolation. Similarly, Sharbel et al. (2000) have suggested isolated refugia during the ice ages as the basis for the existing geographic pattern within the species. According to Sharbel et al. (2000), the location of the refugia was in Middle Asia and the Western Mediterranean area. Our data allow a similar conclusion for the chromosomal regions examined by us. Price et al. (1994), Alonso-Blanco and Koornneef (2000), and Aguadé (2001) mentioned Asia as a centre of diversity for *A. thaliana*. Hewitt (1999) suggested these regions as refugias for other plant species. If the pattern found by us for chromosome 2 can be extrapolated to the entire genome, Middle Asia at present is not a centre of diversity for the species but shows a severe bottleneck effect of isolation, probably due to the ice age.

Localized associations between markers and characters tend to suggest the effect of selection. The association between climatic factors and the geographical distribution of the polymorphisms studied by us could point in the same direction. However, since there is an East/West climate gradient parallel to the distribution gradient, the correlation with climatic factors could also be accidental, and the correlation with climate is not stronger than expected on the basis of the general East-West distribution. A much closer geographic sampling in an area with large climatic differences within region of recombination would be interesting. At present, there is no evidence for an adaptive component to the geographic distribution of the haplotypes in chromosome 2 (Hoffmann et al., in review).

Our interpretation raises the question why only 3 of 67 CAPSs from all over the genome showed a clear geographical pattern. There are two explanations, both of which probably contribute to the observation. First, almost all of the initial CAPSs tested by us were identified between „European“ accessions (*Col* and L*er* (Jander et al., 2002) or *Col* and *C24* (Schmid et al., 2003)). Therefore, both „alleles“ of each SNP occur in European accessions. This discriminates against SNPs diagnostic for the European/Asian differences. This and the small number of accessions from the Asian part of the range might have been

the reason why no geographic correlation was detected in previous sequencing studies (Aguadé, 2001 and references therein).

The ascertainment bias is removed when SNPs are detected by comparatively sequencing a sufficiently long region in a representative sample of accessions in order to obtain a statistically reliable set of closely linked SNPs. Substituting clusters of SNPs allows the identification (by non-disjunct markers) of stretches of the genome that have been transmitted as blocks and to compare blocks separated by recombination. A genome-wide survey of individual SNPs is likely to include recombination events among all of the markers.

Second, this fine-scale analysis of stretches of sequences also has shown that there are many independent mutations with low frequencies in between the ones that are diagnostic for the two haplotype groups, especially in the „European" haplotypes (Fig. 2). Only 13 of the 106 SNPs detected in this region had an intermediate frequency in our sample of accessions and were diagnostic for the two haplotype groups. The other (intrahaplotype) single SNPs or short stretches of non-disjunct SNP pattern show no significant correlation with either geography or climate. This indicates that *A. thaliana* does not extend its geographical range by a slow and steady diffusion from the two centers of distribution. In that case, the (post-glacial) mutational history would parallel the geographical range extension. In contrast, even the distribution of the two basic haplotypes and their recombination products is very coarse-grained, and there are nearly pure „Asian" sequences for one (*Kent* in fragment 2) or both (*Sah-0* in fragment 1 and parts of fragment 2 and *Le-0* from Leiden, Netherlands (unpublished data)) in European locations, as there are „European" sequences in Asian accessions (*Kaz-1*). Part of this is undoubtedly (recent) anthropogenic influence, but fine-scale analysis of more genomic regions may reveal a smoother overall pattern of increasing genome-wide recombination events during range expansion.

Even in the two regions of chromosome 2, only one in 8 randomly chosen SNPs shows a statistically significant geographical distribution pattern. This can also be demonstrated when the diagnostic SNPs from chromosome 2 are used as guides to classify accessions as „Asian" or „European". We have made a survey of another 115 SNPs with a genome-wide distribution in 335 accessions (Schmid et al., in review) to look for further sites with the same geographic signal. The distribution of 5 of the 115 new individual SNPs was significantly associated with the Asian/European separation at probability values below 0,02%. At this significance level one expects 2-3 false positives, thus the set of 5 SNPs

probably contains geographically diagnostic ones. These 5 SNPs were situated on chromosomes 1, 2, 3 and 5. We predict that the comparative sequencing of other chromosomal regions in a comparable range-wide sample of about 100 accessions will reveal further and evidence and additional details of the post-glacial meeting and recombination of two subpopulations of *A. thaliana*. Unfortunately, we interpreted the first geographic signals as signs of local adaptive polymorphisms and excluded all other regions from our study. However, this information will become available from ongoing sequencing projects in several laboratories. Our results also show that it will require a much higher sampling density than our Europe-wide set of accessions to obtain unequivocal evidence of selection and local adaptation during the range extension of *A. thaliana*. Combining natural genotype/environment correlations with unbiased marker-assisted studies of genetic segregation and selection in laboratory hybrids will be the method of choice (Mitchell-Olds, 2001).

Acknowledgements

We thank Ralf Horres and Birgit Gemeinholzer for helpful discussions on the manuscript and C. Koch and P. Oswald for technical assistance. Special thanks go to N. Ermakov, who helped collecting *A. thaliana* in Siberia. I. Stehlik collected the plant material from the Republic of Georgia. This work was supported by grants from the Deutsche Forschungsgemeinschaft to K.B.

Genome size variation among accessions of *Arabidopsis thaliana*

Heike Schmuths, Armin Meister, Ralf Horres, Konrad Bachmann. 2004. Annals of Botany, 93: 1-5.

Abstract

Background and Aims Estimates of the amount of nuclear DNA of *Arabidopsis thaliana*, known to be among the lowest within angiosperms, vary considerably. This study aimed to determine genome size of a range of accessions from throughout the entire Eurasian range of the species.

Methods Twenty accessions from all over Europe and one from Japan were examined using flow cytometry.

Key Results Significant differences in mean C-values were detected over a 1,1-fold range. Mean haploid (1C) genome size was 0,215 pg (211 Mbp) for all analysed accessions. Two accessions were tetraploid.

Conclusions A closer investigation of the DNA fractions involved in intraspecific genome size differences in this experimentally accessible species may provide information on the factors involved in stability and evolution of genome sizes.

Keywords: *Arabidopsis thaliana*, genome size, flow cytometry, tetraploid, geographic correlation, C-value, intraspecific variation

Introduction

Since *Arabidopsis thaliana* is believed to possess one of the smallest nuclear genomes among higher plants, the exact determination of the genome size has received much attention. Different techniques were used to estimate the genome size, among them reassociation kinetics (Leutwiler et al., 1984), quantitative gel blot hybridization (Francis et al., 1990), Feulgen photometry (Bennett and Smith, 1991; Krisai and Greilhuber, 1997), and flow cytometry (Galbraith et al., 1991; Arumuganathan and Earle, 1991; Marie and Brown, 1993; Dolezel et al., 1998; Barow and Meister, 2002; Bennett et al., 2003).

Earlier studies estimated the haploid genome size (1C-value) at between 0,051 pg (Francis et al., 1990, from Bennett et al., 2003) and 0,215 pg (Dolezel et al., 1998, laboratory 1). Sequencing studies found about 125 Mbp (*Arabidopsis* Genome Initiative (AGI 2000), which corresponds to 0,128 pg. Determinations based on reassociation kinetics and quantitative gel blot hybridization tend to suggest smaller genome sizes (0,051 pg to 0,082 pg; Francis et al., 1990; Leutwiler et al., 1984) than Feulgen photometry or flow cytometry (0,085 pg to 0,215 pg; Galbraith et al., 1991; Dolezel et al., 1998). Furthermore, older studies tend to estimate lower values than more recent ones.

Considering the small 1C-value, the role of endopolyploidy and its relation to cell size or life cycle has been investigated by several authors (e.g. Melaragno et al., 1993; Kondorosi et al., 2000; Beemster et al., 2002; Barow and Meister, 2003). It was found that *A. thaliana* has a high level of endopolyploidy in nearly all parts of the plant (cotyledon, root, lower leaf stalk, lower leaf, upper stem, upper leaf, flower stalk, sepal, petal (Galbraith et al., 1991; Barow and Meister, 2003).

Most of the investigations of genome size in *A. thaliana* made use of the accession *Columbia*. There are indications that intraspecific variation exists between accessions, for example between *Columbia* and the Cape Verde Islands ecotype *Cvi-0* (Meister, unpubl. res.).

We used flow cytometry to measure genome size of 21 accessions from throughout the entire Eurasian range, including newly collected material from Middle Asia, and one accession from Japan. Since intraspecific genome size differences are subject to much critical discussion (e.g. Greilhuber, 1998), we tried to check our results by measuring each accession ten times and repeating this procedure for accessions with especially large and small genome sizes.

Materials and Methods

Plant material

21 accessions were included in this study, of which 18 were obtained from the Nottingham *Arabidopsis* Stock Centre and 3 were collected in the wild (Table 1). Plants were grown at 24°C (16h day, 8h night) and pots were not allowed to dry out. Leaves from three to five individuals per accession were used for the determinations. Ten determinations were made per accession. This reduces the assumed standard error ratio of 5% by a factor of $1/\sqrt{n}$, i.e. to 1,6% for 10 determinations and 1,1% for 20 determinations. Three accessions with small (*Kn-0*, *Köl*) and large (*Mas*) genome size in the first series of determinations were measured a second time with 10 repeats.

Table 1. List of investigated accessions.

accession	location
Ag-0[1]	Argentat, France
Alc-0[1]	Alcalá de Henares/Madrid, Spain
Chi-0[1]	Chisdra, Russia
Col[1]	Columbia, Poland
Es-0[1]	Espoo, Finland
Kent[1]	Kent, UK
Kly-1[2]	Kolyvan, Russia
Kn-0[1]	Kaunas, Lithuania
Köl[1]	Köln, Germany
La-0[1]	Landsberg/Warthe, Poland
Ll-2[1]	Llagostera, Spain
Mas[2]	Masljacha, Russia
Oph[1]	Ophain, Belgium
Oy-0[1]	Oystese, Norway
Sah-0[1]	Sierra Alhambra, Spain
Sha[1]	Shakdara, Tajikistan
Sij-2[2]	Sijak, Uzbekistan
Stoc[1]	Stockholm, Sweden
Tsu-0[1]	Tsu, Japan
Wa-1[1]	Warschau, Poland
Ws[1]	Wassilewskija/Dnepr, Russia

[1] from Notthingham *Arabidopsis* Stock Centre (NASC)
[2] collected in the wild

Preparation and analysis

We followed the protocol of Barow and Meister (2002) for preparation of the nuclear suspensions and the analysis procedure of DNA contents. A FACStar ^{PLUS} flow cytometer

(Becton Dickinson, San José, CA) equipped with two argon lasers INNOVA 90-5 (Coherent, Palo Alto, CA) was used and the data were analysed with the program CellQuest (Becton Dickinson, San José, CA). Nuclear DNA content was estimated by the fluorescence of the nuclei of the samples stained with propidium iodide relative to the internal standard *Raphanus sativus* (2C=1,38 pg, Dolezel et al., 1998). Usually 10.000 nuclei were measured.

Statistical analysis

Normal distribution and homogeneity of variances of the average genome size per accession were tested with the Kolmogorov-Smirnov-test and the Levene-test, respectively. As the data have no homogeneity of variances, no variance analysis or parametric tests were used for further computations. Instead we used the non-parametric Kruskall-Wallis-test and subsequently calculated comparisons of means with the Dunn test, which can handle the unequal sample size of ten or twenty determinations per accession. Genome size differences among di- and tetraploids were computed with the Mann-Whitney-U-test.

Correlations between mean genome size, geographic coordinates, and seed size parameters were estimated with the Spearman-rank-correlation. Calculations were made using the programs SPSS 10.0 and SigmaStat (Erkrath, Germany). Geographical coordinates were kindly provided by M. H. Hoffmann (University of Halle, Germany).

Results

Ploidy levels

Of the 21 investigated accessions 19 were diploid and two tetraploid. The tetraploid accessions *Stoc* and *Wa-1* were excluded from the correlations of genome size, geographic coordinates, and seed parameters so as to determine the effect of genome size independently from that of ploidy.

Differences within accessions

Replicate measurements within accessions were highly repeatable. The standard deviation of genome size in the diploid accessions ranges between 0,004 pg for the accession *Ws* and

0,014 pg for the accession *Oy-0* (=1,04 to 3,26% of the genome sizes, respectively). Also, the ten replicate measurements of *Kn-0*, *Köl* and *Mas* varied in this range. An example for a typical flow cytometry histogram is shown in Fig. 1, which also illustrates the high rate of endopolyploidy.

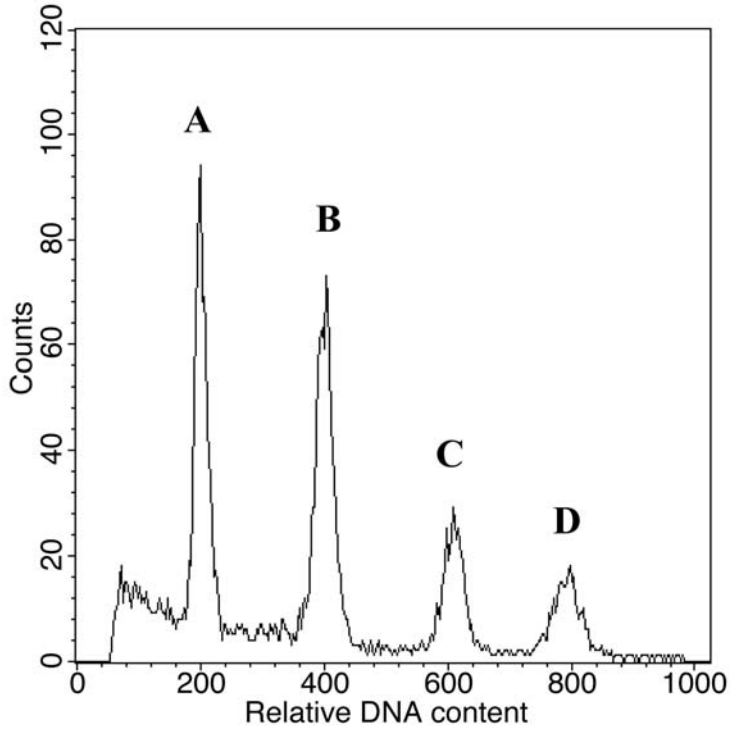


Fig. 1. Flow cytometry histogram of leaf nuclei from the accession *Mas*. Peak A corresponds to 2C, peak B to the 4C-value and peak D to the 8C-value of *A. thaliana*, peak C corresponds to the internal standard *R. sativus*.

Differences among diploid accessions

There is a 1,1-fold difference in genome size between the accessions, with *Col* as the smallest and *Kly-1* as the largest accession. The mean diploid genome size is 2C=0,431 pg. A general overview of genome sizes in all investigated accessions is listed in Table 2.

Genome sizes of the 19 diploid accessions are normally distributed (p=0,956) but homogeneity of variances is rejected (p=0,016). We therefore used the non-parametric Kruskall-Wallis-test with pairwise comparisons of mean genome size by the Dunn test (Table 2). Since this test uses ranks instead of absolute values, differences between accessions are less pronounced than in a parametric test. Despite of this weakness of the non-parametric test, significances (p<0,05) were found between all measurements of the five largest diploids (*Kly-1*, *Mas, Sah-0, Alc-0, Sha*) and the three accessions with the smallest genome sizes (*Col*, *Köl*, *Kn-0*), (Table 2). Accessions with mean genome size

differences smaller than between *Chi-0* and *Col* are not significantly distinct from each other. The percentage difference of mean genome size between *Chi-0* and *Col* is 5,35%. Thus, significant differences between accessions are larger than the standard deviation within accessions.

Table 2. Mean, standard deviation, minimum and maximum of the 21 investigated accessions. Differences were calculated for the 19 diploid accessions with the Dunn test.

accession	number of determinations	mean (pg)	minimum (pg)	maximum (pg)	significance code[1]
Col	10	0,412±0,007	0,403	0,424	a
Köl	20	0,415±0,012	0,390	0,437	a b
Kn-0	20	0,421±0,006	0,403	0,429	a b c
La-0	10	0,425±0007	0,417	0,435	a b c d
Es-0	10	0,426±0,009	0,411	0,439	a b c d e
Sij-2	10	0,429±0,005	0,419	0,438	a b c d e
Kent	10	0,429±0,011	0,412	0,444	a b c d e
Ll-2	10	0,429±0,006	0,421	0,439	a b c d e
Tsu	10	0,430±0,007	0,421	0,443	a b c d e
Ag-0	10	0,432±0,008	0,411	0,444	a b c d e f
Oy-0	10	0,436±0,014	0,419	0,462	a b c d e f
Oph	10	0,434±0,010	0,419	0,445	a b c d e f
Chi-0	10	0,434±0,008	0,416	0,446	b c d e f
Ws	10	0,435±0,004	0,427	0,441	c d e f
Sha	10	0,439±0,007	0,430	0,450	d e f
Alc-0	10	0,440±0,008	0,427	0,449	d e f
Sah-0	10	0,441±0,009	0,427	0,456	d e f
Kly-1	10	0,450±0,014	0,432	0,472	e f
Mas	20	0,448±0,007	0,431	0,458	f
Stoc	10	0,889±0,031	0,856	0,942	
Wa-1	10	0,892±0,045	0,848	0,983	

[1]Same letters indicate a lack of significant differences among accessions (p>0,05). Only accessions that do not have the same letter are significantly different.

Correlation with coordinates and seed parameters

We used the non-parametric Spearman-test to test for correlation of genome size and latitude/longitude. Spearman´s rho resulted in a significant positive correlation between genome size and longitude (rho=0,185, p=0,006, n=220). This suggests a larger mean genome size of the Eastern accessions compared to the Western accessions within the Eurasian distribution range.

Correlation between genome size and latitude resulted in a negative correlation for the 19 diploids (rho=-0,138, p=0,040, n=220). Thus, the genome size decreases slightly but

significantly with increasing longitude (diploid accessions have a larger genome in the South than in the North).

Seed width and length were measured for 10 to 50 seeds per accession. Seed length varied between 0,359 and 0,717 mm with a mean of 0,537 mm, seed width varied between 0,250 and 0,543 mm with a mean of 0,334 mm. Spearman´s rho was used to test for correlation and resulted in a small but significant negative correlation of genome size and seed width of rho=-0,095, p=0,012, n=700, and genome size and seed length of rho=-0,099, p<0,001, n=700.

A Spearman-rank-correlation between genome size and precipitation or temperature in the vegetation period from October to June revealed no significant correlations (data not shown).

Tetraploid accessions

The two tetraploid accessions *Stoc* and *Wa-1* had means of 0,889 pg and 0,892 pg, with standard deviations of 0,031 and 0,045 (=3,44 to 5,08% of the genome size), respectively (Table 2). The 1C DNA amounts of 0,445 pg and 0,446 pg are comparable to those of the 2C-values of the diploid accessions with large genome sizes.

To our surprise the length and width of the tetraploid seeds ranged around the upper level of the diploid seeds rather than being noticeably larger, but a Mann-Whitney-U-test detected significant differences (p<0,001) between diploid and tetraploid seeds. Seed length varied between 0,542 mm to 0,696 mm in *Wa-1* and 0,630 mm to 0,880 mm in *Stoc*, with mean seed lengths of 0,620 mm and 0,708 mm, respectively. Seed width ranged between 0,315 mm to 0,413 mm in *Wa-1* and 0,326 mm to 0,500 mm in *Stoc*, with mean seed widths of 0,376 mm and 0,408 mm, respectively. Thus, single seeds from the diploid accessions *Oph* (0,511 mm) and *Sij-2* (0,543 mm) had larger seed widths than the tetraploid seeds from *Wa-1* and *Stoc*.

Discussion

Our estimated 1C-values using *R. sativus* (2C=1,38 pg) as an internal standard agree with former published values (e.g. Galbraith et al., 1991; Dolezel et al., 1998) but are higher than other estimated values (e.g. the *Arabidopsis* Genome Initiative, 2000 or Bennett et al., 2003). Differences between these studies might be due to different types of flow

cytometers (lamp/laser) (Dolezel et al., 1998), different size standards (Dolezel et al., 1998; Barow and Meister, 2002) or, in the case of sequencing, to a conservative, low estimate of the contribution of centromer regions of the chromosomes of *A. thaliana* (Bennett et al., 2003).

Bennett et al. (2003) compared the incompletely sequenced *A. thaliana* genome against the completely sequenced *Caenorhabditis elegans* genome (1C= approx. 0,1 pg) and determined *A. thaliana* with1C=approx. 0,16 pg. Vilhar et al. (2001), using image cytometry and *Pisum sativum* as standard (2C=8.84 pg), estimated similar C-values for *A. thaliana* (2C=0,32 and 0,33 pg). The value we used for *R. sativus* (Dolezel et al., 1998) was derived by a cascade down from *Allium cepa* (2C=33.5 pg), in which different laboratories used different flow cytometers and had divergent estimates for *A. thaliana*.

Nevertheless, since we always referred the same internal standard, *R. sativus* with 1,38 pg (Dolezel et al., 1998, laboratory 1, which was also used for these measurements), the intraspecific variation differences found in this study for accessions of *A. thaliana* are true.

Measuring accessions from all across the Eurasian range of the species, the Dunn test revealed significant differences. Intraspecific variations of genome size below the species level are supposed to be rare (reviewed by Greilhuber, 1998). Reported differences can be explained by chromosome polymorphisms, spontaneous aberrations but also by technical shortcomings (Greilhuber, 1998; Galbraith et al., 1991). Possible measuring errors in our investigations are limited by using the same flow cytometer and at least ten repetitions, which decreases an estimated error from 5% per measure to 1,6% (n=10) or 1,1% (n=20) per accession mean. Considering the genetic diversity of the accessions found by Schmuths et al. 2004b and the magnitude of the variation in spontaneous aberrations or chromosome polymorphisms found in various angiosperm species (reviewed by Greilhuber, 1998) intraspecific variations in genome size can be expected.

The genome size of the tetraploids ranges between 0,445 pg and 0,446 pg, which is comparable twice that of the diploid accessions with large genome size. Polyploid angiosperms can reduce or increase their genome sizes relative to the ancestral diploids (reviewed by Bennetzen, 2002; Soltis and Soltis, 1995; and Wendel, 2000). Since we do not know the precise diploid ancestor(s) of the tetraploid accessions in *A. thaliana*, we do not know if the relatively high tetraploid genome size deviates from the sum of the parental values. Generally, cell and organ size in plants is correlated with DNA content (p.e. Strasburger et al., 1991). The significant size differences between seeds from di- and tetraploid accessions is therefore expected, while the significance of the negative

relationship between mean genome size of the diploids and seed length and width might be an artefact of the high number of samples (n=700) and the very small Spearman´s rho (less than –0,1). This is also suggested by relatively few accessions with the 10% larger genome sizes as compared with the large variation of the seed parameters. Also, the variation of the seed parameters does not correlate with small or large genome sizes.

Genome sizes of our investigated accessions increased slightly but significantly from North to South and from West to East. Former studies found positive correlations between the duration of vegetation period and genome sizes in various angiosperms (Bennett, 1987; Reeves et al., 1998; Turpeinen et al., 1999). This seems to agree with our data. But since *A. thaliana* can germinate in spring but also in autumn, which prolongs the vegetative period, our correlation disagrees with these results. Most accessions are summer annual; winter annuals are only found in Northern Europe (Laibach, 1951). The geographical separation between winter and summer annuals is a quantitative trend rather than a strict replacement, since even in Siberia, there are summer annuals among a majority of winter annuals (own observation). Thus, the question remains unanswered if the correlation of genome size and duration of vegetation period found in other angiosperms is characteristic for *A. thaliana* or is due to sampling artefacts.

Recent molecular studies of the *Arabidopsis* genome are related to the published *Columbia* sequence and comparisons with parts of the L*er* genome (https://www.ncgr.org/cgi-bin/cereon/cereon_login.pl). We found that there are accessions with a significantly larger genome size than the widely used accession *Columbia*. The increased genome size of some accessions might be due to a larger centromeric region (Bennett et al., 2003) but might also involve coding regions. A closer investigation of the sequences involved in the intraspecific genome size differences of *A. thaliana* might influence the interpretation of comparative studies of natural occurring variation including QTL analysis (Alonso-Blanco and Koornneef, 2000; Mitchell-Olds, 2001; Schmid et al., 2003).

Acknowledgements

We thank C. Koch for her technical assistance. Special thanks go to N. Ermakov and M. H. Hoffmann, who helped collecting *A. thaliana* in Siberia. This work was supported by grants from the Deutsche Forschungsgemeinschaft to K.B. and the Fonds der Chemischen Industrie to A.M.

Phenological and morphological plasticity differ among populations of a world-wide sample of *Arabidopsis thaliana* (L.) Heynh. under different temperature treatments

Matthias H. Hoffmann, Jürgen Tomiuk, Heike Schmuths, Christina Koch, Konrad Bachmann. Acta Oecologica, eingereicht.

Abstract

A world-wide sample of 74 accessions of *Arabidopsis thaliana* was studied at two temperature regimes (14°C and 22°C). Sixteen primary morphological and five phenological characters were scored. The variability of all characters was large and accessions cannot unambiguously be clustered. Correlations among characters are significant across both temperature treatments. The characters, however, are only weakly correlated with single nucleotide polymorphisms. The phenological responses to the temperatures used in the experiment (plasticity: character at 22°C minus character at 14°C) correlate significantly with temperature data from the sites of origin of the accessions (monthly mean temperature of January, April, July, and October). Accessions from winter cold areas develop faster in the warmer treatment. Accessions from winter warm areas are less retarded in their development in colder conditions. The results suggest that the observed patterns have evolved through climatic selection during the postglacial range expansion of *Arabidopsis*.

Keywords: phenotypic plasticity, morphometry, climate

Introduction

Arabidopsis thaliana is a frequently used object for studies of phenotypic plasticity. Environmental factors investigated in such studies were: light (e.g. Dorn et al., 2000; Pigliucci et al., 1995a, b, 1999; Pigliucci and Marlow, 2001; Pollard et al., 2001), nutrient availability (Pigliucci et al., 1995a, b; Pigliucci and Hayden, 2001), CO_2 (Norton et al., 1995), and density, i.e. competition between the plants (Orobovic and Tarasjev, 1999). Climatic factors, such as water availability and temperature were rarely used for studies of phenotypic plasticity although the large distribution range of the species suggests that the plants cope with a wide range of climatic conditions (Hoffmann, 2002). Water availability was used as an environmental signal by Pigliucci et al. (1995a, b) and temperature by Westerman and Lawrence (1971), Westerman (1971), and Lu and Wu (1986). The studies of Westerman and Lawrence (1971) and Westerman (1971) focus mainly on the temperature effect on the expression of four character states in a limited sample of accessions, some of them with unknown origin. These studies showed remarkable phenotypic plasticity in all studied characters. The effect of different thermal conditions on a world-wide sample of accessions from different climatic conditions has not yet been studied. Here, we assess the influence of two temperature regimes on 16 morphological characters and five primary phenological stages in a sample of 74 accessions. The difference of 8°C between the treatments (14 and 22°C) can be approximately related to the differences in the average May temperature found across the total north-south extension of the European range of the species.

We posed the following three questions: 1) What is the magnitude and direction of the change in character expression in the various accessions at the two temperature regimes? 2) Is there a correlation between character states and/or character responses and a set of nuclear genome wide single nucleotide polymorphisms (SNP)? 3) Is phenological plasticity associated with temperature conditions present at the place of origin of the plants?

Material and Methods

Seeds of different accessions were obtained from the Nottingham *Arabidopsis* stock centre. Additional accessions have been collected in the wild (accession names are set in italic in the text, *Sij* - Sijak, Uzbekistan; *Gat* - Gatersleben, Germany; *Sie* – Porta Turfi, Siena, Italy; *Par* – Paris, France; *Bad* – Badetz, Germany; *Han* – Münster Handorf Dorbaum,

Germany; *Gie* – Münster Gievenbeck, Germany; *Duel* –Dülmen-Dernekamp, Germany). The distribution of the 74 accessions initially used for the experiment is shown in Figure 1. Single seeds were placed in 6.5 x 6.5 cm pots filled with the autoclaved Klassmann Substrat2 substrate (Klassmann-Deilmann GmbH, Geeste, Germany), moistened and placed in the dark at 4°C for 7 days. Thereafter, the plants were transferred to growth chambers at constant 14°C and 22°C, respectively. Long-day conditions (20 h light, 4 h dark) were maintained during the experiment. In total, up to five plants per accession were cultivated. To minimise micro-environmental effects, the trays containing the pots were randomly rearranged on the shelves every second day. Pots were not allowed to dry up.

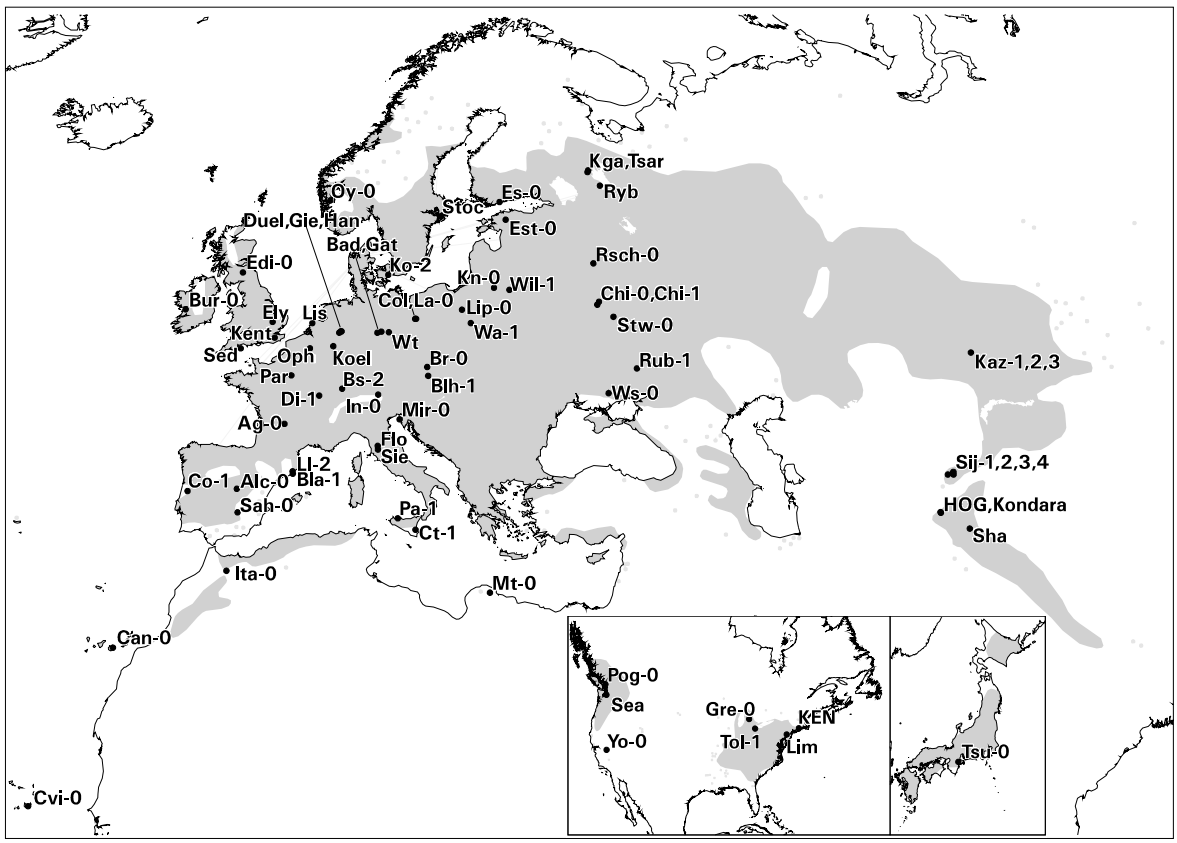


Fig. 1. Distribution of the 73 accessions used in this study and their acronyms. The grey area indicates the distribution range of *Arabidopsis thaliana*. Small grey points are occurrences outside the more or less densely inhabited distribution range. The insets show the accessions from North America and Japan.

Sixteen primary morphological traits were measured. Characters of the leaves: number of rosette leaves at bolting, number of cauline leaves, numbers of hairs on the fifth rosette leaf and on the first cauline leaf (hairs/10 mm^2) in the distal third of the leaf, length and width of the first cauline leaf. Characters of the flower (each character was measured on the lowermost seven flowers): length of the inner sepals, length and width of the petals.

Characters of the fruit (each character was measured on seven fruit in the lower part of the main flowering stem): length of siliqua, width of septum, length of style, number of ripe seeds per locus and number of sterile ovules per locus. From these data the total number of ovules per locus is derived. Furthermore, the numbers of lateral shoots from rosette leaf axils and cauline leaf axils were counted. The measurements were performed using a rule with smallest graduation mark of 0,1 mm. Subsequently, septum width and the length of the style were not considered because of their small size associated with large errors in measurement. The following phenological characters (phenological stages) were recorded: days to germination, days to expansion of first rosette leaf, days to bolting, days to opening of the first flower, and days to opening of the first siliqua. Some developmental phases or periods were derived from these primary phenological characters (phenological phases): vegetative period (days from germination to bolting, or to the opening of the first flower, both measures gave the same results), generative period (days from opening of the first flower to opening of the first fruit) and total life period (days from germination to opening of the first fruit).

Response to different growth condition (plasticity) has been calculated for every accession according the following formula:

Plasticity = Value (22°C) -Value (14°C).

Negative plasticity values indicate that the character score decreases between the 14°C treatment and the 22°C treatment.

The detection of the 67 genome wide SNPs is described in Schmuths et al. (2004b).

Climate data for the collection sites of the various accessions were kindly provided by W. Cramer (CLIMATE database version 2.1, Potsdam Institute of Climate Impact Research, Germany).

Statistical analyses were performed using the program SAS (1990). Analysis of variance (Anova) was done with the SAS procedure GLM using the model (Type III SS) with the effects accession, treatment and accession x treatment interaction. Principal component analysis (PCA, SAS procedure PRINCOMP) was applied to analyze dependencies among the phenological and morphological characters. All morphological and phenological characters were first included in the principal component analysis but some groups of characters were highly independent from each other. Most of the variation observed over all characters could be explained only by accepting many principal components and, therefore, the independent character groups were considered separately. In most cases, only factor loadings of variables of more than 0,5 and less than –0,5 are considered for

discussions (see Röhr, 1993). Linear correlations between temperature data and the phenological response of individuals from the accessions to the treatment were calculated by the SAS procedure REG. In order to compensate an inflating effect of type-one-error through multiple testing, a sequential Bonferroni procedure has been applied (Jaccard and Wan, 1996).

The Geographical Information System (GIS) Arc/Info (ESRI, 1992) was used for geographical analysis (preparation of data sets for comparisons of longitude and latitude with the biometric data).

Results

Plants of some accessions included in the experiment showed developmental disturbances in one of the two temperature treatments: The accessions *Ita-0*, *Yo-0* and *Sah-0* did not reach maturity in the 22°C treatment. *Ct-1*, *Mt-0*, *Oy-0*, *Sij-3*, and *Wt* failed to complete their life cycle in the 14°C treatment.

Variability of Arabidopsis thaliana *between the treatments*

Except the early phenological stages in the 22°C-treatment where the distribution is skewed negatively, i.e. plants of many accessions germinate, bolt and flower simultaneously early in the experiment, all morphological and phenological character are normally distributed. Analysis of variance (Anova) was performed on all original phenological and morphological data (Table 1). Comparisons between the treatments need to be divided into overall effects comparing the pooled data from all accessions between the two temperature regimes, and the comparison of the individual behaviour reactions of the accessions to the treatments (accession x treatment interaction).

1. Variability of pooled accessions between treatments. *Arabidopsis thaliana* (pooled data over accessions) showed significant differences between treatments in all studied morphological and phenological characters (Table 1). As expected, phenological stages were shorter at the higher than lower temperature. The quantitative morphological characters showed no consistent patterns at higher temperature. In the 14°C treatment, leaves were generally more numerous, longer and wider, and more shoots have been produced from rosette leaf axils and cauline leaf axils. The numbers of hairs per 10 mm² leaf area was much higher at 22°C, both for rosette and cauline leaves. The petals were

longer and wider at 14°C, while the sepals were longer at the 22°C. At 14°C the fruit were longer, more ovules were formed and more of them became mature but more sterile ovules were observed at 22°C.

Table 1. ANOVA with 74 accessions and two different temperature regimes. Increasing or decreasing characters between the temperatures are indicated in column 2 (14:22). Degrees of freedom (df) and mean of square sums (MS) are given for temperature, accession and temperature x accession.

trait	14:22	model (p<0,0001)		error		temperature (df = 1, p≤0,005)	accession (df = 73, p<0,001)	temperature x accession (p<0,001)	
		df	MS	df	MS	MS	MS	df	MS
Number of rosette leaves	>	137	1375,121	427	52,777	2870,156	1790,103	63	735,274
Number of cauline leaves	>	136	54,853	428	3,345	328,266	72,078	63	24,181
Number of hairs on cauline leaf	<	137	210,261	427	25,896	3933,247	227,966	63	83,356
Number of hairs on rosette leaf	<	132	20,395	417	2,896	682,891	21,129	58	5,809
Length of 1st cauline leaf	>	137	438,571	428	78,612	6330,269	453,657	63	364,951
Width of 1st cauline leaf	>	137	91,884	427	10,937	1925,233	97,300	63	41,025
Number of basal shoots	>	136	35,756	420	7,267	406,888	35,674	63	21,497
Number of lateral shoots	>	136	33,566	421	3,180	357,381	42,641	62	13,127
Sepal length	<	137	0,216	430	0,054	0,430	0,291	63	0,128
Petal length	>	137	0,915	430	0,138	6,700	1,077	63	0,584
Petal width	>	137	0,177	430	0,019	1,621	0,212	63	0,089
Fruit length	>	136	18,819	417	2,950	498,580	17,311	62	9,231
Number of ripe seeds	>	136	137,987	417	21,668	605,206	108,450	62	46,508
Number of sterile ovules	<	136	39,301	417	12,243	459,790	35,676	62	23,436
Days to germination	>	138	36,605	436	3,562	761,391	35,399	64	11,817
Days to first leaf	>	138	113,996	435	8,912	4687,449	75,163	64	18,661
Days to bolting	>	139	2847,589	434	94,465	72240,257	2594,053	64	527,766
Days to flower	>	137	3617,140	432	96,593	119908,792	2642,196	64	629,189
Days to fruit	>	136	5835,119	415	101,140	262810,274	2979,763	62	656,480

2. Accession-specific temperature responses. The temperature responses of individual accessions differed considerably from the overall pattern in amount and direction. *Post hoc* range tests and multiple comparisons revealed only overlapping groups. The only group that might be visually distinguished is a group of glabrous accessions (*Flo*, *Sie*, *Mir-0*, and *Br-0*; the latter has, however, very few hairs in the 22°C-treatment) vs. all other hairy accessions. This distinction, however, is statistically not significant because some accessions are subglabrous (e.g. *Duel*, *In-0*, *Kn-0*, *Wa-1*) with an average of 1-2 hairs per 10 mm².

There was no accession whose plants differed significantly between the two temperature treatments for all characters. On the other hand, plants of all accession showed at least for one character significant differences between temperatures. Phenological characters of

plants of most accessions differed significantly between the two treatments, only few specimens showed non-significant differences for several stages. Developmental rates were not significantly different between the treatments in the earliest stages (germination, expansion of the cotyledons and first rosette leaf) in some accessions (e.g. *Edi-0*, *Han*, *Lim*, *Pog-0*), but became significantly faster at 22°C in later stages of these accessions. The opposite pattern could be observed in some accessions (e.g. *Bla-0*, *Br-0*, *KEN*) where early stages were passed significantly faster at 22°C but developmental rates converged at later stages. This behaviour was most pronounced in the accession *KEN*, which needed a longer period to bolting and flowering at 22°C despite faster germination.

Vegetative characters showed a similar variability in their responses among accessions, i.e. no common patterns of differences were observed in different temperatures. Accessions formed either more (e.g. *Es-0*, *Flo*, *Mir-0*) or fewer (e.g. *Han*, *Tsar*, *Wil-1*) rosettes leaves at 22°C; some accessions showed no response in this trait (e.g. *Cvi-0*, *KEN*, *Kga*). Only *Bla-1* and *Ryb* had significantly more cauline leaves at the higher temperature, all other accessions with significant differences between the treatments had fewer cauline leaves at 22°C. Significant differences in the numbers of hairs on the rosette and cauline leaves were observed in accessions that had more hairs at 22°C. The leaf size was significantly greater at 22°C in a few accessions (length: *Gre-0*, *Sij-2*; width: *Gre-0*). Generally, at the higher temperature fewer shoots were formed from the axils of rosette and cauline leaves, with the exception of *Stoc* that had significantly more lateral shoots.

Measures of the flower characters revealed that plants from several accessions differed significantly in all of them (e.g. *Kent*, *Ryb*, *Sij-1*), while others did not differ significantly in any of these characters (e.g. *Can-0*, *Cvi-0*, *Kaz-2*). The size of flowers differed significantly among accessions at 22°C. Sepals of *Ag-0*, *Gat*, *Koel*, *Kent*, *Sed*, and *Sij-1* were longer and those of *Duel*, *Ryb*, *Sah-0*, *Stoc*, *Wa-1* and *Ws-0* were significantly shorter at 22°C. Petals were longer at 22°C only in *Kaz-3*, *Kent*, *Koel*, and *La-0*, whereas they were significantly shorter in 17 accessions (e.g. *Sij-1*, *Stoc*, *Ryb*). Significantly wider petals at 22°C were only observed in *Kent* and *Kaz-3* whereas plants of most accessions showed the opposite pattern (e.g. *Stoc*, *Ryb*, *Sij-1*).

The responses of fruit characters to temperature were also highly heterogeneous. A few accessions did not differ significantly in any of the characters between the treatments (e.g. *Ag-0*, *Pog-0*, *Tsar*), and no accession varied significantly in all of the fruit characters. Only the accession *Rsch-0* showed significantly longer fruit at 22°C in contrast to all other

accessions. The total number of ovules and the number of mature seeds was generally lower at 22°C. Only the accessions *Co-1* and *Stoc* had fewer sterile ovules at 22°C.

3. Accession x treatment interaction. The genotype x environment effect is significant for all characters indicating that the reaction norm of the plants is dependent on the genotype and environmental conditions.

Correlations of morphological and phenological character state expression

Including first all characters into the principal component analysis, 90% of the variation of the characters can be explained only by assuming seventeen factors with an eigenvalue larger than 1 (data not shown). The high number of factors indicates a high independence among the characters and cannot simply be explained. Therefore, we decided to split the data set into four independent groups of characters (Table 2): 1) vegetative characters (characters of the leaves and the numbers of basal and lateral shoots), 2) characters of the flower (sepals and petals), 3) characters of the fruit (fruit length and number of sterile and fertile ovules), and 4) phenological characters.

Vegetative characters. The traits summarized as vegetative characters are the most heterogeneous data set of the four complexes. Four principal components are necessary to explain 85% of the variance. The first factor mainly determines the number of rosette and cauline leaves as well as the number of lateral shoots from the leaf axils of the cauline leaves. The size of the cauline leaves is negatively correlated with these traits indicating some trade-off between numbers and sizes of leaves. Factor 2 relates particularly to the width of the cauline leaves that is negatively correlated with the numbers of hairs on the leaves. The length of the cauline leaf loads high on factor 3, while the number of shoots from rosette leaf axils loads high on factor 4.

Flower and fruit characters. The sizes of sepals and petals are highly correlated. Similar high correlations show the fruit characters, whereas the number of ripe seeds is comprehensibly negatively correlated to the number of sterile ovules.

Phenological characters. The PCA across all phenological characters revealed three, apparently more or less independent phenological phases. ‘Phenological phases’ are intervals, e.g. days from germination to bolting, whereas ‘phenological stages’ refer to particular points in the development of the plants, e.g. days to opening of the first flower.

Table 2. Principal Component Analysis (PCA) of morphological and phenological characters of *Arabidopsis*. The original data set is split into four independent groups of characters (1. vegetative characters: characters of the leaves and the numbers of basal and lateral shoots, respectively, 2. characters of the flower: sepals and petals, 3. characters of the fruit: fruit length and number of sterile and fertile ovules, and 4. phenological characters; see text for detail).

	PC 1	**PC 2**	**PC 3**	**PC 4**
	Vegetative characters			
cumulative factor loadings	**0,3515**	**0,6198**	**0,7433**	**0,8528**
Number of rosette leaves	0,503524	0,089634	0,039771	-0,250314
Number of hairs on rosette leaf	0,058335	-0,451488	0,463868	0,456520
Number of hairs on cauline leaf	0,241529	-0,441469	0,390003	0,072168
Length of 1st cauline leaf	-0,298403	0,343782	0,602717	0,047638
Width of 1st cauline leaf	-0,336395	0,439334	0,318879	0,010828
Number of cauline leaves	0,501957	0,283767	0,203453	0,001694
Number of basal shoots	0,135380	0,295151	-0,301820	0,848613
Number of lateral shoots	0,460684	0,338217	0,183586	-0,034515
	Floral characters			
cumulative factor loadings	**0,7580**	**0,9298**	**1,000**	
Sepal length	0,535322	0,806272	0,251705	
Petal length	0,615129	-0,167927	-0,770336	
Petal width	0,578833	-0,567209	0,585856	
	Fruit characters			
cumulative factor loadings	**0,7044**	**0,9269**	**1.000**	
Fruit length	0,569335	0,610080	-0,551053	
Number of ripe seeds	0,640470	0,091087	0,762562	
Number of sterile ovules	-0,515418	0,787086	0,338879	
	Phenological characters			
cumulative factor loadings	**0,7724**	**0,9233**	**0,9817**	
Days to germination	0,56196	0,79390	-0,0079	
Days to first leaf	0,71483	0,64272	0,0303	
Days to bolting	0,96067	-0,125001	-0,222	
Days to flower	0,97696	-0,119063	-0,173	
Days to fruit	0,98975	-0,135723	0,0343	
Vegetative period	0,95826	-0,218112	-0,180	
Generative period	0,79037	-0,149147	0,593	
Total life time	0,97610	-0,213048	0,0363	

The time to germination and to the expansion of the first leaf (earliest stages of the plant development) load together on factor 2, whereas the following stages and the vegetative period and the total lifetime load particularly on factor 1. The generative phase, as measured the time from the first flower to the shedding of the first seeds appears again somewhat independent from the other development loading on factor 3.

Responses of the characters to the temperature treatments. The accessions show different responses of character state expressions to the different treatments (for examples, see Fig. 2). The responses across all accessions resemble a bell-shaped curve and are accession dependent. Positive responses indicate that the character is in the 22°C treatment larger than in the 14°C treatment. Phenological responses are mostly negative, i.e. the development is faster at 22°C than at 14°C.

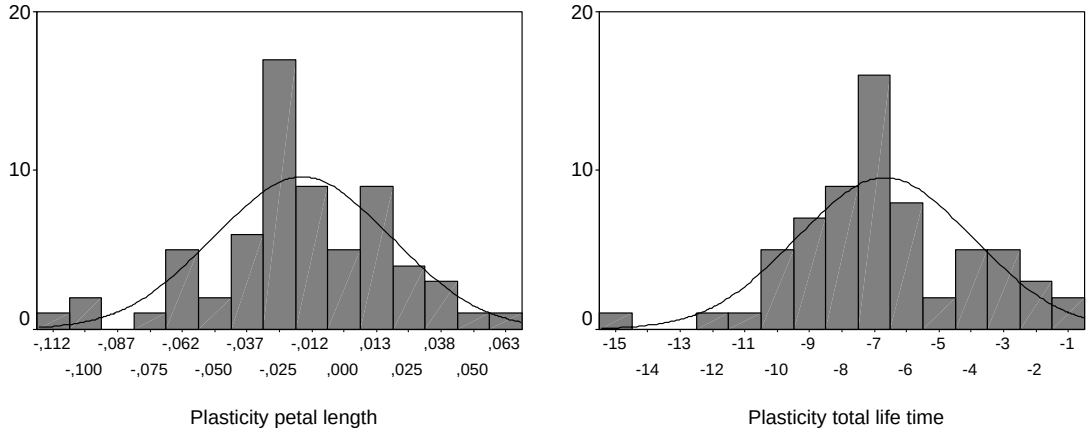


Fig. 2. Examples of the accessions' responses to the different treatments. The normal distribution is superimposed on the data. Positive responses indicate that the character is in the 22°C treatment larger than in the 14°C treatment. Phenological responses are mostly negative, i.e. the development is faster at 22°C than at 14°C.

Correlations of the morphological and phenological characters with climate data and geography

Correlations between the response of the phenological characters explaining most variance in the PCA (Table 2) and a monthly representative of seasonal temperatures have been calculated (Table 3). As representative for seasons we used the temperatures of January, April, July, and October, respectively. Most correlations between January, April and October temperature, respectively, and phenological response are significant. The data indicate that the phenological response of the accessions to the two temperature treatments is positively related to the temperatures at the original accessions site from autumn until spring (October $p<0,022$; January $p<0,019$; April $p<0,064$; July $p<0,220$; see Table 3).

The positive correlation of all characters indicates that accessions from cold winter areas have a tendency to have a more negative phenological plasticity. These accessions developed at 22°C much faster than in the 14°C treatment.

Table 3. P-values of correlations between phenological response and temperature data at the accessions' places of origin. Sequential Bonferroni procedure has been applied separately for each character because of the high correlation among the here defined phenotypic plasticities. Significant type-one-errors are marked by an asterisk.

	January	April	July	October
Days to bolting	0,018	0,064	0,22	0,022
Days to flower	0,019	0,042	0,114	0,015
Days to fruit	0,007*	0,012*	0,054	0,005*
Vegetative period	0,015*	0,029	0,057	0,008*
Total life time	0,005*	0,008*	0,026*	0,002*

The development of accessions from warmer areas seem not as much accelerated at higher temperatures, i.e. the developmental rates are more similar between the temperature treatments. Table 3 shows also that the July (summer) temperature significantly correlates only with the total life time of the plants. This result appears comprehensible because the plants pass the summer mostly as seeds.

Accessions from winter cold sites of origin (e.g. *Tsar*, *Ryb*, *Chi-1*, *Kaz-1*) had a low negative response of time to bolting, flowering, and fruiting, and a low negative plasticity in the vegetative developmental rate. Plasticity of the length of the vegetative period, i.e. the duration from germination to bolting, showed that accessions from cold winter Middle Asia and East Europe had mostly intermediate and highly negative values of plasticity, while accessions from the rather winter warm Mediterranean region exhibited mostly low positive values.

Association of morphological and phenological characters with genome-wide SNPs

Associations between various character states and a genome-wide set of 67 single nucleotide polymorphisms that are described for populations from the original sites of our accessions (Schmuths et al., 2004b) have been analysed with the non parametric Mann-Whithney-U-Test. All morphological and phenological characters were only weakly correlated with genetic variability. Nevertheless, some associations were suggestive: The measures of the petal size correlated negatively with two markers located 140 bp from each other in the *PROLIFERA* gene on chromosome 2 (MASC05398 and MASC05481, $p<0,05$). MASC03341 (located in an unknown gene) correlated positively with the length and width of the first cauline leaf and the plasticity of this character ($p<0,01$).

Discussion

In this study we have chosen the test variable, *temperature*, because it may be one of the most important abiotic factors limiting plant distribution ranges (e.g. Kullman, 1996; Woodward, 1987). Almost all of the investigated characters showed significant differences in their expression under different temperatures. This confirms the high plasticity of *A. thaliana*, which has been reported in response to different environmental factors by other authors (e.g. Westerman and Lawrence, 1970; Pigliucci et al., 1999; Dorn et al., 2000; Pigliucci and Marlow, 2001; Pollard et al., 2001).

Most of the observed character correlations (data not shown) appear intuitively comprehensible. The positive correlation of the number of rosette leaves with the number of lateral shoots at the main flowering stem appear plausible in the light of a possible translocation of nutrients and an increased vigour. With few exceptions (flower size and the number of sterile ovules) correlations among various characters detected among the different accessions are the same at both temperatures. This points to a stable phenotypic adaptation and is in accordance with the conclusion given by Pigliucci and Marlow (2001) who stated that a short-term temperature change does not essentially alter the phenotypic integration. Different light regimes had also little effect on the character correlations of *A. thaliana* (Pigliucci et al., 1999) and other plant species (e.g. Battjes and Bachmann, 1994). Other environmental factors, for example simulated season lengths (Pigliucci and Marlow, 2001), are apparently more likely to change the phenotypic integration in *A. thaliana*.

Many accessions of *A. thaliana* do not show a consistent pattern in their responses of morphology and phenology to temperatures, i.e. values are higher at 22°C than at 14°C in some accessions and lower in others. These results differ from those presented by Westerman and Lawrence (1970) who found that the response to temperature had the same directions and differed only slightly in magnitude. Apparently these results are due to their selected sample of accessions or to the Agar culture medium used in their study. Depending on the investigated character many (phenology) or comparatively few accessions (morphology) behave or express characters significantly differently in the temperature treatments.

The character values in each treatment group and the responses to the temperature treatments are distributed normally. Such a normal distribution has traditionally been interpreted as the cumulative influence of many genes with small and additive effects (Sokal and Rohlf, 1995; Westerman, 1971). This would not be surprising, since 94 genes distributed across the genome have been found to be involved in leaf development in *A.*

thaliana (Berná et al., 1999; Robles and Micol, 2001) and at least 63 QTLs in inflorescence development (Ungerer et al., 2002). On the other hand, only two QTLs explain as much as 38% of the genetic variation among accessions for the timing of the switch from the vegetative to the generative phase (Mitchell-Olds, 1996), even though there are more than 80 genes known influencing this trait (Blázquez et al., 2001).

We have examined a large sample of accessions for plants that show exceptionally high responses to the difference in environmental temperature. No accession is observed whose plants show significant plasticity at all characters. Accessions at the extremes of the plasticity distributions for many characters are *Bla-1*, *Han*, *Mir-0*, and *Ryb*, while *Cvi-0*, *Di-1*, *Est-0*, *Kaz-2*, whereas *Ws* is among the least plastic ones. The original sites of these accessions are geographically widely distributed, and obviously frequently used laboratory strains are not necessarily less plastic than plants collected in the wild.

Other authors observed only in few cases correlations between climate, site of origin of the accessions, morphological traits, and plasticity. Li et al. (1989) found an impact of the latitude of the site on some traits of plants. Differently, most of the statistically significant correlations between character states and latitude and longitude found by us seem to be spurious because of the inconsistent patterns across the data set (data not shown). The discrepancy between results, however, may be accounted for the different characters studied, different growth conditions in the experiments but also for differences in sampled accessions. In this study 24 out of 40 accessions analysed by Li et al. (1989) were studied, but many additional accessions from Middle Asia were included.

A positive correlation between plasticity of phenological characters and climate at the site of origin indicates that development of plants of accessions sampled from areas with cold winter is much more accelerated at high temperatures than that of plants of accessions originating from areas with warm winters (Fig. 3). Positive values of plasticity are very scarce and accessions from areas with warm winter have only slightly negative values of plasticity indicating that different growth temperatures have minor effects on their developmental rates, i.e. there is a minor boost of growth rates if plants grow at the 22°C. Differently, plants from cold areas can use short time windows for their growth during the summer more efficiently than those from warm areas, and this may be the selection factor that determines the different responses. Furthermore, temperature may stimulate more strongly development of plants from accessions sampled from cold areas and, therefore, these plants can pass much faster from the vegetative to the generative phase at the onset of spring than accessions from warm regions.

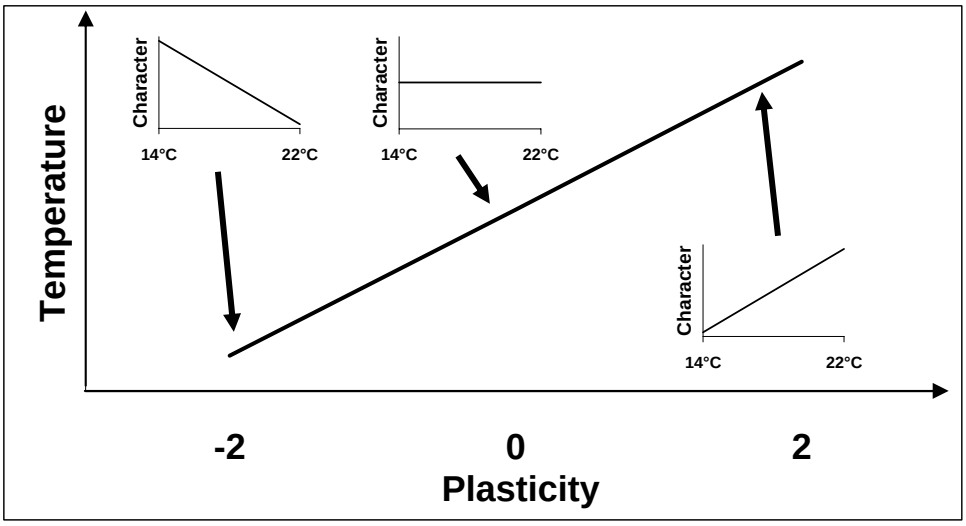


Fig. 3. Schematic representation of a positive correlation between plasticity of a particular morphological or phenological character and the climate data from the accessions' place of origin. The inset diagrams indicate the response of the character to the different treatments. For example, the upper left diagrams may indicate that in this particular accession the size of an organ is longer in the 14°C treatment than in the other one. Zero plasticity (middle inset diagram) means that the character is in both treatments equally developed. This diagram shows that accessions from colder areas have a negative plasticity in that character, i.e. the character is, for example, longer in colder growth conditions (14°C treatment), and that accessions from warmer areas show the opposite pattern.

In consistency with observed geographical patterns at the molecular level (Sharbel et al., 2000; Hoffmann et al., 2003; Schmuths et al., 2004b), the here observed correlations among the magnitude of the phenological response to our temperature treatments and the temperature at the accessions' site of origin suggest that differences are the result of local adaptation after the postglacial range expansion of the species, but the long lasting impact of humans on landscape structures and plant dispersal must also be considered when distribution patterns of ecotypes are analysed.

Acknowledgments

We thank B. Sperling and I. Faustmann for technical assistance, and R. Horres and R. Fritsch for discussions. A. Börner kindly provided us with the climate chamber. The work was supported by the Deutsche Forschungsgemeinschaft.

Evidence for pleistocene refugia and postglacial expansion of *Arabidopsis thaliana* from genome-wide SNP markers

Karl J. Schmid, Ottó Törjék, Rhonda Meyer, Heike Schmuths, Matthias H. Hoffmann, Thomas Altmann. Molecular Biology, eingereicht.

Abstract

To characterize geographic and historical patterns of genetic variation in the model plant *Arabidopsis thaliana*, we have genotyped 115 genome-wide single nucleotide polymorphism (SNP) markers in 351 accessions from the whole species range using MALDI-Tof assays and found isolation by distance and a high proportion of ancestral alleles among *A. thaliana* accessions from the Iberian Peninsula and Central Asia. The analysis of genetic relationships among accessions revealed a star-like dendrogram with low bootstrap support, but also numerous accessions that are genetically nearly identical. In a model-based inference of population structure, accessions from southern Europe and Central Asia were grouped into two separate populations, whereas central and eastern European accessions appear to represent admixed individuals whose genomes were reshuffled by historical recombination events.

In summary, the SNP data suggest that *A. thaliana* colonized Central and Eastern Europe from Pleistocene refugia in southern Europe and, possibly, from Central Asia. However, the low level of gene diversity and the unusual relationship between geographic and genetic distance in Central Asia are also consistent the hypothesis that this region may not have been a glacial refugium, but was colonized postglacially by a small population of migrants from a western refugium, such as the Balkans. The present data underline the importance of a thorough understanding of historical and geographic effects on studies of genetic and phenotypic variation.

Keywords: *Arabidopsis thaliana*, single nucleotide polymorphisms, population structure, postglacial colonization, biogeography

Introduction

The mouse-ear cress *Arabidopsis thaliana* (L.) Heynh. has become an important model system for studies of plant evolution and ecology (Mitchell-Olds, 2001) and it is characterized by abundant naturally occurring genetic and phenotypic diversity (Alonso-Blanco and Koornneef, 2000). *A. thaliana* is a highly self-fertilizing species (Redéi, 1975; Abbott and Gomes, 1989). This mode of reproduction has several important consequences for the structure of genetic variation. For example, in selfing species a single seed is sufficient to colonize a new habitat because no mating partners need to be found. Effective recombination rates are reduced in selfing species and larger regions of the chromosome share the same evolutionary history. Coadapted gene complexes are stable over long periods of time and may facilitate local adaptation (Allard et al., 1968) although in inbred populations, the origin of coadapted genes is limited to rare admixture events and novel mutations.

Genetic diversity in populations of *A. thaliana* was examined using a variety of genetic markers, including allozymes (Kuittinen et al., 1997), microsatellites (e.g., Innan et al., 1997; Kuittinen et al., 1997; Loridon et al., 1998), cleaved amplified polymorphic sequences (CAPS) (King et al., 1993; Hanfstingl et al., 1994; Hardtke et al., 1996; Lättig et al., 2003; Barth et al., 2002), and the DNA sequencing of multiple nuclear and mitochondrial loci (e.g., Hanfstingl et al., 1994; Hardtke et al., 1996; Kawabe et al., 1997; Bergelson et al., 1998; Kawabe and Miyashita, 1999; Purugganan and Suddith, 1999; Kuittinen and Aguadé, 2000; Savolainen et al., 2000; Haubold et al., 2002; Nordborg et al., 2002). Observed patterns of genetic variation are consistent with a self-fertilizing organism. Genetic diversity in local populations is low and most genetic variation is present between populations (Hanfstingl et al., 1994; Todokoro et al., 1995; Bergelson et al., 1998; Breyne et al., 1999; Miyashita et al., 1999; Barth et al., 2002; Clauss et al., 2002), which is expected when local populations are founded from a single or few highly inbred individuals. Heterozygosity is strongly reduced and likely a consequence of the repeated inbreeding of parental lines. Genome-wide levels of linkage disequilibrium (LD) extend up to 250 kbp and are also consistent with a reduced effective recombination rate due to selfing (Nordborg et al., 2002).

A. thaliana occurs naturally in Eurasia and North Africa (Price et al., 1994; Al-Shebaz and O'Kane, 2002), and has expanded its species range into East Asia and North America where favourable climatic conditions exist for this species (Hoffmann, 2002). Despite numerous attempts to characterize the genetic population structure and expansion history

of *A. thaliana*, no consistent model has yet emerged. Most studies have found no correlation between genetic and geographic distance (e.g., King et al., 1993; Todokoro et al., 1995; Bergelson et al., 1998; Miyashita et al., 1998; Barth et al., 2002). The absence of a geographic signal has often been explained by the biological features of *A. thaliana* such as a lack of competitiveness and a preference for disturbed habitats (e.g. Bergelson et al., 1998; Mauricio, 1998). Since agricultural fields and other sites influenced by human activity are highly disturbed habitats, a human-induced admixture of geographically separated populations may be responsible for the lack of a relationship between genetic and geographic distance. In contrast, an AFLP study of 142 accessions (Sharbel et al., 2000), a meta-analysis of previous sequence surveys of individual loci (Hoffmann et al., 2003b), and a sequence survey of a region on chromosome 2 (Schmuths et al., 2004b) uncovered a large-scale population structure in *A. thaliana*. In particular, populations were isolated by distance in the Mediterranean region and in Central Asia possibly reflecting the legacy of past ice ages. The climatic and vegetational changes of the Pleistocene affected the distribution of many plant species in Europe, and both palaeoclimatic and genetic data indicate that glacial refugia were located in Southern Europe (the Iberian Peninsula, Italy, and the Balkans) and in Central Asia (Willis, 1996; Comes and Kadereit, 1998; Hewitt, 1999). Many species appear to have expanded from these refugia into Central Europe during the past 18,000 years. Based on these observations, Sharbel et al. (2000) postulated that the observed genetic population structure of *A. thaliana* may result from postglacial population expansion out of Pleistocene refugia and a subsequent admixture of populations in Central and Eastern Europe.

In this study, we present a genome-wide survey of genetic variation among natural accessions of *A. thaliana* using single nucleotide polymorphisms (SNPs) as genetic markers. Single nucleotide polymorphisms are abundant in the genome of *A. thaliana* (Jander et al., 2002; Schmid et al., 2003), and several methods are available for high throughput genotyping at moderate cost. SNPs are not only excellent markers with which to rapidly genotype newly created populations to be used in genetic mapping experiments (Cho et al., 1999; Törjék et al., 2003), but they are also useful for studying the evolutionary history of a species and for estimating population parameters of genetic diversity (reviewed by Brumfield et al., 2003). We utilize a set of framework SNP markers (Schmid et al., 2003; Törjék et al., 2003) that are evenly distributed throughout the genome, separated by an average distance of 1,13 Mbp, to genotype nearly all accessions that are currently available from the public *Arabidopsis* stock centers and an additional set of accessions that

was recently collected in Central Asia. The major goals of this study are (i) to identify redundant (i.e., genetically identical) accessions in stock collections; (ii) to evaluate the evolutionary relationship among accessions; (iii) to identify a core collection of accessions that are useful as parental lines or control populations in genetic mapping studies; and (iv) to examine the relationship between geographic and genetic distance.

Materials and Methods

Plant material

Seeds of accessions analysed in this study were obtained from various sources. Accessions available through the *Arabidopsis* stock centers include *Col-0* from G. Rédei (Univ. of Missouri-Columbia, USA); *C24* from JP Hernalsteens (Vrije Universiteit Brussels, Belgium); Landsberg *erecta* from M. Koornneef (Wageningen University, Netherlands); *Ag-0*, *An*, *Bch-1*, *Bur*, *Cal*, *Co*, *Cvi*, *Ei*, *Eil-0*, *Gr*, *Hi*, *Lip-0*, *Lm*, *Lu*, *Ob-0*, *Old-1*, *Per*, *Oy*, *Sue*, *Sg-1* and *Te* from from S. Misera (Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany); a further 286 accessions from the Nottingham *Arabidopsis* Stock Center (NASC), the *Arabidopsis* Biological Resource Center (ABRC), Columbus, Ohio, and the SENDAI *Arabidopsis* seed stock center (SASSC), Sendai, Japan. After receipt of seeds from the above sources, accessions were propagated in bulk at the Institut für Genbiologische Forschung GmbH, Berlin, Germany, and at the Max-Planck-Institute of Molecular Plant Physiology, Golm, Germany. Prior to this analysis a single representative individual was selected for each accession from which progeny was amplified. An additional 41 accessions were included that originate from various locations in Russia and Uzbekistan. These accessions were collected in 2001 and 2002 by H. S. and M. H. and have not yet been deposited in stock centers. Accessions were grouped into the following seven geographical regions: Scandinavia; British Isles, Central Europe (Northern coast to the Alps); Iberian Peninsula; Southern Italy (south of the Po valley); Asia and Africa. These regions reflect the physical barriers that have separated geographical regions during glaciation and include several refugial areas (Hewitt, 1999).

SNP markers

SNP markers used in this study are a subset of polymorphisms detected in a survey of 13 different *A. thaliana* accessions (Schmid et al., 2003; Törjék et al., 2003), which are accessible via the GABI-MASC SNP database (http://www.mpiz-koeln.mpg.de/masc/). A framework set of 115 SNPs was identified from all available markers based on the following criteria (Törjék et al., 2003): (i) SNPs are polymorphic between the *C24* and *Col-0* accessions as they are also utilizied for genotyping mapping populations derived from these two accessions; (ii) SNPs should be spaced ca. 1,15 Mbp; (iii) if possible, SNPs should be polymorphic between L*er* and *Col-0* in addition to *C24*/*Col*. Six of these markers are located between framework marker sites and are separated from the nearest neighboring marker by 400 to 52661 bp.

DNA isolation and genotyping

Genomic DNA was extracted from 50 mg leaf tissue of 2-3 plants each using the NucleoSpin Multi-96 Plant kit (Macherey-Nagel, Düren, Germany) or the DNAeasy extraction kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. The collected accessions were genotyped with the set of 115 SNP markers, established for MALDI-ToF analysis (performed by GAG-Bioscience GmbH, Bremen, Germany). A subset of SNPs was genotyped in a smaller number of accessions using the SNaP- shot method (Applied Biosystems). Both MALDI-ToF analysis and SNaPshot reactions were carried out as described (Törjék et al., 2003).

Data analysis

The level of polymorphism at a marker was estimated as gene diversity,

$$H = \frac{n}{n-1}(1 - \sum^{n} p_i{}^2),$$

where n is the number of alleles and p_i is the relative frequency of allele *i* of a single SNP (Nei, 1987). Polymorphism across loci in a group of accessions was estimated as the mean of gene diversity among all SNP markers. A neighbor-joining (NJ) tree (Saitou and Nei, 1987) was constructed with the *neighbor* program of the PHYLIP package (Felsenstein, 1989) using a matrix of the proportion of uncorrected pairwise differences. SNPs with missing data in pairwise comparisons were excluded (pairwise deletion). Levels of

pairwise gametic linkage disequilibrium were estimated as r^2 (Hill and Robertson, 1968). Tests for significant pairwise LD were conducted with the χ^2 statistic (Weir, 1996) and sequential Bonferroni correction of *P*-values (Sokal and Rohlf, 1995). Correlations between matrices of genetic and geographic distances were calculated as the normalized Mantel statistic *r* (Mantel, 1967) with 999 permutations, using R PACKAGE 4.0 (www.fas.umontral.ca/BIOL/Casgrain/en/labo/R). Geographic regions with a high degree of homogeneity (i.e., a high proportion of shared nonvariable SNPs) were identified by the following *ad hoc* method. Individual SNPs were transformed into a matrix in which one state was encoded as „1" and the other as „2". For each of the two alleles of a SNP, a grid consisting of 50 rows and 214 columns, which covers that Eurasian distribution range, was generated. The two grids were added together. Grid cells containing accessions from the same locality having different SNPs were assigned the value „3" (e.g., *Bla-1* and *Bla-4*). Spatial homogeneity was calculated using the FOCALSTD (focal standard deviation) function of the Arc/Info geographical information system package (Environmental Systems Research Institute, 1992), which calculates the standard deviation (SD) for every cell in the grid from a defined number of neighborhood focal cells. If a SNP in a focal region is homogeneous, SD=0; otherwise SD>0. The grids obtained from the FOCALSTD calculation were transformed for subsequent calculations. In a last step, the grids for the individual SNPs were added and the spatial homogeneous zones for all data determined manually. Our calculations used different focal sizes: large ones were used to determine large-scale homogeneous regions (e.g., Central Asia) and small ones to define homogeneous zones at the regional level, e.g., within Central Europe.

The population structure was also examined with a model-based approach using the *structure* program, which uses a Bayesian framework to infer the number of populations in a sample and to assign individuals to different populations (Pritchard et al., 2000). We used a model with admixture and allowed 10,000 runs for „burn-in" and an additional 20,000 repetitions to estimate the parameters. The number of populations inferred ranged from 1 to 30 and five simulation runs were conducted for every population size.

SNP alleles were classified as ancestral or derived by comparison with outgroup species in which the SNPs were also genotyped. The outgroup species include *Arabidopsis lyrata* ssp. *petraea*, *Arabidopsis cebenennsis*, *Arabidopsis halleri*, and *Boechera drummondii*. *A. thaliana* SNP alleles were classified as „ancestral" if they were identical to the outgroup alleles and as „derived" if they differed from the outgroup alleles. SNPs with variable

outgroup alleles were excluded. The proportion of the ancestral SNP alleles in the species distribution range was interpolated using the inverse distance weight (IDW) function of Arc/Info with a power of two and a sample size of twelve points.

To identify a core collection of accessions, we ranked sampled accessions with respect to their contribution to the overall gene diversity of the collection. First, the two accessions with the largest pairwise number of differences were identified in the input set and then transferred to the output set. All accessions in the input set were then individually added to the output set and the one that most increased its genetic diversity was kept in the output set. This procedure was repeated until all accessions were moved to the output set. The „robustness" of the core collection with respect to the markers used was tested by bootstrap analysis (10,000 iterations). The „efficiency" of the algorithms was evaluated by comparing gene diversity of the core collection of a given size with the diversity of 1,000 random sets of the same size.

If not indicated otherwise, statistical calculations were conducted with the R statistical package (www.r-project.org).

Data availability

Summary information about the accessions and SNP markers used in this study, and the genotype data in tabular form are available at http://www.mpimp-golm.mpg.de/arab-diversity. A physical map of the SNP markers is available in chapter 9, Supplementary Data (Fig. 1). The SNP genotypes were also deposited in the GABI-MASC SNP database (http://www.mpiz-koeln.mpg.de/masc) and in NCBI dbSNP (accession numbers: XXX - XXX).

Results

Summary of genotyping

Three hundred and fifty-one *A. thaliana* accessions were genotyped for 115 SNPs using the MALDI-ToF assay. From these, a set of 335 accessions was included in further analyses. Seven accessions (*Fi-0*, *Fr-7*, *Ga-0*, *Kl-2*, *Kr-0*, *Cnt-1*, and *Dar*) were excluded because of a high proportion of missing data (> 23 missing genotypes, 20%), and nine accessions were left out (*Sed*, *Lis*, *Bla-14*, *Po-0*, *Pr-0*, *Ts-5*, *Tsu-0*, *Wl-0* and *Old-1*) because more than 5% of markers were heterozygous.

A total of 115 SNPs × 335 accessions = 38, 525 nucleotide changes were included in the analyses. One SNP was triallelic (MASC01582) and all others biallelic. Among all genotypes, 3,324 (8,6%) constitute missing data due to failure of the target region PCR amplification and 78 (0,2%) are heterozygous. Genotypes were classified as being of *Col-0* type or *C24* type. The mean frequency of the *Col-0* type genotypes (37,2%) was less than the frequency of the *C24* type genotypes (53,5%); the mean allele frequency of the *C24* type alleles among all markers was significantly higher than the frequency of the *Col-0* type alleles (Wilcoxon's signed rank test, V=4040,5, P=0,04913). Fig. 1 shows a histogram of allele frequencies. Low-frequency alleles range from 0,009 (*C24* specific allele; MASC02644) to 0,497 (*Col-0* specific allele; MASC03340). Forty-eight SNPs (41,74%) are rare polymorphisms with a frequency of less than 10% of the minor allele.

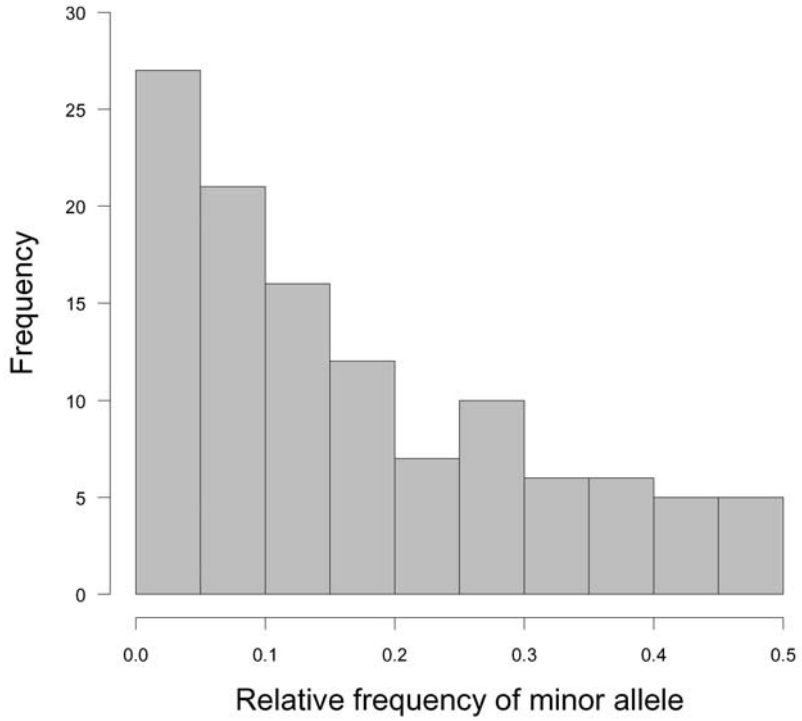


Fig. 1. Frequency distribution of the minor allele of 115 SNP markers among 335 accessions of *A. thaliana*.

Most SNP markers are located in intergenic regions (64%). A small fraction (8%) is located in coding regions and causes changes in the amino acid pattern (Table 1). SNPs of the latter type have the lowest gene diversity among all SNP types, but the difference compared to other SNP types is not significant (one-way ANOVA: F=0,2271, d.f.=3, P=0,87), which indicates that the different SNP types can be combined in analyses of variation.

Table 1. Number and mean gene diversity, *H*, of different SNP types.

SNP type	Count	H (SD)
Coding replacement	9	0,206 (0,052)
Coding synonymous	22	0,233 (0,085)
Intergenic	74	0,250 (0,186)
Noncoding	10	0,247 (0,059)
Total	115	0,245 (0,015)

Error rate in SNPs typing

We assessed the reliability of the MALDI-ToF genotypes using two approaches. First, five randomly chosen accessions (*Ag-0*, *An-2*, *Bla-5*, *Nok-2*, and *Np-0*) were genotyped twice by MALDI-ToF with the whole set of SNP markers. Among the five duplicate sets (5 x 115=575 comparisons), 521 genotypes (86,6%) could be obtained from both duplicates and 23 genotypes (4,0%) from only one duplicate, and 54 genotypes (9,4%) failed in both duplicates. Among 521 pairwise comparisons, only one (0,2%) differed between two duplicates. Second, we resequenced 29 SNPs from 16 accessions (464 genotypes) with the SNaPshot method. From the resequenced genotypes, 408 (87,9%) could be obtained from both replicates and 55 genotypes (11,7%) from only one duplicate, and one genotype (0,22%) failed in both duplicates. Of genotypes from both duplicates, four (0,98%) differed between the MALDI-ToF and SNaPshot replicates, and among those, three were determined to be heterozygous by either MALDI-ToF or SNaPshot assays. These controls indicate an overall error rate in this analysis of less than 1%.

Linkage disequilibrium (LD)

Pairwise LD was estimated for all pairs of SNPs with a frequency of >0,1 of the minor allele. The mean r^2 for all 2080 pairwise comparisons was 0,0183 and the median was 0,0080. Despite a low average r^2, 372 (18%) pairwise comparisons still exhibited significant LD after Bonferroni correction (P<0,05). A correlation of r^2 and physical distance was not significant (not shown). However, five of the six pairwise comparisons with r^2>0,2 include markers that are located adjacent to each other. Two markers (MASC0790 and MASC01171) are very close to each other (a distance of 400 bp) and are essentially in complete LD (r^2=0,98). Marker MASC4123, which is located on the tip of chromosome 4, is insignificant LD with both neighboring markers (MASC02820 and MASC04123), with r^2 values of 0,29 and 0,21, respectively. One pair of markers with high

r^2 is located on two different chromosomes: MASC09216 on chromosome 4 and MASC03470 on chromosome 5 (r^2=0,33). Since values of r^2 are generally low, it remains to be determined whether a significant linkage also exists between other polymorphisms from these regions. This assessment of LD may be conducted by sequencing regions harboring these rather distantly spaced markers.

Relationships among accessions

A neighbor-joining tree that was based on the pairwise genetic distance (Fig. 2) shows that the two accessions from which the SNPs were ascertained (*Col-0* and *C24*) are maximally distant from each other.

With the exception of the Central Asian accessions, the tree reveals the previously observed „star phylogeny" (Sharbel et al, 2000) with low bootstrap support of internal branches. A small number of accessions are highly similar to *C24* or *Col-0*. For example, the *Col-2*, *Col-3*, *Col-5* accessions are close to *Col-0*. It is interesting to note that the *Co* accession originating from *Coimbra* (Portugal) clusters closely with *C24*, indicating that the *C24* accession, whose origin is unknown, is probably derived from *Co* and not from *Col-0*, as is sometimes suggested in the literature (e.g. Loridon et al., 1998). The widely used Landsberg *erecta* accession is similar to the *DijonG*, L*er*-1 and *S96* accessions. Accessions whose genotypes do not differ by more than 2 SNPs (the expected error rate of the MALDI-ToF assay) can be considered genetically identical. Using this criterion, 96 accessions can grouped into 35 sets of essentially identical accessions (9. Supplementary Data, Fig. 3). Several pairs consist of accessions from distant geographic origin, for example: *Ct-1* (Catania, Italy) and *EnkD* (Enkheim, Germany); *Bu-2* (Burghaun, Germany); and *Bla-4* (Blanes, Spain) and *Mt-0* (Martuba, Libya) and *Ma-0* (Marburg, Germany).

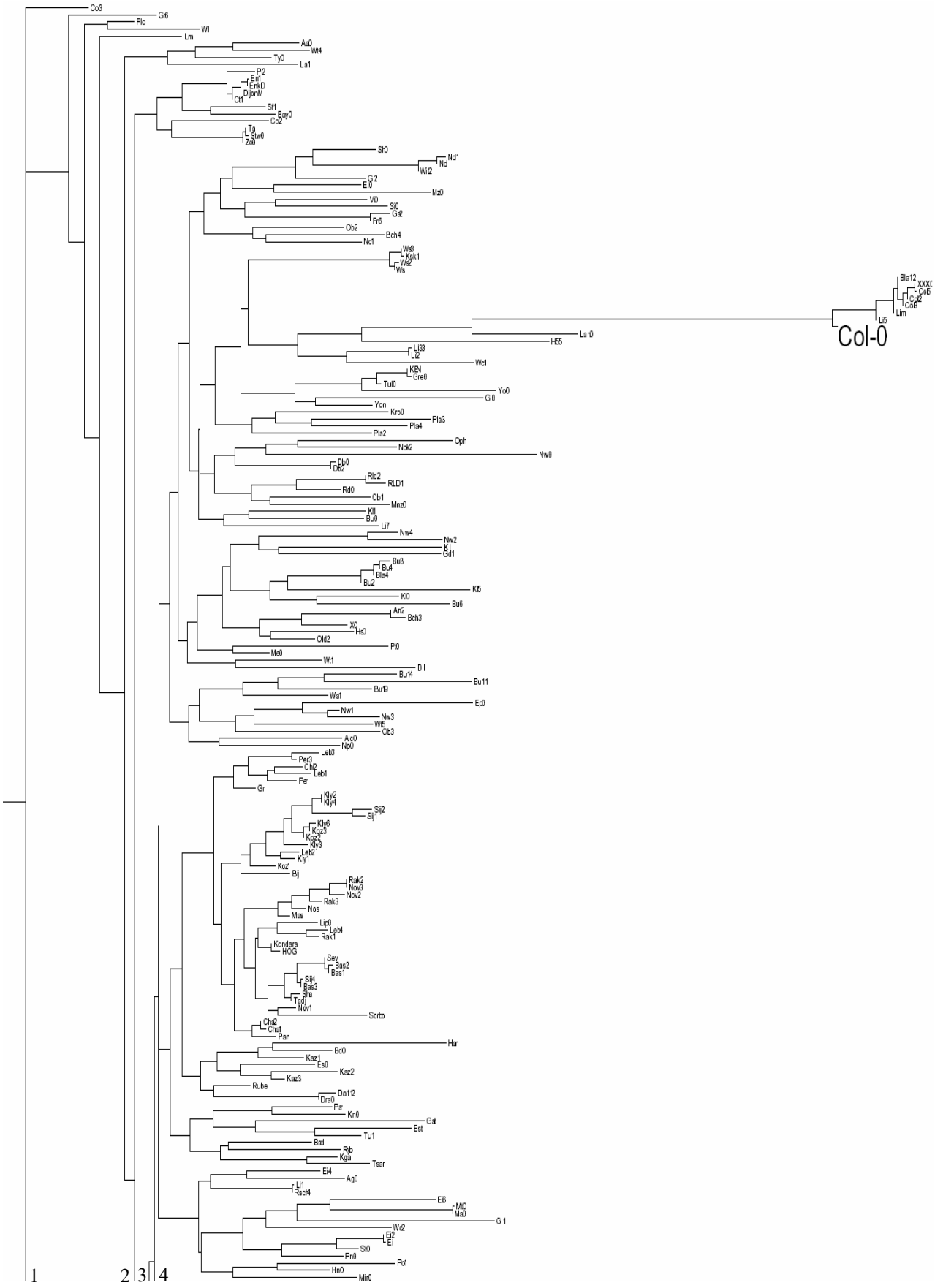


Fig. 2. Dendrogram of 335 accessions based on the neighbor-joining clustering of a pairwise distance matrix. The tree was rooted at the midpoint.

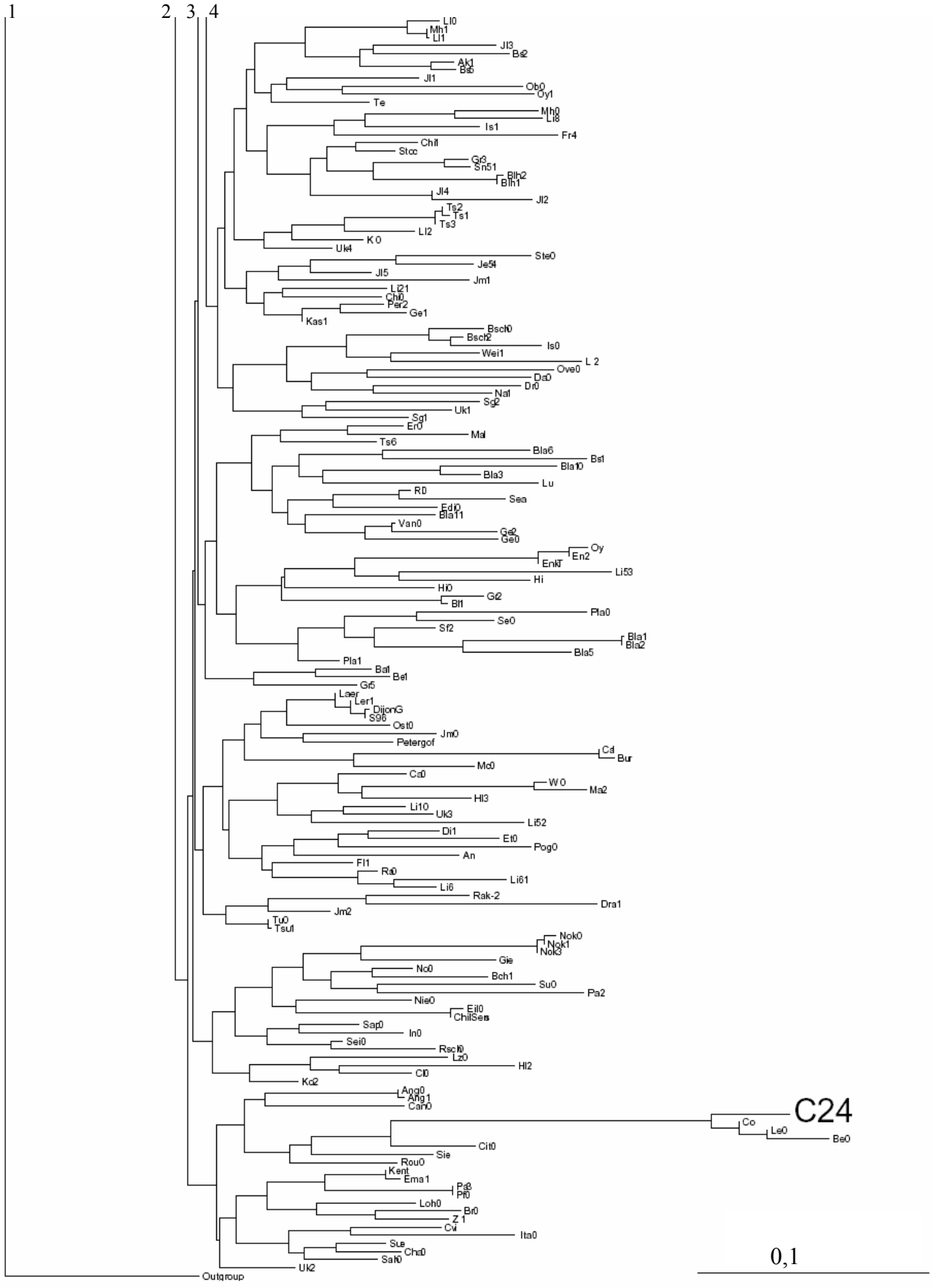


Fig. 2. (continued) Dendrogram of 335 accessions based on the neighbor-joining clustering of a pairwise distance matrix. The tree was rooted at the midpoint.

Population structure

To investigate the extent of population structure, we first compared allele frequencies among accessions from different geographic regions (Table 2). Within Eurasia, gene diversity was lowest in Central Asian accessions (H=0,095) and highest in Iberian accessions (H=0,252). Differences in gene diversity between Eurasian geographic regions are significant (F=12,06, d.f.=6, P<0,0001).

Table 2. Genetic diversity in different geographic regions.

Region	n	*H* (SE)	Proportion of missing data (SE)	Proportion of ancestral alleles (SE)
America	11	0,329 (0,018)	0,097 (0,051)	0,656 (0,053)
Africa	4	0,206 (0,026)	0,067 (0,029)	0,735 (0,036)
Iberian Peninsula	32	0,252 (0,015)	0,092 (0,005)	0,673 (0,021)
British Isles	11	0,232 (0,020)	0,087 (0,011)	0,691 (0,017)
Central Europe	188	0,245 (0,016)	0,083 (0,002)	0,686 (0,005)
Southern Italy	9	0,221 (0,020)	0,102 (0,014)	0,719 (0,017)
Scandinavia	10	0,221 (0,019)	0,091 (0,010)	0,706 (0,020)
Eastern Europe	22	0,185 (0,019)	0,102 (0,006)	0,684 (0,010)
Central Asia	40	0,095 (0,015)	0,080 (0,005)	0,754 (0,005)
Eastern Asia	1	-	0,043 (-)	0,765 (-)
Unknown Origin	7	0,385 (0,015)	0,096 (0,056)	0,666 (0,065)
Total	335	0,245 (0,015)	0,086 (0,002)	0,69 (0,005)

In pairwise t-tests between regions, all comparisons that include the Central Asian accessions are significant after sequential Bonferroni correction (not shown), but none of the other comparisons are. When Central Asian accessions are removed from the one-way ANOVA, the test is only marginally significant (F=1,98, d.f.=5, P=0,081). There is no significant difference in the proportion of missing data between geographic regions (F=1,86, d.f.=6, P=0,111). Thus, estimates of genetic diversity should not be biased due to different levels of missing data.

We examined the presence of isolation by distance in a Mantel test in which only the Eurasian accessions were included (n=308). The overall Mantel correlation of the genetic and geographical distances was negative (r=-0,0645; P=0,035). Therefore, the genetic distance decreases with increasing geographical distance, suggesting no isolation by distance in the entire data set. However, significant positive correlations exist for all comparisons between Central Asia and all other regions except Central Europe, and for the comparisons between the Iberian Peninsula with Central Europe and the British Isles, respectively (Table 3). Fig. 3 shows the presence of significant isolation by distance among

Table 3. Mantel test results r (*P*-values) for correlations between different geographic regions (999 iterations, *n* = 114 biallelic characters included). Boldface values are significant at a 5 % level after sequential Bonferroni correction for multiple tests.

Comparison	*n*	Mantel´s *r* (*P*)
All samples together	308	-0,0645 (0,035)
Central Europe	175	0,0366 (0,202)
Central Europe-Asia	214	-0,0607 (0,083)
Central Europe-Iberian Peninsula	205	0,0938 (0,049)
Central Europe-Southern Italy	178	0,0271 (0,273)
Central Europe-British Isles	186	0,0365 (0,203)
Central Europe-Scandinavia	182	-0,0039 (0,536)
Central Europe-Eastern Europe	218	-0,1070 (0,003)
Asia-Iberian Peninsula	69	0,5143 (0,001)
Asia-Southern Italy	42	0,7392 (0,001)
Asia-British Isles	50	0,7268 (0,001)
Asia-Scandinavia	46	0,7069 (0,001)
Asia-Eastern Europe	82	0,4668 (0,001)
Asia	39	0,3487 (0,001)
Iberian Peninsula-Southern Italy	33	0,2579 (0,060)
Iberian Peninsula-British Isles	41	0,2937 (0,005)
Iberian Peninsula-Scandinavia	37	0,1828 (0,094)
Iberian Peninsula-Eastern Europe	73	0,3100 (0,067)
Iberian Peninsula	30	0,1282 (0,053)
Southern Italy-British Isles	14	0,0195 (0,418)
Southern Italy-Scandinavia	10	0,2436 (0,056)
Southern Italy-Eastern Europe	46	-0,0093 (0,450)
Southern Italy	3	-0,9968 (n/a[a])
British Isles-Scandinavia	18	0,0056 (0,449)
British Isles-Eastern Europe	54	0,1091 (0,073)
British Isles	11	-0,0494 (0,430)
Scandinavia-Eastern Europe	50	-0,0442 (0,304)
Scandinavia	7	0,0523 (0,395)
Eastern Europe	43	-0,0618 (0,202)

[a] A *P* -value cannot be calculated for less than 4 values.

Iberian (R^2=0,177, *P*<0,0001) and Central Asian accessions (R^2=0,323, *P*<0,0001). Among Central Asian accessions, genetic and geographic distance are positively correlated for accessions within a range of 1,000 km, but negatively correlated for larger distances indicating the presence of genetically similar accessions in geographically distant Central Asian locations such as Siberia and Uzbekistan.

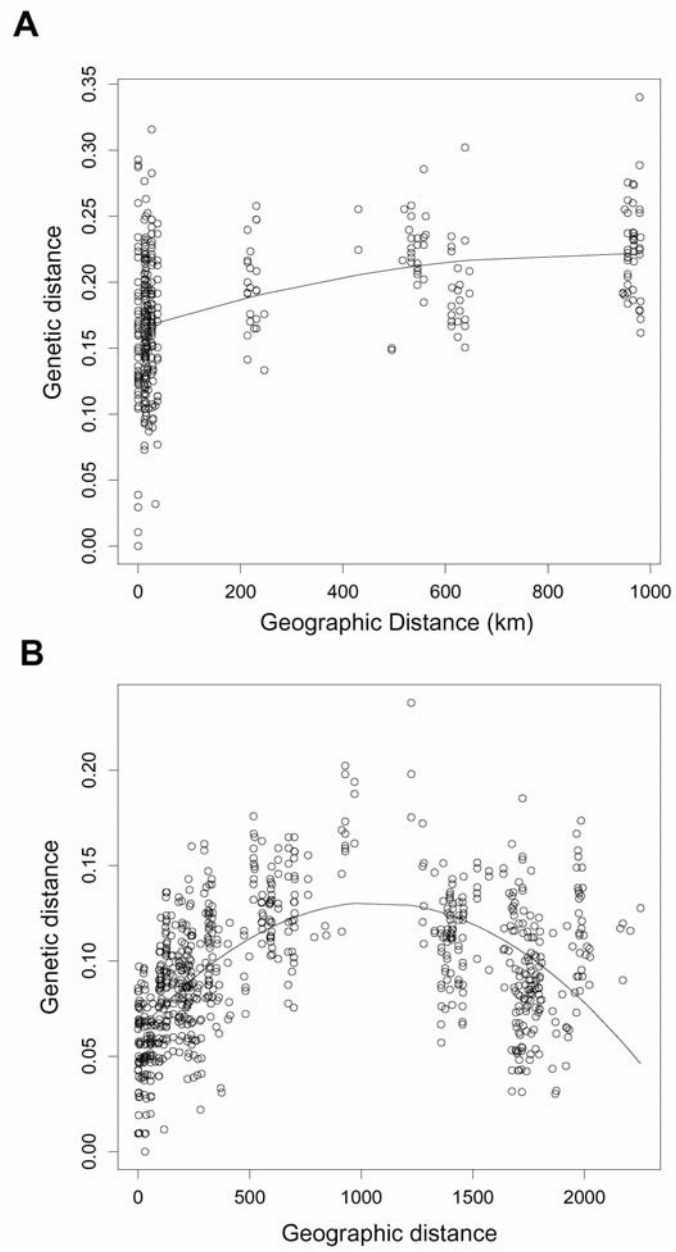


Fig. 3. Correlation between geographical and genetic distance indicates isolation by distance in *Arabidopsis thaliana* on the Iberian Peninsula (A) and in Central Asia (B). The line corresponds to the fitted curve of a quadratic regression.

To further characterize geographic patterns of genetic diversity, we determined the degree of homogeneity among SNP markers within the Eurasian distribution area (Fig. 4). Homogeneity is defined as the number of invariant SNPs among accessions from a geographic region (see Materials and Methods). Two spatial gradients with different levels of homogeneity were identified (Fig. 4). One gradient stretches from Central Asia (high degree of homogeneity) to Central Europe (little homogeneity) and the other from Africa

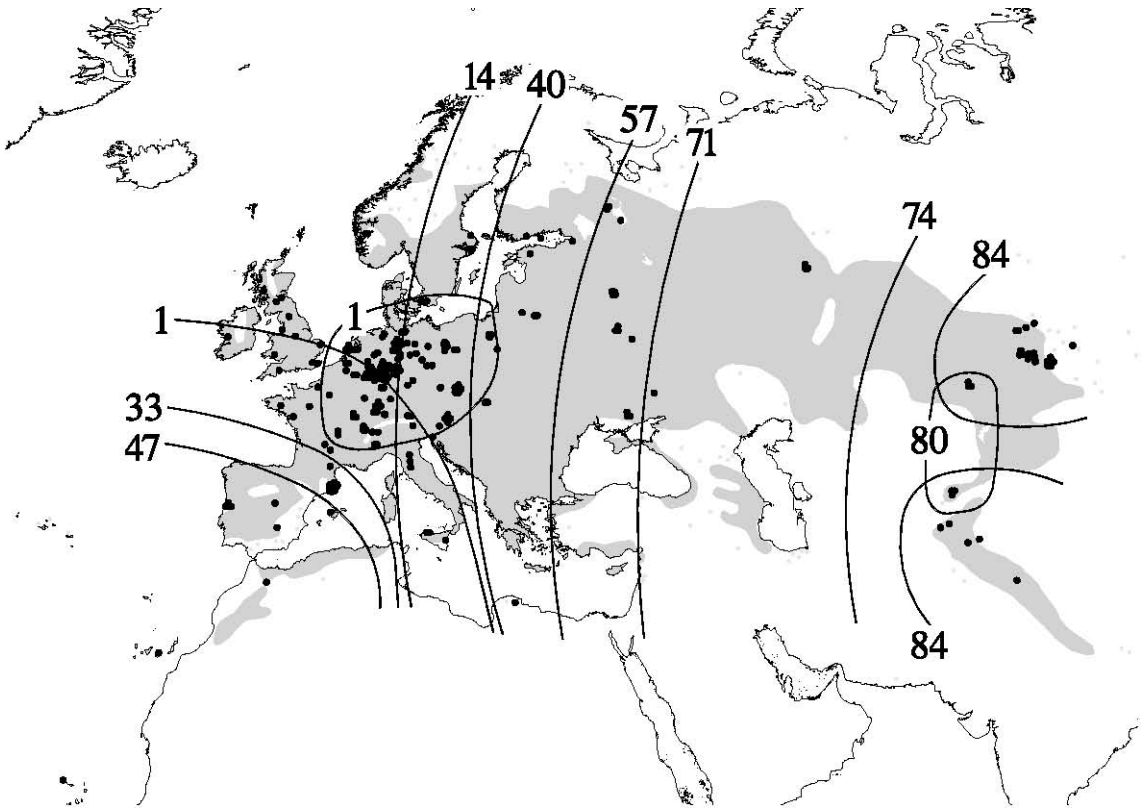


Fig. 4. Distribution of homogeneity of 115 SNPs markers among *A. thaliana* accessions across the Eurasian range. The composite map shows the locations of the accessions included in the analyses and the isolines show the absolute number of SNPs that are invariant (i.e., not polymorphic) in a geographic region.

and the Iberian Peninsula (high homogeneity) to Central Europe. Among central European accessions, the degree of homogeneity is very low with only one invariant SNP marker.

We estimated the number of populations, *K*, that is most consistent with our data according to the model-based approach of Pritchard et al. (2000). This is achieved by comparing the likelihood values of simulations with different numbers of populations used in the simulations. Likelihood values for *K*, when *K* ranges from 1 to 30, increase until *K*=15; at that point they remain the same up to *K*=30. The value of *K* with the highest likelihood should reflect biological reality (presuming the model is true); in our case this would be *K*=15. Every accession is proportionally assigned to one of the inferred populations, and this information can be used to differentiate between original and admixed populations (Fig. 5 and Fig. 6). In our data, the accessions originating from the Iberian Peninsula and from Central Asia form distinct populations, whereas accessions from Central and Eastern Europe appear to belong to admixed populations. It is important to note that some of the assumptions of the model-based inference are violated in a selfing species, which may lead to an overestimation of the number of populations.

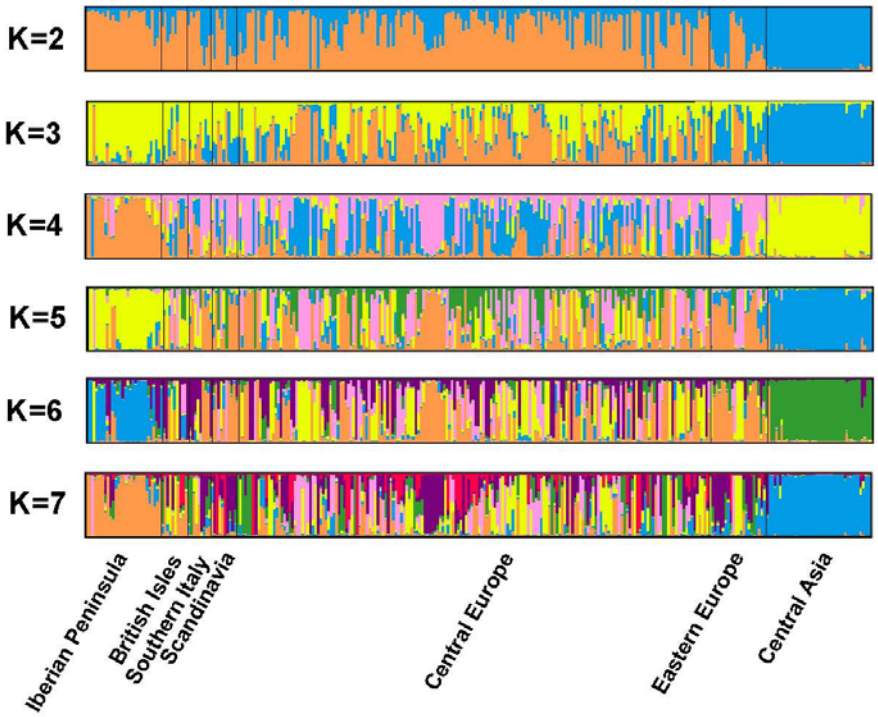


Fig. 5. Model-based inference of population structure of *A. thaliana*. Every individual is represented by a vertical line and the relative proportion of the estimated membership to the different populations is indicated by the color-coding of the vertical bar. *K* corresponds to the number of populations that are used in the estimation process. The figure indicates that the assignment of accessions from the Iberian Peninsula and from Central Asia into distinct populations is robust with different values of *K*. The figure was produced with the *distruct* program (http://www-hto.usc.edu/ ~noahr/distruct.html).

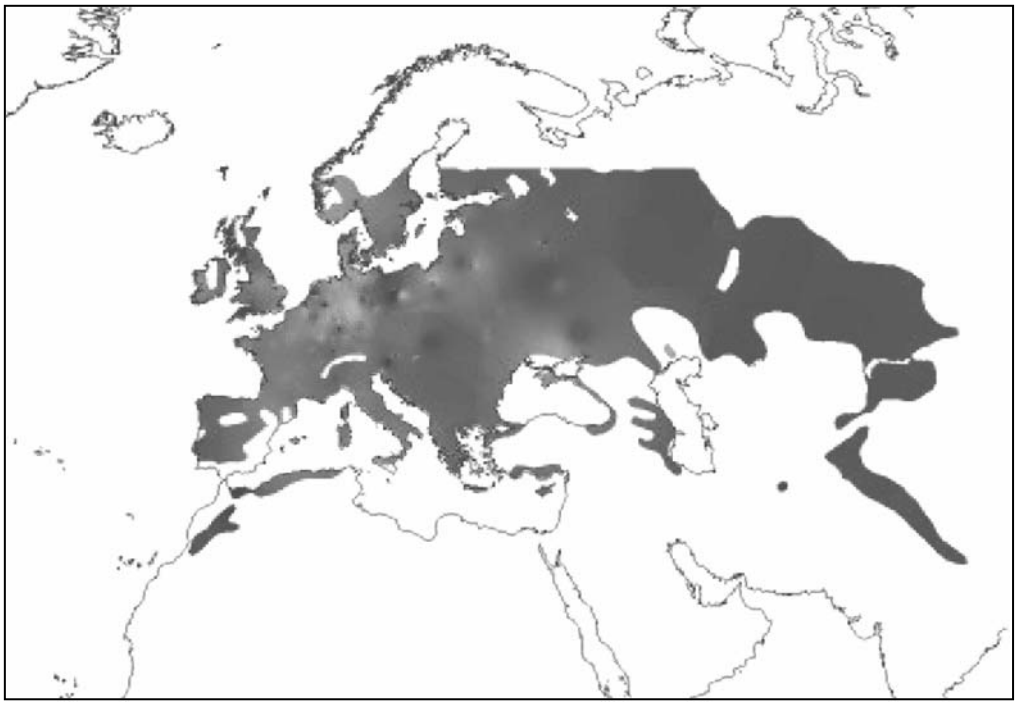


Fig. 6. Geographic pattern of model-based assignment of *A. thaliana* accessions to different populations. Population structure was inferred with *K*=3 using the *structure* program. The colors represent the assignment of accessions to each of the three populations, which was geographically extrapolated using the IDW function of the Arc/Info geographical information system.

Outgroup comparisons

Using outgroup information, the evolutionary state of 49 of the 115 SNPs (43%) could be determined. The mean relative frequency of the ancestral allele among 335 accessions is 0,69 and 0,31 of the derived allele. Ancestral alleles are significantly more frequent than derived alleles (Pairwise Wilcoxon signed-rank test, V=635, P=0,0004). By combining all outgroup alleles into a genome-wide, ancestral genotype, we determined the degree of similarity of every *A. thaliana* accession in our dataset to a hypothetical outgroup genotype (9. Supplementary Data, Fig. 2). The accession most similar to the outgroup haplotype is *Coimbra-3*, with 41 out of 47 SNPs (87%) showing the ancestral state; the least similar accession is *Columbia-5*, with 14 out of 48 SNPs (29%) in the ancestral state. The *C24* accession has twice as many ancestral alleles (70%) as *Col-0* (35%). We also observed differences in the relative frequency of ancestral alleles among geographic regions (Table 2). The Central Asian accessions have a higher proportion of ancestral alleles than accessions from any other region within the sampling range (one-way ANOVA, F=16,325, d.f.=6, P<0,0001). Pairwise comparisons between geographic regions show that Central Asian accessions have different proportions of ancestral alleles compared to Central European (P<0,0001), Eastern European (P<0,0001) and Iberian accessions (P=0,014) after sequential Bonferroni correction. This result remains significant when the five accessions with a high proportion of derived alleles are removed (not shown). When the Central Asian accessions are removed, the ANOVA is not significant anymore (F=0,878, d.f.=5, P=0,507). The geographic distribution of accessions with a high proportion of ancestral alleles shows that these accessions tend to occur in the putative glacial refugia; the accessions with many derived alleles occur in the putative postglacial colonization areas (Fig. 7).

Identification of a core set of accessions

For some purposes it is desirable to have a core collection that harbors a large proportion of the segregating genetic variation in a small number of accessions. Using an algorithm that maximizes gene diversity, H, for a given number of accessions, genetic diversity decreased almost linearly with an increasing number of accessions in the collection (results not shown). There was no obvious set of accessions that would define such a core collection but gene diversity of accessions identified by our method was always higher than of a randomly chosen set of the same size. However, a bootstrap analysis indicates

that the composition of a core collection depends heavily on the markers used, because on average less than 5% of accessions included are found in the bootstrapped set.

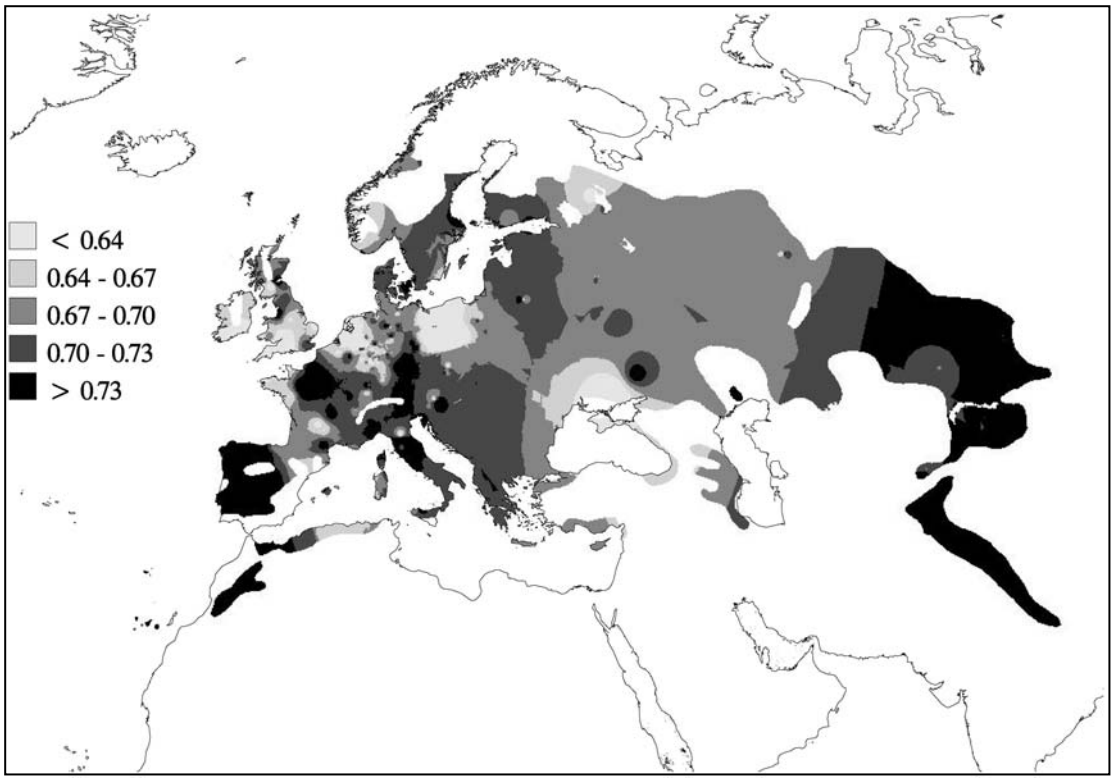


Fig. 7. Geographic distribution of the proportion of ancestral SNP alleles among the Eurasian distribution range of *A. thaliana*. The numbers indicate the proportions of ancestral alleles among up to 49 SNPs for which the ancestral allele could be determined by outgroup comparison.

Discussion

Quality control of genotyping

The markers used for this study were identified and genotyped using automated high-throughput genomics approaches. In such large-scale genotyping projects, the quality of the data is an important issue. Recent studies have suggested that no genotyping method is 100% accurate and that up to 5% of individual genotypes could be miscalled (Bray et al., 2001). To account for this potential of error, we performed two different checks. The error rate we found of less than 1% agrees with the error rate of 0,4% observed upon establishment of the marker system (Törjék et al., 2003). The overall percentage of missing data was 8,63. Although we did not further investigate the cause of missing genotypes, we believe that a considerable proportion of missing data results from unrecognized polymorphisms in the primer binding sites, because the marker set was identified by the

pairwise comparison of *Col-0* and *C24* accessions. Using the same markers for MALDI-ToF genotyping of 130 BC3F1 plants of a mapping population derived from these two accessions, only 2,93% of data was (O. T. and T. A., unpublished observation). However, since the proportions of missing data are not different among geographic regions (Table 2), they should not affect the inference of population structure. A small number of individuals have more than 5% heterozygous genotypes. Possible reasons this observation could be contamination of seed stocks or DNA samples, the duplication of loci or recent outcrossing in the natural environment. In the latter case, heterozygous accessions may represent local populations that were recently subject to admixture or have higher outcrossing rates and are worth to be studied in greater detail.

Possible effects of sampling and ascertainment bias

Our data suggest the existence of a large-scale population structure and isolation by distance throughout the Eurasian distribution range. Before discussing the biological significance of these results, it is important to consider the potential effects of sampling and marker ascertainment bias. Sampling bias is the unequal sampling of different geographic regions and local populations. Most accessions currently available from the *Arabidopsis* stock centers were collected by a number of different researchers and do not represent a carefully designed sampling scheme. Central Europe has been extensively sampled, the Iberian Peninsula and Central Asia to a lesser degree; however, the Western part of Russia and potential glacial refugia such as Italy and the Balkans are not well represented in current collections (Fig. 4). Nevertheless, despite the different numbers of accessions representing the geographic regions, levels of gene diversity in Europe are not correlated with sample size (Table 2) and thus sampling bias does not seem to be a major confounding factor. Ascertainment bias is a more serious concern, because the SNP markers used in this study were selected to differentiate between two accessions of European origin (*Col-0* and *C24*) and were used to analyse non-European populations, which may lead to incorrect estimates. For example, estimates of genetic diversity of northern East Asian human populations inferred from genotyping Y-chromosomal SNPs were significantly lower with SNP markers identified in southern East Asian populations (Su et al., 1999) than with SNP markers identified in a global human sample (Karafet et al., 2001). It is possible to correct for ascertainment bias in estimates of LD (Nielsen and Signorovich, 2003) or genome-wide recombination rates (Clark et al., 2003), but it is

difficult to discern how such a correction can be achieved for estimates of genetic diversity in other populations. Since we cannot observe SNP markers that are polymorphic in Central Asian but fixed in European accessions, the observed gene diversity in the Central Asian accessions may be an underestimate. However, despite the ascertainment bias of SNP markers, the same pattern of geographic isolation by distance in the Central Asian population was observed by Sharbel et al. (2000), who used unbiased AFLP markers. Additionally, a sequence survey of two genomic regions on chromosome 2 using a set of European and Central Asian accessions found a lower level of diversity in the Central Asian than in the European accessions (Schmuths et al., 2004b).

Evidence for population structure in A. thaliana

Levels of pairwise LD among SNP markers are low and most markers are effectively unlinked from each other, although some evidence exists for the presence of genome-wide LD. The combined genotypes of all 115 SNPmarkers constitute a genomic „fingerprint" that can be used to uniquely identify individual accessions. Furthermore, simulation studies suggest that the relatively small number of SNPs used in this study is sufficient to detect the presence of population structure (Turakulov and Easteal, 2003).

The previously observed lack of a large-scale population structure in *A. thaliana* was interpreted to be the result of human disturbance. Although we also find some evidence for human-induced long-distance migration (or mix-ups during the handling of seeds), the absence of a geographic pattern in previous studies may result from the small numbers of accessions and markers analyzed. In contrast, the present SNP analysis provides strong evidence of a large-scale population structure as indicated by different levels of genetic diversity among geographic regions (Table 2), the geographic distributions of SNP homogeneity (Fig. 4), the existence of isolation by distance (Table 3), and the model-based inference of geographic structure (Fig. 5). Geographic patterns of SNP variation suggest that the current distribution of genetic diversity of *A. thaliana* was mainly shaped by climatic and vegetational changes during and after Pleistocene glaciation. A popular model assumes that subpopulations of a species were separated in different glacial refugia during cold periods (Hewitt, 1999). In Eurasia, potential refugia include the Iberian Peninsula, Italy, the Balkans, and portions of Central Asia (Comes and Kadereit, 1998). The occupation of these refugia over long time spans (possibly several glaciation cycles) may have contributed to the genetic differentiation between refugia either by genetic drift or

local adaptive evolution. After the end of the last cold period, those regions of Europe which could not sustain *A. thaliana* populations were then rapidly recolonized by individuals originating from the different refugia. Genetic analysis of various plant and animal species indicates that the recolonization of Central Europe from refugia is a general pattern but that species differ with respect to refugia used and recolonization routes taken (Taberlet et al., 1998; Hewitt, 1999). Our data support such a model for *A. thaliana* with the Iberian Peninsula and Central Asia as glacial refugia and a postglacial recolonization of Central and Eastern Europe by migrants from these refugia. The accessions from the Iberian Peninsula and Central Asia belong to distinct and differentiated populations, whereas accessions from Central and Eastern Europe appear to constitute recently admixed populations. Both the geographic patterns of SNP homogeneity and of the model-based inference of the population structure suggest that the admixture zone includes the northern part of Central Europe and Eastern Europe, which corresponds to the location of postglacial „suture zones" observed in other species (Hewitt, 1999). Furthermore, the proportion of ancestral alleles is lower in recolonized regions than in putative refugia (Fig. 7), which may reflect either the fixation of new alleles by genetic drift or positive selection in migrating populations. The geographic pattern of the model-based inference with $K = 3$ indicates that the southern part of Central Europe was recolonized from the Iberian refugium and Eastern Europe from the Central Asian refugium (Fig. 6). Interestingly, the area of the third inferred population represents roughly the maximal expansion of the ice sheets and tundra vegetation during the last glacial cold period, about 18,000 years ago (Frenzel et al., 1992), suggesting that there was little if any gene flow back into the refugial areas - a theory, which is consistent with genetic models (Hewitt, 1996). Levels of outbreeding appear to have been high enough to reshuffle genetic variation between migrants from previously separated refugia. Such a scenario is supported by the star-like phylogeny of European accessions observed in this and in other studies (Breyne et al., 1999; Miyashita et al., 1999; Sharbel et al., 2000) and it also explains the difficulty in defining a distinct core collection of accessions.

There are, however, several caveats that require further investigation. First, although it appears that most of Central Europe was recolonized from the Iberian refugium, we were not able to investigate the importance of the refugia in Italy and the Balkans, because few if any accessions have been collected from these areas. This is a major limitation because in many animal and plant species, these regions were important sources of genomes during the recolonization of Europe (Willis, 1996; Comes and Kadereit, 1998; Hewitt, 1999).

Second, the decrease in SNP homogeneity along the putative recolonization routes appears to be in contrast to genetic models of dispersal, which propose that genetic diversity should be high in the glacial refugium and decrease due to genetic drift and bottlenecks in migrating populations (Hewitt, 1996). We do not observe such a pattern, as gene diversity (*H*) is nearly as high in Central Europe as on the Iberian Peninsula and significantly higher than in Central Asia. The high diversity in the suture zone may be the result of the admixture of individuals from different refugial populations. Another explanation may be found in the colonization history of local habitats. If *A. thaliana* is able to generate large populations from single arriving seeds, parallel or subsequent colonization events of locations by different „stocks" can increase the genetic heterogeneity of local populations. This has been observed in different genes (e.g., Purugganan and Suddith, 1998, 1999) for which accessions from the Brittany (France) occur in different clades of the phylogenetic tree. The high selfing rate and the rather low frequency of flower visitation by insects (Hoffmann et al., 2003a) may explain why in some populations no genetic homogenisation occurs even over a longer time period. Finally, several lines of evidence suggest that Central Asia was not a Pleistocene refugium as proposed by Sharbel et al. (2000) but instead, was colonized in postglacial times from a single Western refugium (i.e., the Balkans). We observe isolation by distance and a low level of gene diversity in Central Asia, which may represent the loss of genetic variation due to genetic drift along recolonization routes that was not compensated by the admixture with individuals from a different refugial population as was likely the case in Central Europe. A recent colonization of Central Asia by *A. thaliana* is furthermore supported by the presence of multiple, genetically similar accessions in very distant (>1000 km) locations of Central Asia (Fig. 3). This pattern may reflect the recolonization of the northern and southern regions of Central Asia from a relatively small pool of migrants from a Western refugium that may have split during the colonization of these areas and have maintained their genetic similarity until present times. On the other hand, there is also support for the existence of a Central Asian refugium. Paleovegetation data indicate that the environmental conditions of this region during the last glaciation period were similar to the Mediterranean refugia, although temperatures were lower in Central Asia (Frenzel et al., 1992). Furthermore, the proportion of ancestral alleles in Central Asian accessions is high (Fig. 7) and similar to the Iberian accessions. The low heterozygosity of Central Asian accessions may result from ascertainment bias as discussed above.

The present study and the previous work by Sharbel et al. (2000) underline the importance of a thorough understanding of geographical and historical patterns of genetic variation, if it is utilized in genetic and functional studies (Alonso-Blanco and Koornneef, 2000). Future work on the population structure of *A. thaliana* should include the characterization of genetic variation in putative refugial areas such as Italy and the Balkans. Since there are currently only few accessions available from these regions, the systematic sampling and subsequent genetic analysis of accessions from these regions will be crucial.

Acknowledgements

This work was funded by the BMBF as part of the GABI project (#0312275A) to T. A. and by the Emmy-Noether program of the DFG (Schm 1354-2/2) to K. J. S. We are grateful to Thomas Heinze for assistance with programming and to Maik Zehnsdorf and Melanie Lück for excellent technical assistance. We also thank K. Bachmann, M. Clauss, B. Haubold, M. Koornneef, A. Lawton-Rauh, T. Mitchell-Olds, S. Ramos Onsins and E. Wheeler for discussions and comments on the manuscript.

4. Schlussfolgerungen

Innerhalb dieser Arbeit wurden zum ersten Mal nicht anonyme Marker für die Korrelation von genetischer Variabilität und geographischer Herkunft bei *A. thaliana* gefunden. Bisher hatten Sharbel et al. (2000) mit anonymen aber genomweiten und nicht im Voraus an einem Testset erfassten und für die Analyse ausgewählten AFLPs Hinweise für eine genetische Isolation durch geographische Distanz erhalten. Eine Metadatenanalyse von Hoffmann et al. (2003b) mit allen zu dieser Zeit verfügbaren Sequenzdaten, die hauptsächlich an europäischen Akzessionen erfasst wurden, ließ ebenfalls eine Korrelation zwischen DNA-Polymorphismen und Geographie vermuten.

Die in dieser Arbeit detektierten SNPs (Artikel 1 und 4) lassen den Schluß zu, dass es zwei deutliche Gruppen von Haplotypen innerhalb von *A. thaliana* gibt, die durch einen jeweiligen Consensushaplotyp charakterisiert werden können. Diese zwei Gruppen sind nicht wie angenommen im europäischen Teil des Areals zu finden (Aguadé, 2001), sondern in Asien und Europa. Die Überlappungs(hybrid)zone der zwei Gruppen liegt, wie schon Sharbel et al. (2000) vermuteten, in Osteuropa (~15° Länge). Die Auswertung der Polymorphismen lässt weiterhin den Rückschluß zu, dass es zwei getrennte glaziale Refugien gab: auf der Iberischen Halbinsel/Italien und in Asien. Das steht im Einklang mit vorher hypothesierten Refugien anderer Arten (Hewitt, 1999).

Die dreizehn, für den europäischen und asiatischen Consensushaplotyp „diagnostischen" Basenzustände (Artikel 1), haben im arealabdeckenden Akzessionsset eine annähernd ausgeglichene Frequenz. Diese ausgeglichene Basenfrequenz im repräsentativen Datensatz kann historisch durch alte Polymorphismen im Stammbaum entstanden (Konnert und Bergmann, 1995) oder durch Selektion am natürlichen Standort bedingt sein (Akashi, 1999). Die darauf basierenden zwei großen Haplotypengruppen innerhalb von *A. thaliana* (Artikel 1) lassen sich innerhalb des gesamten Genoms wiederfinden (Artikel 4) und scheinen miteinander zu rekombinieren (Artikel 1). Innerhalb der europäischen Gruppe gibt es eine Vielzahl von Mutationen. Eine genauere geographische Assoziation der SNPs innerhalb der europäischen und asiatischen Gruppe ist nicht möglich, da vor allem in Europa das natürliche Verbreitungsgebiet stark anthropogen beeinflußt ist.

In Artikel 4 sind die Auswirkung der Auswahl von bekannten Polymorphismen aus einem begrenzten Ausgangsset auf die Ergebnisse zu erkennen. Die verwendeten SNPs wurden an zwei (*Col* und C*24*) bzw. drei (*Col*, *C24* und L*er*) jeweils europäischen Akzessionen erfasst. Dadurch wurde gegen mögliche „asiatische Allele" selektiert. Aussagen über die

Gesamtpopulation können jedoch nur anhand jeweils repräsentativer Stichproben, sowohl des Pflanzenmaterials, als auch der genetischen Marker gemacht werden. Durch diese systematische Einschränkung der untersuchten SNPs im Versuchsansatz kann möglicherweise die große Einheitlichkeit der untersuchten mittelasiatischen Akzessionen erklärt werden.

Andererseits zeigten 29 Akzessionen aus Usbekistan, Tadjikistan, Russland und Polen auch in einem durch Sequenzierung gewonnenen Datensatz (Artikel 1) keinerlei Variation. Price et al. (1994) stellten aufgrund von vergleichenden morphologischen Untersuchungen und geographischen Auswertungen von Fundorten anderer Brassicaceae die Hypothese auf, dass das Diversitätszentrum von *A. thaliana*, ähnlich dem verwandter Arten, in Mittelasien liegt. Die vorliegenden Ergebnisse lassen jedoch die Frage aufkommen, ob Mittelasien tatsächlich das Diversitätszentrum von *A. thaliana* ist und die Pflanzen dort erst kürzlich an Variation verloren haben oder ob Asien zu unrecht als Diversitätszentrum gilt.

Der Verlust an genetischer Variabilität in den mittelasiatischen Akzessionen könnte durch zufälligen Drift z.B. durch Flaschenhalseffekte (bottlenecks) entstanden sein. Bottlenecks werden durch unvorteilhafte Bedingungen, welche die Größe der Population und damit auch die genetische Variabilität über mehrere Generationen senken, hervorgerufen. Diese Senkung der Variabilität ist permanent und kann nicht durch anschließende Zunahme der Populationsgröße gesteigert werden (Falconer und Mackay, 1997). Eine weitere Möglichkeit zur Erklärung der geringen genetischen Diversität in den mittelasiatischen Akzessionen ist natürliche Selektion. Genetisch an ihre Umwelt schlecht angepasste Genotypen haben oft eine geringere Reproduktionsrate und sterben unter gleichbleibenden Umweltbedingungen somit langfristig aus.

Der Verlust der genetischen Variabilität durch Selektion scheint bei *A. thaliana* unwahrscheinlich, da Akzessionen mit hohem Anteil an asiatischen Markern auch im Mittelmeergebiet gefunden wurden (Artikel 1). Zum anderen gibt es eine Akzession (*Kaz-1*), die die gefundene Korrelation durchbricht und mit reinem europäischem SNP-Muster in Kasachstan vorkommt. Eine Erklärung dafür kann der große anthropogene Einfluß innerhalb des Areals oder eine Verwechslung der Samen bei der Vermehrung sein. Ein anderer Rückschluß ist, dass die heutige Verbreitung von *A. thaliana* in erster Linie historisch durch die eiszeitlichen Refugien entstanden ist und nur gering oder gar nicht durch Selektion und genetische Adaption an den derzeitigen Standort beeinflußt wird.

A. thaliana ist molekularbiologisch sehr gut bearbeitet, aber über innerartliche Genomgrößenunterschiede und natürlich vorkommende autotetraploide Pflanzen bzw. Populationen ist nur wenig bekannt. Bisher wurden polyploide Akzessionen nur nach induzierten Mutationen durch Röntgenstrahlung (Redéi, 1964), Colchizinbehandlung (Bouharmount, 1965) oder Transformation von Embryonen (Sangwan et al., 1991) oder Kotyledonen bzw. Wurzelzellen (Altmann et al., 1994) beobachtet. Die erstmals in dieser Arbeit gefundenen natürlichen tetraploiden Akzessionen *Stoc* (Stockholm, Schweden) und *Wa-1* (Warschau, Polen) sind wahrscheinlich aus natürlichen diploiden Akzessionen hervorgegangen. Ob sie aus diploiden Akzessionen mit großem oder kleinem Genom entstanden sind, bleibt Spekulation, da die Vorfahren nicht eindeutig bekannt sind. Die Genomgröße der tetraploiden entspricht der Summe zweier „großer" diploider Genome.

Die natürlichen tetraploiden Akzessionen *Stoc* und *Wa-1* waren in den morphologischen und phänologischen Untersuchungen (Artikel 3) normal fertil und zeigten einen normalen Phänotyp. Auch Redéi (1969) bonitierte bei induzierten tetraploiden Akzessionen einen normalen Phänotyp. Altmann et al. (1994) beobachteten bei tetraploiden Akzessionen größere Blätter und Blüten und einen zwei bis drei Wochen früheren Blühbeginn (bei einem Temperaturregime von 16 h 20°C und 8 h 17°C) im Vergleich zu den von ihnen untersuchten diploiden Akzessionen. In der vorliegenden Arbeit konnten diese Beobachtungen in Bezug auf größere Blüten bei den tetraploiden Akzessionen bestätigt werden.

Die morphologischen und phänologischen Merkmale (Artikel 3) zeigten innerhalb der untersuchten Akzessionen eine enorme Variabilität. Pigliucci (2001) stellte eine Klassifikation von Reaktionsnormen auf, um plastische Reaktionen und das Maß der Spezialisierung der Pflanzen an ihre jeweilige Umwelt zu vereinen. Dabei unterteilte er die Genotypenreaktionen in vier Gruppen:

I) den Spezialisten mit kleiner Plastizität für das beobachtete Merkmal und hoher Fitness in der spezifischen Umwelt, der speziell an ein extremes Habitat angepasst ist (z.B. alpine Bedingungen),

II) den Generalisten mit einer hohen Plastizität für ein beobachtetes Merkmal, der somit die Fähigkeit hat, verschiedenartige Umwelten zu besiedeln oder großen Veränderungen in der Umwelt zu widerstehen (z.B. Arten, die sowohl über als auch unter Wasser leben können),

III) den Spezialisten, der plastisch für das beobachtete Merkmal ist, aber adaptiert ist an ein spezielles Set von Umweltbedingungen. Die Plastizität wäre somit neutral oder nur

fälschlicherweise angenommen und der Organismus sollte weiterhin als Spezialist bezeichnet werden. Das Auftreten dieser Situation ist viel wahrscheinlicher als Situation I), und

IV) den nicht plastischen Generalisten. Das ist ein Genotyp, der trotz fehlender Plastizität in einer Vielfalt von für ihn suboptimalen Umwelten lebt und somit jeweils einen ihm möglichen und wahrscheinlich unterschiedlichen „durchschnittlichen" Phänotyp annehmen würde.

Einige untersuchte Genotypen (Akzessionen) aus Artikel 3 entsprechen nach dieser Einteilung dem nicht plastischen Generalisten. Diese Kategorie spiegelt die Idee wider, dass Evolution nicht durch das Überleben des Fittesten, sondern durch das Überleben von dem, was funktioniert, abläuft. Das würde bedeuten, Selektion ist kein omnipotenter Druck und ihre Effizienz ist abhängig von historischen Vorfällen und einer Vielzahl anderer Einflüsse, z.B. Drift, Migration und Mutation (den Boer, 1999). Das widerspricht jedoch der Theorie, dass Organismen ideal an ihre jeweiligen Umweltbedingungen adaptiert sind (Gould und Lewontin, 1979).

Andererseits werden in der Evolution auch viele neutrale Mutationen angesammelt, die keinen Einfluß auf die Ausprägung der Eigenschaften einer Pflanze haben.

Für die Zuordnung einiger Genotypen in die Gruppe IV (Pigliucci, 2001) spricht, dass keine Korrelation zwischen Plastizität und geographischer Herkunft gefunden wurde, die ausnahmslos für alle Akzessionen zutraf (Artikel 3). Zum Beispiel wurde eine Korrelation der Rosettenblätter bei 14°C und 22°C mit winterwarmen bzw. winterkalten Regionen gefunden. Die Akzessionen aus winterkalten Gebieten hatten bei 22°C weniger Rosettenblätter als bei 14°C. Die Akzessionen aus winterwarmen Gebieten hatten bei 22°C mehr Rosettenblätter als bei 14°C. Diese Korrelation wurde von zwei Akzessionen durchbrochen: *Ryb* (Rybreka, Russland) und *Es-0* (Estland) kommen aus winterkalten Gebieten und hatten bei 22°C mehr Rosettenblätter als bei 14°C. Die zwei Akzessionen zeigten keine Besonderheiten innerhalb der detektierten SNPs (Artikel 1 und 4) und hatten diagnostische SNPs entweder für die europäische (*Ryb*) oder für die mittelasiatische (*Es-0*) Haplotypengruppe.

Die vorläufige Einordnung vieler Akzessionen aufgrund der morphologischen Vergleiche zwischen 14°C und 22°C in die Gruppe des nicht plastischen Generalisten, der unter einer Vielzahl von Umweltbedingungen jeweils nur einen „durchschnittlichen" aber niemals optimalen Phänotyp annimmt, könnte eine Erklärung für den Rückgang von *A. thaliana* auf ackerbaulich genutzten Flächen sein. Konkurrenzstärkere Brassicaceae wie z.B. *Capsella*

bursa-pastoris scheinen gut adaptierte Ökotypen auszubilden, die je nach Standort eine hohe oder geringe Plastizität zeigen (Neuffer und Meyer-Walf, 1996).

In der Pflanzenzüchtung werden Genotypen bevorzugt, die stark an spezifische Umwelten adaptiert sind, jedoch oft eine geringe Plastizität aufweisen. Die dazu vorausgesetzte geringe genetische Varianz ist jedoch nur in der fertigen Sorte und nicht im Ausgangsmaterial erwünscht. Innerhalb der Sorte entstehen somit konstante und vorhersagbare Merkmale, die eine geringe Schwankungsbreite aufweisen. Das Ertragspotential der angebauten Sorten kann zu einem großen Teil ausgeschöpft werden, da die Landwirtschaft durch Maßnahmen wie z.B. Bodenbearbeitung, Düngung, Pflanzenschutzmaßnahmen und teilweise Bewässerung optimale Wachstumsverhältnisse erzeugt. Dieser Status der Homöostasie, in dem ein Organismus aktiv in einem festen Zustand gehalten wird, kann jedoch aufgrund von äußeren Einflüssen nicht immer sichergestellt werden. Wechselnde klimatische Bedingungen oder starker Krankheitsdruck können zu suboptimalen Wachstumsbedingungen führen, die durch die oft geringe Plastizität der Sorten nur teilweise abgefangen werden können. Im Erntejahr 2003 gab es in West- und Mitteleuropa hauptsächlich bedingt durch ungünstige Witterung hohe Ernteeinbußen von 37% bei Roggen, 9% bei Körnermais und 8% bei Raps im Vergleich zum Vorjahr (ZMP Jahresbericht 2003/2004; http://www.zmp.de/info/jahresbericht/getreide_prognose.asp). Eine Erhöhung der Plastizität des Ertrages in den landwirtschaftlich genutzten Sorten könnte stabilere Ernteerträge auch unter ungünstigen Umweltbedingungen garantieren.

Bisherige Untersuchungen an Plastizität an Kulturpflanzen zeigen deutlich, dass es Ressourcen für Plastizität im vorhandenen Pflanzenmaterial gibt: z.B. Lein (*Linum usitatissimum* L.) (Ernteertrag und Physiologie, Diepenbrock und Porksen, 1992); Sojabohne (*Glycine max* L.), Mais (*Zea mays* L.) und Sonnenblume (*Helianthus annus* L.) (Samen pro Pflanze und Wachstumsrate während des Samenansatzes, Vega et al., 2001); Erbse (*Pisum sativum*), Rotklee (*Trifolium pratense*), Luzerne (*Medicago sativa* L.), Gerste (*Hordeum vulgare*), Roggen (*Secale cereale*), deutsches Weidelgras (*Lolium perenne* L.), Raps (*Brassica napus olifera*) (Wurzelwachstum bei unterschiedlicher Verfügbarkeit von Kalium, Hogh-Jensen und Pedersen, 2003). Dabei bleibt zu beachten, dass in verschiedenen Umwelten unterschiedliche Gene für die Ausprägung des selben Merkmals aktiviert werden können (z.B. Battjes und Bachmann, 1994).

Veränderte Umweltbedingungen können von Pflanzen auch durch die unterschiedliche Regulation der Genaktivität umgangen werden. Dabei reagiert die Pflanze durch unterschiedliche Expression ansonsten identischer Gene z.B. mithilfe von

Transkriptionsfaktoren auf die sie umgebenden Umweltbedingungen. Dubcovsky et al. (1994) untersuchten die genetische Organisation der Reaktion auf Salzstress in dem Gras *Lophopyrum elongatum* und seinem salz-sensitiven Verwandten *Triticum aestivum*. Sie waren in der Lage, elf Gene zu identifizieren, die wenige Stunden nach einem induzierten Salzstress eine erhöhte RNA-Akkumulation hatten. Diese Gene werden im salz-sensitiven Weizen zu einer geringeren Rate transkribiert. In diesem Fall mag die Resistenz nicht auf neue Gene oder einer Mutation desselben Gens, sondern einfach durch eine unterschiedlich starke Expression der vorhandenen Gene zurückgehen. Die Reaktion auf Salz scheint außerordentlich komplex bei Pflanzen zu sein, die normalerweise nicht auf salzigen Böden wachsen, wie z.B. *A. thaliana*. Liu et al. (2001) demonstrierten, dass bei Salzstress in der Wurzel permanent ein SOS-2 Gentranskript (kodierend für eine Proteinkinase) aufreguliert wird und Quesada et al. (2000) identifizierten 17 Loci, verstreut über alle 5 Chromosomen von *A. thaliana*, die die Salztoleranz beeinflussen. Die Resistenz von Pflanzen gegenüber Salzstress ist vor allem in Regionen, in denen viel bewässert wird und somit die Gefahr der Bodenversalzung besteht, von großer Bedeutung.

Ein bereits in der landwirtschaftlichen Praxis angewendetes Beispiel für Plastizität ist der Vernalisationsbedarf bei Weizen (*T. aestivum*). Der sogenannte Wechselweizen hat einen geringen Vernalisationsbedarf (z.B. die Sorten *Thasos*, *Monsun* oder *Triso*) und kann sowohl im Herbst, als auch im Frühjahr mit nur geringen Ertragseinbußen bei einer späteren Aussaat ausgebracht werden.

Dies verdeutlicht, dass die züchterische Bearbeitung von Merkmalen für eine breitere Spanne von Umweltbedingungen zur Anpassung an die jeweiligen Produktionsbedingungen führen kann. Das Verständnis von Plastizität und Reaktionsnormen ist dabei von großer Bedeutung. Durch die Aufklärung von Reaktionsnormen an Modellpflanzen wie *A. thaliana* könnten Merkmale wie z.B. Bestockung, Salztoleranz oder Qualität an Kulturpflanzen in der Züchtung gezielter bearbeitet werden. Eine daraus resultierende Anpassung der Sorten an eine größere Spanne von Umwelten kann Ertragsverluste verringern und somit die landwirtschaftliche Produktion unabhängiger gegenüber äußeren Einflüssen machen.

Nachdem durch gezielte Maßnahmen die Pflanzengesundheit und damit die Ertragsfähigkeit der Kulturpflanzen nahezu maximiert werden kann, ist es an der Zeit, diesen durch Technik hervorgerufenen homöostatischen Zustand durch die Ausschöpfung der vorhandenen natürlichen Plastizität zu ergänzen und auszubauen, um die dauerhafte Sicherung hoher Erträge gewährleisten zu können.

5. Ausblick

Anhand der in dieser Arbeit identifizierten diagnostischen SNPs für geographische Herkunft auf Chromosom 2 können jetzt entsprechende Marker auf anderen Chromosomen gefunden werden. Damit können die postulierten Ausgangshaplotypen und ihre Rekombinationsmuster bei den verschiedenen Akzessionen festgestellt werden. Somit könnte getestet werden, ob alle denkbaren möglichen Rekombinanten in der Natur auftreten oder ob sich Selektion auf Rekombination in bestimmten Chromosomen-bereichen nachweisen lässt. Auf der Ebene einzelner Marker gibt es bereits Anzeichen für nicht-zufällig fehlende Rekombinanten (siehe unten). In einigen ausgewählten Gegenden Europas sollte einmal sehr dicht gesammelt werden, um die lokale Variabilität und ihr vielleicht adaptives Verbreitungsmuster festzustellen. Eine entsprechende Studie könnte auch zu einer genaueren Lokalisierung der geographischen Durchmischungszone der zwei Haplotypen-gruppen führen und weitere Rückschlüsse auf die eiszeitlichen Refugien und die derzeitige Adaptation der Akzessionen am natürlichen Standort zulassen.

Die in der Durchflußzytometrie gemessenen intraspezifischen Genomgrößenunterschiede lassen die Frage aufkommen, welche genomischen Bereiche bei den diploiden Akzessionen mit großem Genom in Bezug zu der bisher sequenzierten Akzession *Columbia*, die ein kleines diploides Genom besitzt, variieren. Vorstellbar wären entweder kodierende Regionen oder nicht kodierende Bereiche, z.B. in den Centromerregionen. Die Tatsache, dass Genomgrößenunterschiede bei derart kleinen Genomen jetzt verlässlich festgestellt werden können, erlaubt es, Hypothesen zur Genomgrößenevolution am Modellobjekt *Arabidopsis* zu untersuchen.

Für eine genauere Untersuchung der Morphologie und Phänologie der bisher untersuchten Akzessionen wurden drei weitere Temperaturversuche (10°C, 18°C, 26°C) innerhalb des DFG-Projektes durchgeführt. Assoziationsstudien mit den erfassten molekularbiologischen Merkmalen zeigen eine temperaturabhängige Korrelation einiger SNPs. So korrelieren z.B. mit der Anzahl der Rosettenblätter bei tiefen Temperaturen (10°C, 14°C und 18°C) andere SNPs als bei höheren Temperaturen (22°C und 26°C). Das bedeutet, dass Selektion auf hohe oder niedrige Anzahl von Rosettenblättern in den beiden Temperaturbereichen zu völlig verschiedenen Allelausprägungen führen könnte. Rekombinationen von Allelen, die sowohl bei tiefen wie bei hohen Temperaturen besonders viele oder besonders wenige

Rosettenblätter haben, können entworfen werden. Aber gerade diese Kombinationen kommen im untersuchten Material nicht vor. Daraus ergeben sich interessante Fragen über die Gen-Interaktionen und über den Grund der Abwesenheit gerade dieser Genotypen bei anderweitig allgemeiner Rekombination zwischen Genotypen in der Natur.

Für weitergehende Untersuchungen dieser und anderer Assoziationen sollten Kreuzungen zwischen Akzessionen mit unterschiedlicher Merkmalsausprägung für ein bestimmtes Merkmal gemacht werden, um die Wirkung der identifizierten Loci genauer zu untersuchen und eventuelle weitere QTL (quantitative trait loci, kosegregierende Chromosomenorte) zu entdecken. Empfehlen würden sich dabei besonders Kreuzungen zwischen den typischsten „asiatischen“ und „europäischen“ Akzessionen, um Kosegregation von diagnostischen Markern und adaptiven Merkmalen nachweisen zu können. Entsprechende Versuche, zum Beispiel zur Temperaturempfindlichkeit von Merkmalen könnten aufgrund der bisherigen Daten gezielt durchgeführt werden. Gerade die Beziehung zwischen eventuellen genetisch festgelegten adaptiven Merkmalen und der phänotypischen Plastizität bei der Anpassung der Pflanzen ist eigentlich erst in letzter Zeit als eine wichtige Frage erkannt worden, zu der die entscheidenden Versuche noch ausstehen.

6. Referenzen

6.1. *Curriculum vitae*

Name	Schmuths
Vorname	Heike
Geboren am	18.05.1978 in Magdeburg

Ausbildung

1984 – 1991	Polytechnische Oberschule Komarow, Magdeburg
1991 – 1996	Humboldt-Gymnasium, Magdeburg
1996 – 2001	Studium der Agrarwissenschaften an der Martin-Luther-Universität Halle-Wittenberg, Halle
2000 - 2001	Diplomarbeit „Züchtung auf Resistenz gegen den Erreger des Halmbruches (*Pseudocercosporella herpotrichoides*) bei Weizen“ an der Fakultät für Pflanzenzüchtung und Pflanzenschutz der Martin-Luther-Universität Halle-Wittenberg, Halle
2001 – andauernd	wissenschaftliche Mitarbeiterin und Promotionskandidatin am IPK Gatersleben in der Abteilung Taxonomie, Gatersleben

Titel der Dissertation „Genetische Variabilität und phänotypische Plastizität der Modellpflanze *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae)“

6.2. Publikationsliste

Artikel

Hoffmann M.H., **Schmuths H.** (2001) *Arabidopsis thaliana* as a tool for biosystematics: studies in molecular phylogeography. IOPB Newsletters Band 33, 17-21.

Hoffmann M.H., Glaß A.S., Tomiuk J., **Schmuths H.**, Fritsch R.M., Bachmann K. (2003) Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). Molecular Ecology 12: 1007-1019.

Schmuths H., Meister A., Horres R., Bachmann K. (2004a) Genome size variation among accessions of *Arabidopsis thaliana*. Annals of Botany 93: 1-5.

Schmuths H., Hoffmann M.H., Bachmann K. (2004b) Geographic distribution and recombination of genomic fragments on the short arm of chromosome 2 of *Arabidopsis thaliana*. Plant Biology, 6: 128-139.

Hoffmann M.H., **Schmuths H.**, Koch C., Bachmann K. Site of origin of a world-wide sample of *Arabidopsis thaliana* (L.) Heynh. influences phenological and morphological plasticity in two different temperature treatments. Acta Oecologica, submitted.

Schmid K.J., Törjék O., Mayer R., **Schmuths H.**, Hoffmann M.H., Altmann T. Evidence for pleistocene refugia and postglacial expansion of *Arabidopsis thaliana* from genome-wide SNP markers. Molecular Ecology, submitted.

Ritz C.M., **Schmuths H.**, Wissemann V. Beautiful bastards: European dogroses originated by multiple hybridisation across the genus *Rosa*. Journal of Heredity, in press.

Vorträge

Schmuths H. Methoden der SNP-Detektion. (2002) Landwirtschaftliches Kolloquium, Martin-Luther-Universität Halle-Wittenberg.

Hoffmann M.H., **Schmuths H.**, Horres R., Bachmann K. (2003) Molecular ecology and biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae).
Symposium Biodiversity, spatial structures and habitat fragmentation, Giessen.

Schmuths H., Horres R., Hoffmann M.H., Bachmann K. (2003) Continued sequence variation on chromosome 2 reveals geographic and recombinant groups in *Arabidopsis thaliana*. Biodiversity and Evolutionary Biology of the German Botanical Society (DBG), Frankfurt.

Poster

Hoffmann M.H., **Schmuths H.**, Fritsch R.M., Bachmann K. (2001) Molecular and geographical analysis of world-wide accessions of *Arabidopsis thaliana* (L.) Heynh. GPZ-Jahrestagung, Halle, Germany.

Fritsch R.M., Hoffmann M.H., **Schmuths H.** (2002) *Arabidopsis* and *Arabis* in Usbekistan – ecologically caused replacement or taxonomic tangle? 26th Plant Life of Southwest Asia Symposium, Van, Turkey.

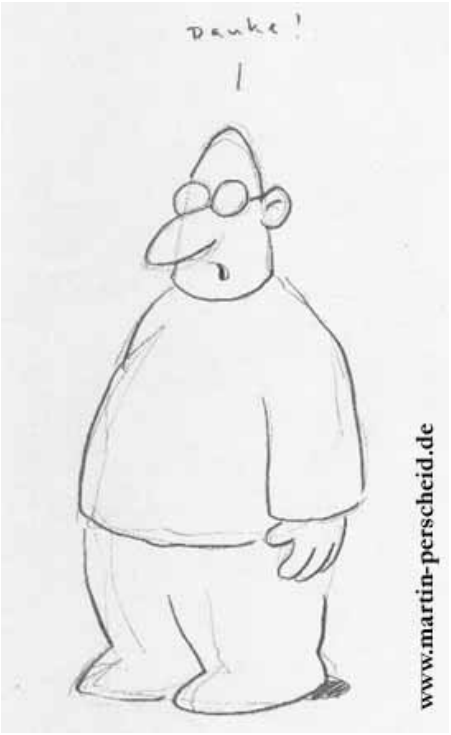
Fritsch R.M., Hoffmann M.H., Horres R., **Schmuths H.** (2002) Distribution of molecular and phenotypic characters throughout the range of *Arabidopsis thaliana* (L.) Heynh. IPK-Institutstag Gatersleben, Germany.

Hartung F., Blattner F.R., **Schmuths H.**, Chu H.H., Puchta H. (2003) Strikingly different mutation rates among the three SPO11 genes of various *A. thaliana* ecotypes. 7th International Congress of Plant Molecular Biology (ISPMB), Barcelona, Spain.

Horres R., **Schmuths H.**, Koch C., Hoffmann M.H., Fritsch R.M., Bachmann K. (2003) Strain specific phenotypic plasticity in *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). Biodiversity and Evolutionary Biology of the German Botanical Society (DBG), Frankfurt, Germany.

Schmuths H. (2003) Geographical variation in *Arabidopsis thaliana* arises from recombination. IPK-Institutstag Gatersleben, Germany.

6.3. Danksagungen



Die vorliegende Arbeit wurde in der Abteilung Taxonomie des Institutes für Pflanzengenetik und Kulturpflanzenforschung (IPK) in Gatersleben angefertigt.
Sie wäre ohne die Überlassung des Themas durch Prof. Dr. Konrad Bachmann und seine unermütliche Unterstützung und seinen Einsatz während der gesamten Dienstzeit innerhalb dieser Zeitspanne nicht möglich gewesen. Für seine Ideen und Anregungen zu meiner wissenschaftlichen Entwicklung und für seine Motivation, die immer zur rechten Zeit kam, möchte ich mich recht herzlich bedanken.

Den derzeitigen und in den letzten drei Jahren verabschiedeten Mitarbeitern der Abteilung Taxonomie danke ich für die lebhaften Diskussionen und für die vielen Hilfestellungen, die teilweise auch Schokolade, Kuchen und Tee beinhalteten.

Namentlich bedanken möchte ich mich bei Christina Koch, die bei der Bonitur der Pflanzen, der Zusammenstellung der vielen Daten und sonstigen Schandtaten eine sehr große Hilfe war.
Petra Oswald und Birgit Wohlbier haben mich im Labor tatkräftig unterstützt.

Der Deutschen Forschungsgemeinschaft (DFG) danke ich für die finanzielle Unterstützung der Arbeit.

Ursula Tiemann danke ich für die Hilfe bei der Erstellung der Abbildungen.

Mein besonderer Dank gilt Herrn Prof. Dr. W. Eberhard Weber für die Diskussionen über meine Arbeit aus pflanzenzüchterischer Sicht und den sich daraus ergebenden anregenden Gesprächen.

Ganz großer Dank geht an Nikolai Ermakov und Vassily Koval, die die Sammelreisen durch Sibirien trotz wiederkehrender Mückenplagen so angenehm wie möglich gestaltet haben.

Und wenn Du denkst es geht nicht mehr, kommt irgendwo ein Lichtlein her - meiner Familie möchte ich für die Unterstützung, die nun schon ein viertel Jahrhundert andauert, ganz besonders danken.

6.4. Eigenständigkeitserklärung

Hiermit erkläre ich, dass diese Arbeit von mir angefertigt wurde.
Es wurden keine anderen als die in der Literatur zitierten Hilfsmittel benutzt.

Halle/Saale, den 29.03.2004

Heike Schmuths

7. Literatur

Abbott R.J., Gomes M.F. (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Hederity 62: 411-418.

Aguadé M. (2001) Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes, in *Arabidopsis thaliana*. Molecular Biology and Evolution 18: 1-9.

Aird P.L. (1994) Conservation for the sustainable development of forests worldwide: a compendium of concepts and terms. Forestry Chronicle 70: 666-674.

Akashi H. (1999) Within- and between-species DNA sequence variation and the 'footprint' of natural selection. Gene 238: 39-51.

Allard R., Jain S., Workman P. (1968) The genetics of inbreeding populations. Advances in Genetics 14: 55-131.

Alonso-Blanco C., Koornneef M. (2000) Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. Trends in Plant Science 5: 22-29.

Al-Shebaz I.A., O´Kane L., Price R.A. (1999) Generic placement of species excluded from *Arabidopsis* (Brassicaceae). Novon 9: 296-307.

Altmann T., Damm B., Frommer W.B., Martin T., Morris P.C., Schweizer D. et al. (1994) Easy detection of ploidy level in *Arabiodpsis thaliana* plants by means of pollen measurement. Plant Cell Reports 13: 652-656.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.

Arumuganathan K., Earle E.D. (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9: 208-218.

Avise J.C. (2000) Phylogeography: the history and formation of species. Havard University Press.

Bandelt H.-J., Dress A.W.M. (1992) A canonical decomposition theory for metrics on a finite set. Advances in Mathematics 92: 47-105.

Barow M., Meister A. (2002) Lack of correlation between AT frequency and genome size in higher plants and the effect of non-randomness of base sequences on dye binding. Cytometry 47: 1-7.

Barow M., Meister A. (2003) Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. Plant, Cell and Environment 26: 571-584.

Barth S., Melchinger A., Lübberstedt T. (2002) Genetic diversity in *Arabidopsis thaliana* L. Heynh. investigated by cleaved amplified polymorphic sequence (CAPS) and inter-simple sequence repeat (ISSR) markers. Molecular Ecology 11: 495-505.

Battjes J., Bachmann K. (1994) Phenotypic plasticity of capitulum morphogenesis in *Microseris pygmaea* (Asteraceae-Lactuceae). Annals of Botany 73: 299-305.

Beemster G.T.S., De Vusser K., De Tavernier E., De Bock K., Inzé D. (2002) Variation in growth rate between *Arabidopsis* ecotypes is correlated with cell division and A-type cyclin-dependent kinase activity. Plant Physiology 129: 854-864.

Bender R., Lange S. (1999) Multiple test procedures other than Bonferroni's deserve wider use. British Medical Journal 318: 600.

Bennett M.D., Leitch I.J. (1995) Nuclear-DNA amounts in angiosperms. Annals of Botany 76: 113-176.

Bennett M.D., Leitch I.J., Price H.J., Johnston J.S. (2003) Comparisons with *Caenorhabditis* (~100 Mb) and Drosophila (~175 Mb) using flow cytometry show

genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* Genome Initiative estimate of ~125 Mb. Annals of Botany 91: 1-11.

Bennett M.D., Smith J.B. (1991) Nuclear DNA amounts in angiosperms. Philosophical Transactions of the Royal Society of London Series B – Biological Sciences 334: 309-345.

Bennett M.D. (1987) Variation in genomic form in plants and its ecological implications. New Phytologist 106: 177-200.

Bennetzen J.L. (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115: 29-36.

Bergelson J., Stahl E., Dudek S., Kreitman M. (1998) Genetic variation within and among populations of *Arabidopsis thaliana* ecotypes. Genetics 148: 1311-1323.

Berger B. (1965) The taxonomic confusion within *Arabidopsis* and allied genera. *Arabidopsis* research (ed. G. Röbbelen), pp. 19-25. Universität Göttingen, Göttingen.

Berná G., Robles P., Micol J.L. (1999) A mutational analysis of leaf morphogenesis in *Arabidopsis thaliana*. Genetics 152: 729-742.

Blázquez M., Koornneef M., Putterill J. (2001) Flowering on time: genes that regulate the floral transition. Workshop on the molecular basis of flowering time control. EMBO Reports 2: 1078-1082.

Bouharmount J. (1965) Action of X-rays on the apical meristems of diploid and polyploid *Arabidopsis*. *Arabidopsis* Information Service 2: 13-14.

Bray M., Boerwinkle E., Doris P. (2001) High-throughput mulitplex SNP genotyping with MALDI-Tof mass spectrometry: practice, problems and promise. Human Mutation 17: 296-304.

Breyne P., Rombaut D., van Gysel A., van Montagu M., Gerats T. (1999) AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes. Molecular and General Genetics 261: 627-634.

Brumfield R.T., Beerli P., Nickerson D.A., Edwards S.V. (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18: 249-156.

Caicedo A.L., Schaal B.A., Kunkel B.N. (1999) Diversity and molecular evolution of the RPS2 resistance gene in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the USA 96: 302-306.

Casgrain P., Legendre P. (1999) The R PACKAGE for multivariate and spatial analysis, version 4.0. User´s manual. Department of Biological Sciences, University of Montreal, Montreal.

Cho R., Mindrinos M., Richards S., Sapolsky R., Anderson M., Drenkard E. et al. (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature Genetics 23: 203-207.

Clark A.G., Nielsen R., Signorovitch J., Matise T.C., Glanowski S., Heil, J. et al. (2003) Linkage disequilibrium and inference of ancestral recombination in 538 single-nucleotide polymorphism clusters across the human genome. American Journal of Human Genetics 73: 285-300.

Clauss M., Cobban H., Mitchell-Olds T. (2002) Cross-species microsatellite markers for elu- cidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). Molecular Ecology 11: 591-601.

Clement M., Posada D., Crandall K.A. (2000) TCS: a computer program to estimate gene genealogies. Molecular Ecology 9: 1657-1660.

Collins F.S., Brooks L.D., Chakravarti A. (1998) A DNA polymorphism discovery resource for research on human genetic variation. Genome Research 8: 1229-1231.

Comes H., Kadereit J. (1998) The effect of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science 3: 432-438.

Comps B., Gömöry D., Letouzey J., Thiebaut B., Petit R.J. (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonisation in the European beech. Genetics 157: 389-397.

den Boer P.J. (1999) Natural selection, or the non-survival of the non-fit. Acta Biotheoretica 47: 83-97.

Diepenbrock W., Porksen N. (1992) Phenotypic plasticity in growth and yield components of linseed (*Linum usitatissimum* L.) in response to spacing and N-nutrition. Journal of Agronomy and Crop Science-Zeitschrift für Acker- und Pflanzenbau 169: 46-60.

Dolezel J., Greilhuber J., Lucretti S., Meister A., Lysak M.A., Nardi L. et al. (1998) Plant genome size estimation by flow cytometry: inter-laboratory comparison. Annals of Botany 82 (Suppl. A): 17-26.

Dorn L.A., Hammond Pyle E., Schmitt J. (2000) Plasticity to light cues and resources in *Arabidopsis thaliana*: testing for adaptive value and cost. Evolution 54: 1982-1994.

Dubcovsky J., Galvez A.F., Dvorak J. (1994) Comparison of the genetic organization of the early salt-stress-response gene system in salt-tolerant *Lophopyrum elongatum* and salt-sensitive wheat. Theoretical and Applied Genetics 87: 957-964.

Environmental Systems Research Institute R Inc (1992) Arc/Info.

Erschadi S., Haberer G., Schöninger M., Torres-Ruiz R.A. (2000) Estimating genetic diversity of *Arabidopsis thaliana* ecotypes with amplified fragment length polymorphisms (AFLP). Theoretical and Applied Genetics 100: 633-640.

ESRI – Environmental Systems Research Institute, Inc. (1992) Arc/Info™. Redlands.

Falconer D.S., Mackay T.F.C. (1997) Introduction to quantitative genetics. (4th Edition). Longman, Essex, UK.

Felsenstein J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166.

Francis D.M., Hulbert S.H., Michelmore R.W. (1990) Genome size and complexity of the obligate fungal pathogen, *Bremia lactucae*. Experimental Mycology 14: 299-309.

Frenzel B., Pécsi M., Velichko A. (1992) Atlas of paleoclimates and paleoenvironments of the northern hemisphere. Gustav Fischer, Stuttgart.

Galbraith D.W., Harkins K.R., Knapp S. (1991) Systemic endopolyploidy in *Arabidopsis thaliana*. Plant Physiology 96: 985-989.

Gould S.J., Lewontin, R.C. (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptionist programme. Proceedings of the Royal Society of London B205: 581-598.

Greilhuber J. (1998) Intraspecific variation in genome size: a critical reassessment. Annals of Botany 82 (Suppl. A): 27-35.

Gut, I.G. (2001) Automation in genotyping in single nucleotide polymorphisms. Human Mutation 17: 475-492.

Haff L.A., Smirnov I.P. (1997) Multiplex genotyping of PCR products with mass tag.labeled primers. Nucleic Acids Research 25: 3749-3750.

Hagenblad J., Nordborg M. (2002) Sequence variation and haplotype structure surrounding the flowering time locus FRI in *Arabidopsis thaliana*. Genetics 161: 289-298.

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

Hanfstingl U., Berry A., Kellogg E.A., Costa III J.T., Rüdiger W., Ausubel F.M. (1994) Haplotypic divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: roles for both balancing and directional selection? Genetics 138: 811-828.

Hardtke C., Müller J., Berleth T. (1996) Genetic similarity among *Arabidopsis thaliana* ecotypes estimated by DNA sequence comparison. Plant Molecular Biology 32: 915-922.

Haubold B., Kroymann J., Ratzka A., Mitchell-Olds T., Wiehe T. (2002) Recombination and gene conversion in a 170-kb genomic region of *Arabidopsis thaliana*. Genetics 161: 1269-1278.

Hauser M.-T., Harr B., Schlötterer C. (2001) Trichrome distribution in *Arabidopsis thaliana* and its close relative *Arabis lyrata*: molecular analysis of the candidate gene *GLABROUS1*. Molecular Biology and Evolution 18: 1754-1763.

Hein, J. (1990) Reconstructing evolution of sequences subject to recombination using parsimony. Mathematical Biosciences 98: 185-200.

Hein, J. (1993) A heuristic method to reconstruct the history of sequences subject to recombination. Journal of Molecular Evolution 36: 396-406.

Hewitt G.M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58: 247-276.

Hewitt G.M. (1999) Post-glacial re-colonization of European biota. Biological Journal of the Linnaean Society 68: 87-112.

Hill W., Robertson A. (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38: 226-331.

Hoffmann M.H. (2002) Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). Journal of Biogeography 29: 125-134.

Hoffmann M.H., Bremer M., Schneider K., Burger F., Stolle E., Moritz G. (2003a) Flower visitors in a natural population of *Arabidopsis thaliana*. Plant Biology 5: 491-494.

Hoffmann M.H., Glaß A.S., Tomiuk J., Schmuths H., Fritsch R.M., Bachmann K. (2003b) Analysis of molecular data of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) with Geographical Information Systems (GIS). Molecular Ecology 12: 1007-1019.

Hoffmann M.H., Schmuths H., Koch C., Bachmann K. Site of origin of a world-wide sample of *Arabidopsis thaliana* (L.) Heynh. influences phenological and morphological plasticity in two different temperature treatments. Acta Oecologica, submitted.

Hogh-Jensen H., Pedersen M.B. (2003) Morphological plasticity by crop plants and their potassium use efficiency. Journal of Plant Nutrition 26: 969-984.

Hoskins R.A., Phan A.C., Naeemuddin M., Mapa F.A., Ruddy D.A., Ryan J.J. et al. (2001) Single nucleotide polymorphism markers for genetic mapping in Drosophila melanogaster. Genome Research 11: 1100-1113.

Hudson R.R. , Kaplan N.L. (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111: 147-164.

Huson D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14: 68-73.

Innan H., Tajima F., Terauchi R., Miyshita N.T. (1996) Intragenic recombination in the *Adh* Locus of the wild plant *Arabidopsis thaliana*. Genetics 143: 1761-1770.

Innan H., Terauchi R., Miyashita N.T. (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 146: 1441-1452.

Jaccard J., Wan C.K. (1996) LISREL approaches to interaction effects in multiple regressions. Sage Publications. Thousands Oaks, CA.

Jack T. (2001) Relearning our ABCs: new twists on an old model. Trends in Plant Science 6: 310-316.

Jander G., Norris S., Rounsley S., Bush D., Levin I., Last R. (2002) *Arabidopsis* map-based cloning in the post-genome era. Plant Physiology 129: 440-450.

Liu J., Ishitani M., Halfter U., Kim C.-S., Zhu J.-K. (2001) The *Arabidopsis thaliana SOS2* gene encodes a protein kinase that is required for salt tolerance. Proceedings of the National American Society of USA 97: 3730-3734.

Johannsen W. (1911) The genotype concept of heredity. American Naturalist 45: 129-159.

Karafet T., Xu L., Du R., Wang W., Feng S., Wells R. et al. (2001) Paternal population history of East Asia: Sources, patterns and microevolutionary processes. American Journal of Human Genetics 69: 615-628.

Kawabe A., Miyashita N.T. (1999) DNA variation in the chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. Genetics 153: 1445-1453.

Kawabe A., Innan H., Terauchi R., Miyashita N.T. (1997) Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Molecular Biology and Evolution 14: 1303-1315.

King G., Nienhuis J., Hussey C. (1993) Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theoretical and Applied Genetics 86: 1028-1032.

Kliebenstein D.J., Figuth A., Mitchell-Olds T. (2002) Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. Genetics 161: 1685-1696.

Knippers R., Philippsen P., Schäfer K.P., Fanning E. (1990) Molekulare Genetik, 5. Aufl., Thieme Verlag, Stuttgart.

Koch M.A., Al-Shebaz I.A., Mummenhoff K. (2003) Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). Annals of the Missouri Botanical Garden 90: 151-171.

Koch M.A., Haubold B., Mitchell-Olds T. (2001b) Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *chs* sequences. American Journal of Botany 88: 534-544.

Koch M.A., Weisshaar B., Kroymann J., Haubold B., Mitchell-Olds T. (2001a) Comparative genomics and regulatory evolution: conservation and function of the *Chs* and *Apetala3* promotors. Molecular Biology and Evolution 18: 1882-1891.

Kondorosi E., Roudier F., Gendreau E. 2000. Plant cell-size control: growing by ploidy? Current Opinion in Plant Biology 3: 488-492.

Konnert M., Bergmann F. (1995) The geographical distribution of genetic variation of silver fir (*Abies alba*, Pinaceae) in relation to its migration history. Plant Systematics and Evolution 196: 19-30.

Krisai R., Greilhuber J. (1997) *Cochlearia pyrenaica* DC. das Löffelkraut, in Oberösterreich (mit Anmerkungen zur Karyologie und zur Genomgrösse). Beiträge zur Naturkunde Oberösterreichs 5: 151-160.

Kuittinen H., Mattila A., Savolainen O. (1997) Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana.* Heredity 79:144-152.

Kuittinen H., Aguadé M. (2000) Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. Genetics 155: 863-872.

Kuittinen H., Aguadé M., Charlesworth D., Haan A.D.E., Lauga B., Mitchell-Olds T. et al. (2002) Primers for 22 candidate genes for ecological adaptations in Brassicaceae. Molecular Ecolology Notes 2: 258-262.

Kuittinen H., Salguero D., Aguadé M. (2002) Parallel patterns of sequence variation within in between populations at three loci of *Arabidopsis thaliana*. Molecular Biology and Evolution 19: 2030-2034.

Kullman L. (1996) Recent cooling and recession of Norway spruce (*Picea abies* (L.) Karst.) in the forest-alpine tundra ecotone of the Swedish Scandes. Journal of Biogeography 23: 843-854.

Kwok P. (2001) Methods for genotyping single nucleotide polymorphisms. Annual Reviews of Genomics and Human Genetics 2: 235-258.

Lättig H.U.K., Brennicke A., Knoop V. (2003) Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. Plant Molecular Biology 33: 37-45.

Laibach F. (1940) Die Ursachen der Blütenbildung und das Blühhormon. Natur und Volk 70: 55-65.

Laibach F. (1951) Über sommer- und winteranuelle Rassen von *Arabidopsis thaliana* (L.) Heynh. Ein Beitrag zur Ätiologie der Blütenbildung. Beiträge zur Biologie der Pflanzen 28: 173-210.

Leutwiler L.S., Hough-Evans B.R., Meyerowitz E.M. (1984) The DNA of *Arabidopsis thaliana*. Molecular and General Genetics 194: 15-23.

Li B., Suzuki J.I., Hara T. (1998) Latitudinal variation in plant size and relative growth rate in *Arabidopsis thaliana*. Oecologia 115: 293-301.

Lin J.-Z., Morrell P.L., Clegg M.T. (2002) The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley (*Hordeum vulgare ssp. spontaneum*). Genetics 162: 2007-2015.

Lindblad-Toh K., Winchester E., Daly M.J., Wang D.G., Hirschhorn J.N., Laviolette J.P. et al. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nature Genetics 24: 381-386.

Lockhart P.J., Steel A.W.D., Waddell P.J., Penny D. (1996) Evolution of chlorophyll and bacteriophyll: the problem of invariant sites in sequence analysis. Proceedings of the National Academy of Sciences of the USA 93: 1930-1934.

Loridon K., Cournoyer B., Goubely C., Depeiges A., Picard G. (1998) Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*. Theoretical Applied Genetics 97: 591-604.

Lu H.Y., Wu H.P. (1986) Studies on the genotype-environment interaction of *Arabidopsis thaliana*. Botanical Bulletin Academia Sinica 27: 187-207.

Lukens L., Zou F., Lydiate D., Parkin I., Osborn T. (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. Genetics 164: 359-372.

Mantel N. (1967) The detection of disease clustering and gerneralized regression approach. Cancer Research 27: 209-220.

Marie D., Brown S.C. (1993) A cytometric exercise in plant DNA histograms, with 2C-values for 70 species. Biology of Cell 78: 41-51.

Mauricio R. (1998) Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. American Naturalist 151: 20-28.

Maynard Smith J. (1999) The detection and measurement of recombination from sequence data. Genetics 153: 1021-1027.

Melaragno J.E., Mehrotra B., Coleman A.W. (1993) Relationship between endoploidy and cell size in epidermal tissue of *Arabidopsis*. Plant Cell 5: 1661-1668.

Meyerowitz E.M. (1989) *Arabidopsis*, a useful weed. Cell 56: 263-269.

Mitchell-Olds T. (1996) Genetic constraints on life-history evolution: Quantitative-trait loci influencing growth and flowering in *Arabidopsis thaliana*. Evolution 50: 140-145.

Mitchell-Olds, T. (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. Trends in Ecology 16: 693-700.

Miyashita N.T., Kawabe A., Innan H. (1999) DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. Genetics 152: 1723-1731.

Myers S.R., Griffiths R.C. (2003) Bounds on the minimum number of recombination events in a sample history. Genetics 163: 375-394.

Nei M. (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Neuffer B., Meyer-Walf M. (1996) Ecotypic variation in relation to man made habitats in *Capsella*: field and trampling area. Flora 191: 49-57.

Nielsen R., Signorovich J. (2003) Correcting for ascertainment biases when analyzing SNP data: applitcations to the estimation of linkage disequilibrium. Theoretical Population Biology 63: 245-255.

Nordborg M., Borevitz J.O., Bergelson J., Berry C.C., Chory J., Hagenblad J. et al. (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics 30: 190-193.

Nordborg M., Tavaré S. (2002) Linkage disequilibrium: what history has to tell us. Trends in Genetics 18: 83-90.

Norton L.R., Firbank L.G., Watkinson A.R. (1995) Ecotypic differentiation of response to enhance CO_2 and temperature levels in *Arabidopsis thaliana*. Oecologia 104: 394-396.

Oden N.L., Sokal R.F. (1986) Directional autocorrelation: an extension of spatial autocorrelograms to two dimensions. Systematic Zoology 35: 608-617.

Orbovic V., Tarasjev A. (1999) Genetic differences in plastic responses to density between ecotypes of *Arabidopsis thaliana*. Russian Journal of Genetics 35: 528-536.

Perneger T.V. (1998) What's wrong with Bonferroni adjustments. Brit. Med. J. 316: 1236-1238.

Pigliucci M. (2001) Phenotypic plasticity. The Johns Hopkins University Press, Baltimore and London.

Pigliucci M., Cammell K., Schmitt J. (1999) Evolution of phenotypic plasticity - a comparative approach in the phylogenetic neighbourhood of *Arabidopsis thaliana*. Journal of Evolutionary Biology 12: 779-791.

Pigliucci M., Hayden K. (2001) Phenotypic plasticity is the major determinant of changes in phenotypic integration in *Arabidopsis*. New Phytologist 152: 419-430.

Pigliucci M., Marlow E.T. (2001) Differentiation for flowering time and phenotypic integration in *Arabidopsis thaliana* in response to season length and vernalization. Oecologia 127: 501-508.

Pigliucci M., Schlichting C.D., Whitton J. (1995a) Reaction norms of *Arabidopsis*. II. Response to stress and unordered environmental variation. Functional Ecology 9: 537-547.

Pigliucci M., Whitton J., Schlichting C.D. (1995b) Reaction norms of *Arabidopsis*. I. Plasticity of characters and correlations across water, nutrient and light gradients. Journal of Evolutionary Biology 8: 421-438.

Pollard H., Cruzan M., Pigliucci M. (2001) Comparative studies of reaction norms in *Arabidopsis*. I. Evolution of response to daylength. Evolutionary Ecology Research 3: 129-155.

Posada, D. (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. Molecular Biological Evolution 19: 708-717.

Posada D., Crandall K.A., Holmes E.C. (2002) Recombination in evolutionary genomics. Annual Review of Genetics 36: 75-97.

Price R.A., Al-Shebaz I.A., Palmer J.D. (1994) Systematic relationships of *Arabidopsis*: a molecular and morphological approach. In: *Arabidopsis* (eds Meyerowitz E, Somerville C), pp. 7-19. Cold Spring Habor Press, Cold Spring Habor, NY.

Pritchard J., Stephens M., Donelly P. (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Purugganan M.D., Suddith J.I. (1998) Molecular population genetics of the *Arabidopsis* CAULIFLOWER regulatory gene: Nonneutral evolution and naturally occurring variation in floral homeotic function. Proceedings of the National Academy of Sciences of the USA 95: 8130-8134.

Purugganan M.D., Suddith J.I. (1999) Molecular population genetics of floral homeotic loci: departures from equilibrium-neutral model at the *APETALA3* and *PISTILLATA* genes of *Arabidopsis thaliana*. Genetics 151: 839-848.

Quesada V., Ponce M.R., Micol J.L. (2000) Genetic analysis of salt-tolerant mutants in *Arabidopsis thaliana.* Genetics 154: 421-436.

Redéi G.P. (1964) Crossing experiments with polyploids. *Arabidopsis* Information Service 1: 13.

Redéi G.P. (1969) *Arabidopsis thaliana* (L.) Heynh. a review of the genetics and bilogy Bibliographia Genetica 21: 1-151.

Redéi G.P. (1975) *Arabidopsis* as a genetic tool. Annual Review of Ecology and Systematics 9: 111-127.

Reeves G., Francis D., Davies M.S., Rogers H.J., Hodkinson T.R. (1998) Genome size is negatively correlated with altitude in natural populations of *Dactylis polygama*. Annals of Botany 82 (Suppl. A): 99-105.

Robles P., Micol J.L. (2001) Genome-wide linkage analysis of *Arabidopsis* genes required for leaf development. Molecular Genetics and Genomics 266: 2-19.

Röhr M (1993) Statistische Strukturanalysen. Fischer, Stuttgart.

Rozas J., Rozas R. (1999) DNASP version3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15: 174-175.

Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G. et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933.

Sachs L. (1997) Angewandte Statistik - Anwendung klassischer Methoden. 8.Auflage. Berlin, Heidelberg, Springer Verlag.

Saitou N., Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Sangwan R.S., Bourgeois Y., Sangwan-Norreel B.S. (1991) Molecular and General Genetics 230: 475-485.

SAS (1990) SAS/STAT User's Guide, version 6, 4th edition. SAS Institute, Cary, N.C.

Savolainen O., Langley C., Lazzaro B., Freville H. (2000) Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabdopsis lyrata* and the selfing *Arabidopsis thaliana*. Molecular Biology and Evolution 17: 645-655.

Schlichting C.D., Pigliucci M. (1993) Evolution of phenotypic plasticity via regulatory genes. American Naturalist 142: 366-370.

Schlichting C.D., Pigliucci M. (1995) Gene regulation, quantitative genetics and the evolution of reaction norms. Functional Ecology 9: 154-168.

Schmid K.J., Rosleff Sörensen T., Stracke R., Törjék O., Altmann T., Mitchell-Olds T. et al. (2003) Large-scale identification and analysis of genome-wide single-nucleotide-polymorphisms for mapping in *Arabidopsis thaliana*. Genome Research 13: 1250-1257.

Schmid K.J., Törjék O., Mayer R., Schmuths H., Hoffmann M.H., Altmann T. Evidence for pleistocene refugia and postglacial expansion of *Arabidopsis thaliana* from genome-wide SNP markers. Molecular Ecology, submitted.

Schmuths H., Meister A., Horres R., Bachmann K. (2004a) Genome size variation among accessions of *Arabidopsis thaliana*. Annals of Botany 93: 317-321.

Schmuths H., Hoffmann M.H., Bachmann K. (2004b) Geographic distribution and recombination of genomic fragments on the short arm of chromosome 2 of *Arabidopsis thaliana*. Plant Biology, 6: 128-139.

Sharbel T.F., Haubold B., Mitchell-Olds T. (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Molecular Ecology 9: 2109-2118.

Sokal R., Rohlf F.J. (1995) Biometry. The principles and practice of statistics in biological research. Freeman, New York.

Soltis D.E., Soltis P.S. (1995) The dynamic nature of polyploid genomes. Proceedings of the National Academy of Sciences of the USA 92: 8089-8091.

SPSS for Windows. (1999) SPSS Inc., Chicago, USA.

Stahl E. A., Dwyer G., Mauricio R., Kreitman M., Bergelson J. (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400: 667-671.

Stebbins G.L. (1971) Chromosomal Evolution in Higher Plants, Edward Arnold Ltd., London

Stehlik I., Blattner F.R., Holderegger R. (2002) Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. Molecular Ecology 11: 2027-2036.

Stenoien H.K., Fenster C.B., Kuittinen H., Savolainen O. (2002) Quantifying latitudinal clines to light responses in natural populations of *Arabidopsis thaliana* (Brassicaceae), American Journal of Botany 89: 1604-1608.

Strasburger E., Noll F., Schenk H., Schimper A.F.W. (1991) Lehrbuch der Botanik für Hochschulen. 33. Auflage. Stuttgart, Jena, New York: Gustav Fischer Verlag.

Su B., Xiao J., Underhill P., Deka R., Zhang W. Akey J., Huang W. et al. (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last ice age. American Journal of Human Genetics 65: 1718-1724.

Syvänen A.-C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2: 930-942.

Taberlet P., Fumagalli L., Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453-464.

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

Templeton A.R., Crandall K.A., Sing C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619-633.

Todokoro S., Terauchi R., Kawano S. (1995) Microsatellite polymorphisms in natural popula- tions of *Arabidopsis thaliana* in Japan. The Japanese Journal of Genetics 70: 543-554.

Törjék O., Meyer R.C., Müssig C., Schmid K.J., Rosleff Sörensen T., Weisshaar B. et al. (2003) Establishment of a high-efficiency SNP-based framework marker set for *Arabidopsis*. The Plant Journal 36: 122-140.

Turakulov R., Easteal S. (2003) Number of SNPs loci needed to detect population structure. Human Mutation 55: 37-45.

Turpeinen T., Kulmala J., Nevo E. (1999) Genome size variation in *Hordeum spontaneum* populations. Genome 42: 1049-1099.

Ullrich H., Lättig K., Brennicke A., Knoop V. (1997) Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. Plant Molecular Biology 33: 37-45.

Ungerer M.C., Halldorsdottir S.S., Modliszewski J.L., Mackay T.F.C., Purugganan M.D. (2002) Quantitative trait loci for inflorescence development in *Arabidopsis thaliana*. Genetics 160: 1133-1151.

Vander Zwan C., Brodie S.A., Campanella J.J. (2000) The intraspecific phylogenetics of A.t. in worldwide populations. Systematic Botany 25(1): 47-59.

Vega C.R.C., Andrade F.H., Sadras V.O., Uhart S.A., Valentinuz O.R. (2001) Seed Number as a Function of Growth. A Comparative Study in Soybean, Sunflower, and Maize. Crop Science 41: 748-754.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G. et al. (2001) The sequence of the human genome. Science 291: 1304-1351.

Via S. (1993) Adaptive phenotypic plasticity: target or by-product of selection in variable environment? American Naturalist 142: 353-365.

Vilhar B., Greilhuber J., Koce J.D., Temsch E.M., Dermastia M. (2001) Plant genome size measurement with DNA image cytometry. Annals of Botany 87: 719-728.

Wakeley J., Nielsen R., Liu-Cordero S.N., Ardlie K. (2001) The discovery of single-nucleotide polymorphisms - and inferences about human demographic history. American Journal of Human Genetics 69: 1332-1347.

Wang D.G., Fan J.B., Siao C.J., Berno A., Young P., Sapolsky R. et al. (1998) Large scale identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. Science 280: 1077-1082.

Wares J.P. (2002) Community genetics in the Northwestern Atlantic interdial. Molecular Ecology 11: 1131-1144.

Weir B. (1996) Genetic Data Analysis II. Sinauer Associates, Sunderland, MA.

Wendel J.F. (2000) Genome evolution in polyploids. Plant Molecular Biology 42: 225-249.

Westerman J.M. (1971) Genotype-environment interaction and developmental regulation in *Arabidopsis thaliana*. II. Inbred lines; Analysis. Heredity 26: 93-106.

Westerman J.M., Lawrence M.J. (1970) Genotype-environment interaction and developmental regulation in *Arabidopsis thaliana*. I. Inbred lines; Description. Heredity 25: 609-627.

Wicks S.R., Yeh R.T., Gish W.R., Waterston R.H., Plasterk R.H.A. (2001) Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. Nature Genetics 28: 160-164.

Willis K. (1996) Where did all the flowers go? The fate of temperate European flora during glacial periods. Endeavor 110: 110-114.

Wolffe A.P., Matzke M.M. (1999) Epigenetics: regulation through repression. Science 286: 481-486.

Wolterek R. (1909). Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer Artunterschiede bei Daphnien. Versuche der Deutschen Zoologischen Gesellschaft 19: 110-172.

Woodward F.I. (1987) Climate and plant distribution. Univ. Press, Cambridge.
Wricke G. (1971). Populationsgenetik. Sammlung Göschen de Gruyter, Berlin.
Wright S. (1931) Evolution of mendelian populations. Genetics 16: 97-159.

Internetseiten

http://www.arabidopsis.org/info/aboutarabidopsis.jsp
http://www.arabidopsis.org/Cereon/index.html
http://www.mipz-koeln.mpg.de/masc/
http://nasc.nott.ac.uk/
http://mips.gsf.de
http://www.genome.wi.mit.edu/genome_software/other/primer3
http://www.technelysium.com.au/chromas.html
http://www.fas.umontral.ca/BIOL/Casgrain/en/labo/R
http://www.r-project.org/
http://www.mpimp-golm.mpg.de/arab-diversity
http://www.mpiz-koeln.mpg.de/masc/
http://www-hto.usc.edu/noahr/distruct.html
http://www.zmp.de/info/jahresbericht/getreide_prognose.asp

8. Verzeichnis der Abkürzungen und Fachtermini

AFLP	amplified fragment length polymorphism
ANOVA	Varianzanalyse (analysis of variance)
bp	Basenpaar (baise pair)
°C	Grad Celsius
CAPS	cleaved amplified polymorphic sequence
1C-Wert	Größe des haploiden Genoms (wird oft in [pg] angegeben)
DNA	Desoxyribonukleinsäure (deoxyribonucleic acid)
h	Stunde
InDel	Insertion/Deletion
MALDI-ToF	Matrix Assiated Laser Desorption/Ionisation Time-of-Flight
Mbp	Megabasenpaare (mega base pairs)
kbp	Kilobasenpaare (kilo base pairs)
PCA	Hauptkomponentenanalyse (principal component analysis)
PCR	polymerase chain reaction
pg	Pikogramm (picogramm)
QTL	quantitative trait loci
RNA	Ribonukleinsäure (ribonucleic acid)
RFLP	restricted fragment length polymorphism
Singleton	einmalig im Datensatz vorkommender Einzelbasenunterschied
SNP	Einzelbasenunterschied (single nucleotide polymorphism)
var.	Varietät

9. Anhang – Supplementary Data

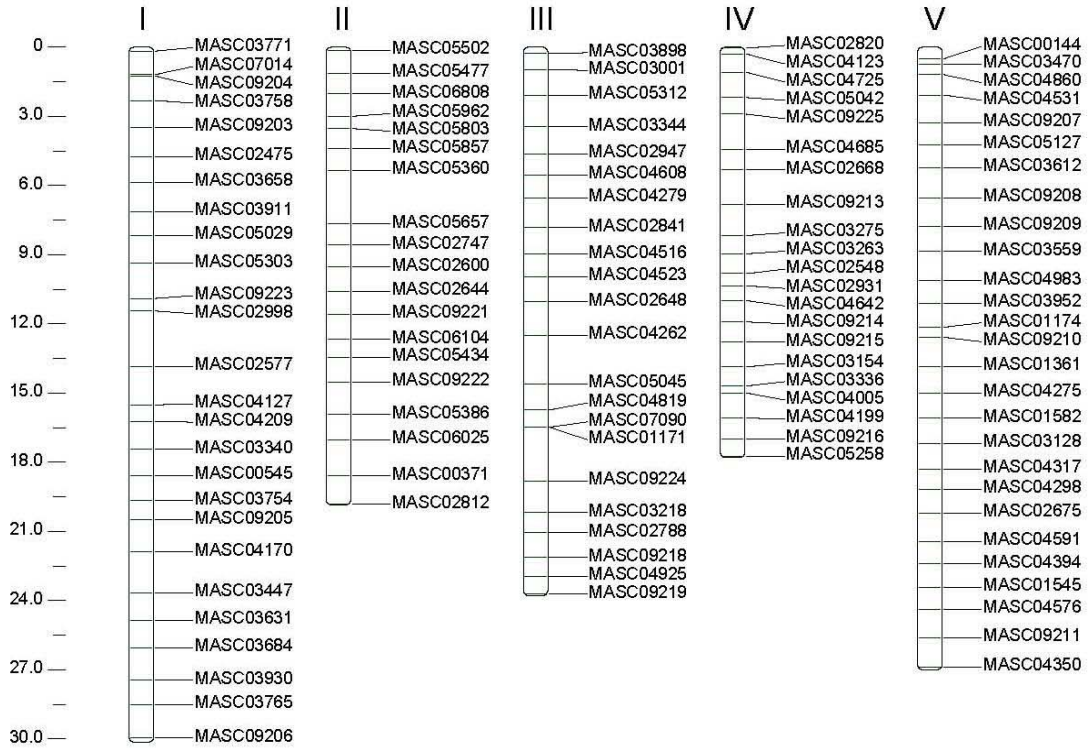


Fig. 1. Physical map of the 115 SNP markers used for this study. Identifiers can be used to retrieve more information about the SNPs from the MASC database at http://www.mipz-koeln.de/masc.

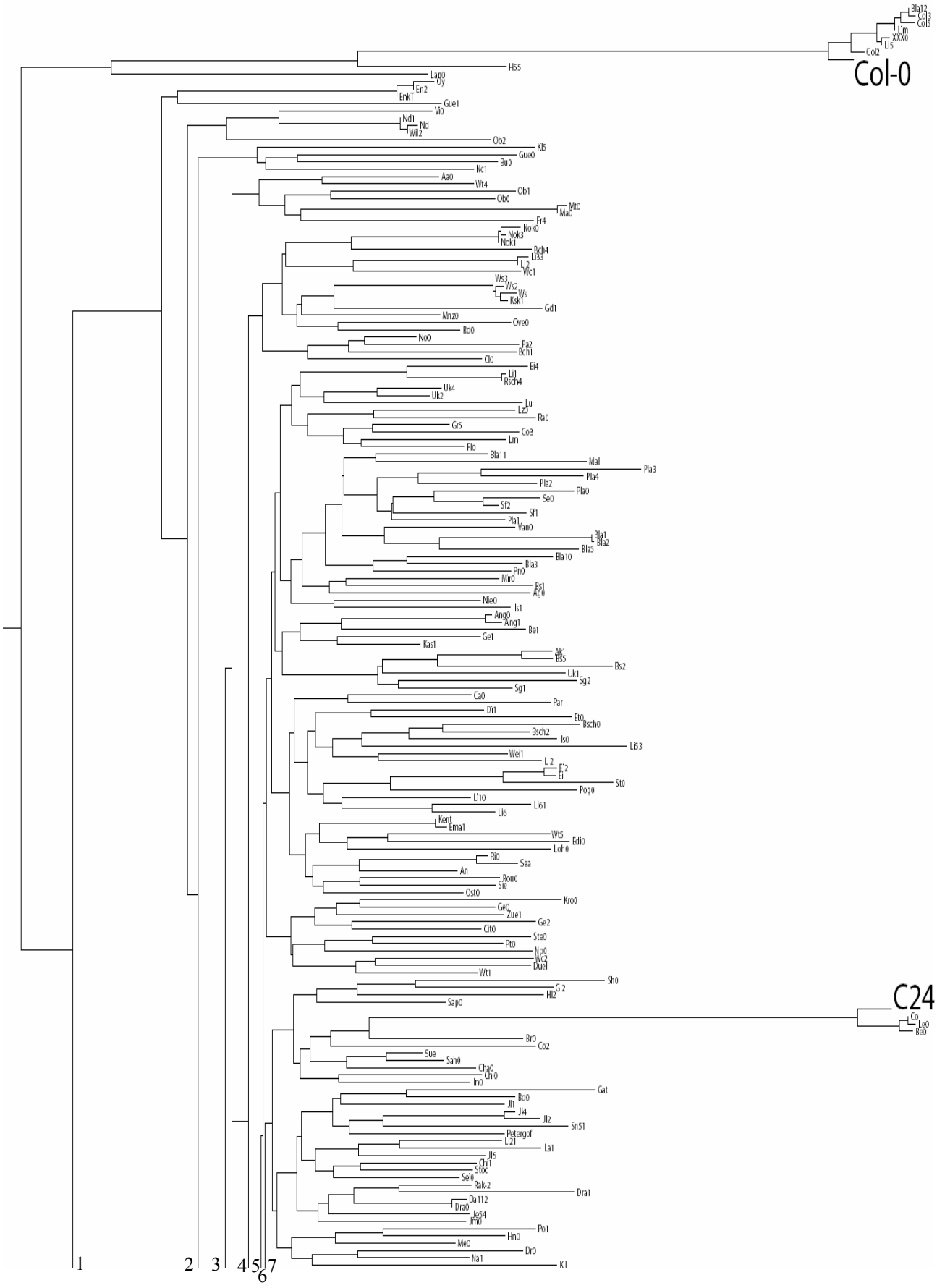


Fig. 2. Neighbor-joining distance tree of 335 accessions that is based on 49 SNP markers for which the evolutionary distance state could be determined by outgroup comparision. The tree is rooted with the outgroup.

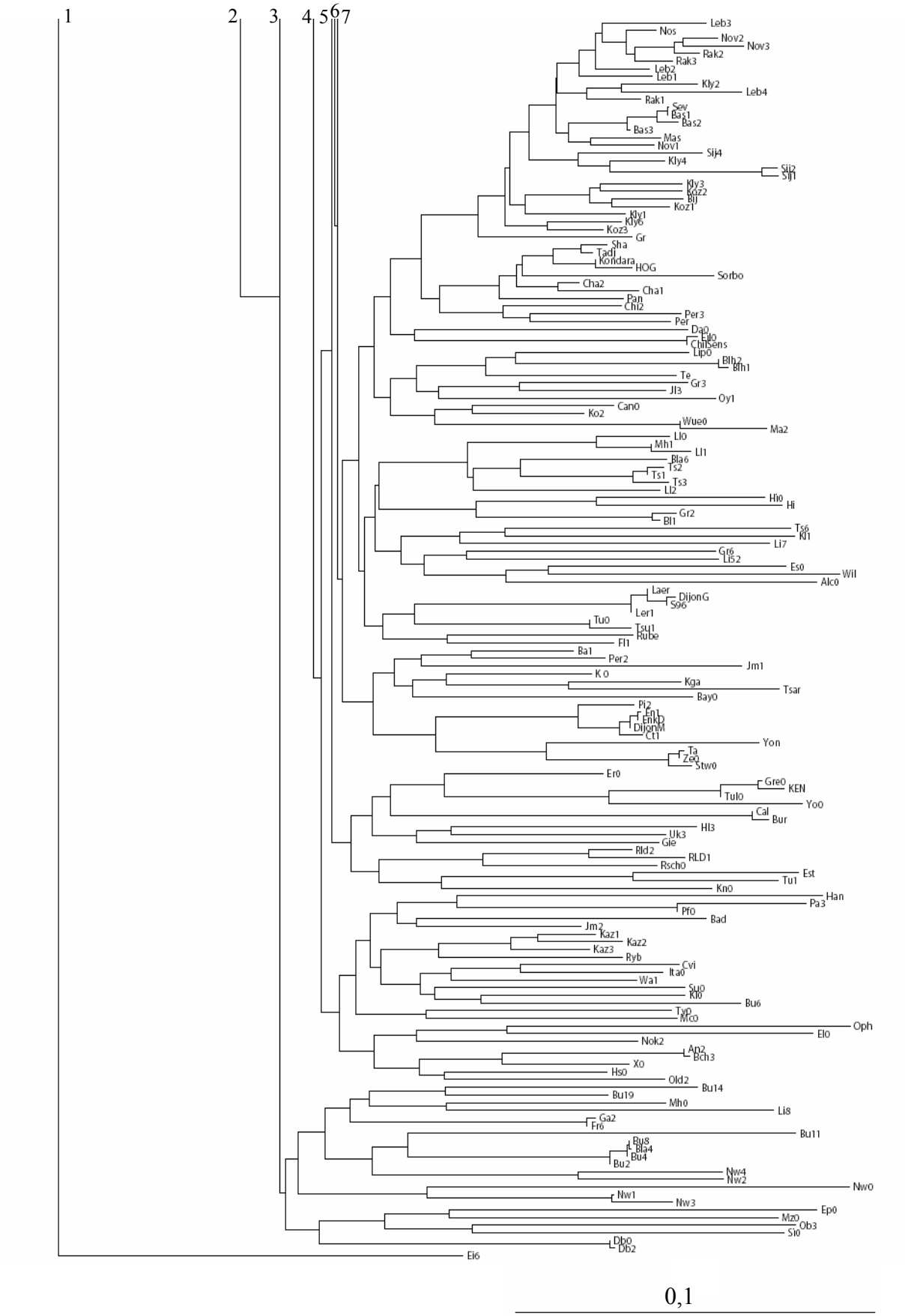


Fig. 2. (continued) Neighbor-joining distance tree of 335 accessions that is based on 49 SNP markers for which the evolutionary distance state could be determined by outgroup comparision. The tree is rooted with the outgroup.

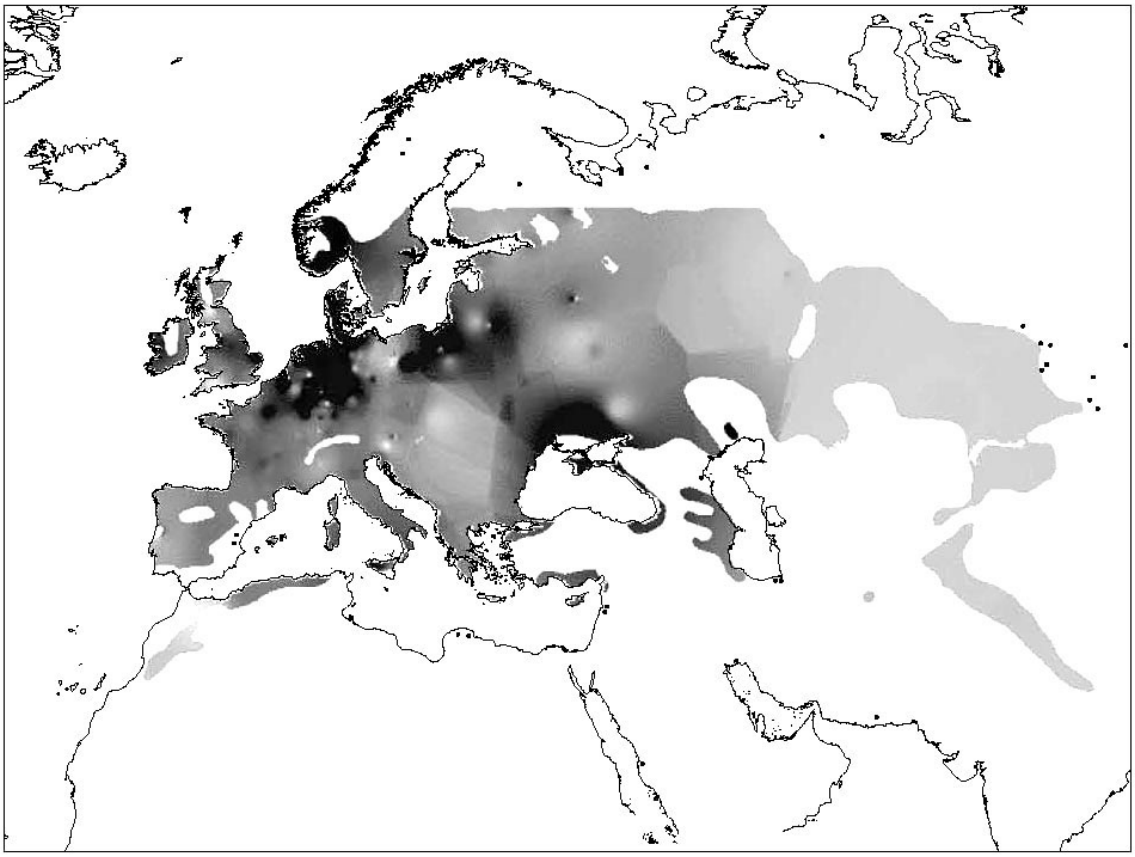


Fig. 3. Geographic pattern of model-based assignment of *A. thaliana* accessions to different populations. Population structure was inferred with K=3 using the *structure* program. The colors represent the assignment of accessions to each of the three populations, which was geographically extrapolated using the IDW function of the Arc/Info geographical information system.