Theor Appl Genet (2008) 117:1041–1053
DOI 10.1007/s00122-008-0843-y

ORIGINAL PAPER

Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits

Andreas Achleitner · Nicholas A. Tinker · Elisabeth Zechner · Hermann Buerstmayr

Received: 22 January 2008 / Accepted: 28 June 2008 / Published online: 17 July 2008

Abstract One hundred and fourteen oat (*Avena sativa* L.) varieties of worldwide origin were evaluated for genetic diversity based on 77 molecular polymorphisms produced by eight selective AFLP primer combinations. Genetic similarity, calculated using the DICE coefficient, was used for cluster analysis and principal component analysis was applied. In addition population structure was explored to identify discrete subpopulations based on allele frequency. Although clustering and population structure showed relationships with region and country of origin, there was no obvious relationship to hull presence or hull colour. Oat varieties originating from European breeding programs showed less diversity than varieties originating from North and South America. Associations between AFLP markers and agronomic traits (grain yield, groat yield, panicle emergence, plant height, and lodging) as well as kernel quality traits (kernel weight, test weight, screening percent and groat percent) were also investigated. Marker-trait associations were tested using a naïve simple regression model and five additional models that account for population structure. Significant associations were found for 23 AFLP markers, with many of these affecting multiple traits. This study demonstrates that diversity can be significantly enhanced using a global collection, and provides evidence for marker-trait associations that can be validated in segregating populations and exploited through marker-assisted selection.

A. Achleitner and N. A. Tinker contributed equally to this work.

Communicated by J. Yu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0843-y) contains supplementary material, which is available to authorized users.

A. Achleitner · H. Buerstmayr (✉)
Department for Agrobiotechnology, IFA-Tulln, Institute for Biotechnology in Plant Production, University of Natural Resources and Applied Life Sciences, Vienna, Konrad Lorenz Str. 20, Tulln 3430, Austria
e-mail: hermann.buerstmayr@boku.ac.at

Present Address:
A. Achleitner
Landwirtschaftskammer Steiermark, Hamerlinggasse 3, Graz 8010, Austria

N. A. Tinker
AAFC-AAC, ECORC, 960 Carling Ave, Ottawa, ON, Canada K1A-0C6

E. Zechner
Saatzucht Edelhof, Edelhof 1, 3910 Zwettl, Austria

Introduction

Oat (*Avena sativa* L.) is a cereal crop that is used throughout the world for human food and animal feed. Compared to other cereal crops, oat is reputed to be better suited for production under marginal environments, including cool-wet climates and soils with low fertility (Hoffmann 1995). Many oat varieties flower and mature quickly in short seasons with long daylength regimes, thus oat is an important crop in many northern countries. The major oat growing areas lie between latitudes 40° and 60°N in America, Europe and Asia, while a smaller proportion of the oat world production originates from the southern hemisphere, i.e. South America, Australia and New Zealand (Forsberg and Reeves 1992). In most countries, including European countries, the area cultivated with oat has declined during the past decades. This is partly attributed to the decline in

oat production for on-farm feed, and to the increase in high-input crops such as maize, wheat, and soybean.

In recent years, the demand in oat for human consumption has increased, particularly because of demonstrated dietary benefits of whole grain and soluble fibre (Food and Drug Administration 1997). Because the oat acreage is much lower than that of most other cereals, the investment in oat breeding is also lower. For example, in 1973 there were 17 breeding companies performing active oat breeding in Germany and 7 in Austria (Aufhammer and Fischbeck 1973). By 2007, these numbers shrank to 5 (Germany) and 1 (Austria).

Plant breeders usually develop populations for variety development from crosses within regionally adapted germplasm. Exchange of germplasm among breeders is common, but occurs predominantly within regions of adaptation. Exchange of germplasm across continents is less common. As a consequence, European spring oat breeding programs may be relying on a small fraction of the available genetic variation in this crop species. Buerstmayr et al. (2007) assessed variation in agronomic and quality traits in a set of 120 spring oat varieties of worldwide origin and found that for several physical grain quality traits, germplasm from outside Europe exhibited superior trait expression compared to European oats.

Genetic diversity in North American oat has been studied using pedigree information (Rodgers et al. 1983; Souza and Sorrells 1989), plant traits (Souza and Sorrells 1991a, b) and RFLP markers (O'Donoughue et al. 1994). In recent years, molecular markers have become increasingly useful for evaluating genetic diversity in plant populations (Mohammadi and Parasanna 2003). Li et al. (2000) developed and used SSR markers to analyse genetic relationships among different *Avena* species. Genetic diversity in Canadian oats has been analysed using AFLPs (Fu et al. 2003, 2004). Paczos-Grzeda (2004) compared AFLP and RAPD markers for assessing genetic diversity in oat varieties registered in Poland. Baohong et al. (2003) used RAPDs for measuring and comparing genetic diversity in Chinese and European oat accessions. Fu et al. (2005) used AFLPs to characterize a world collection of 670 oat accessions from 79 countries. Although this study addressed global genetic diversity using many historic varieties, a detailed analysis of oat diversity that includes modern European oat varieties has not yet been reported.

One of the limiting factors in genomic analysis of many plant species, including oats, is that most genomic studies have been conducted in experimental populations developed from a bi-parental cross. Thus, while many QTL have been reported, the effects of these QTL often turn out to be unique to a specific genetic background, and there has been limited success in applying the results. Many researchers now consider that association analysis, whereby genes and QTL are detected in a random set of genotypes from a mixed genetic background, is a viable solution to this problem (Gupta et al. 2005, Rostoks et al. 2006, Breseghello and Sorrells 2006). The concept of association analysis has been known for many years, and the strategy has been applied with limited success in oat (Beer et al. 1997). However, the increased availability of molecular markers, and the refinement of statistical tools, has kindled renewed interest in this approach. Although association analysis shows great promise as an efficient and valuable tool for gene discovery, the analysis of marker-trait associations must account for the presence of population structure. Failure to do so can cause the detection of spurious associations between traits and unlinked markers. Current methods for accounting for population structure are described by Yu et al. (2006) and are illustrated in a case study presented by Zhao et al. (2007).

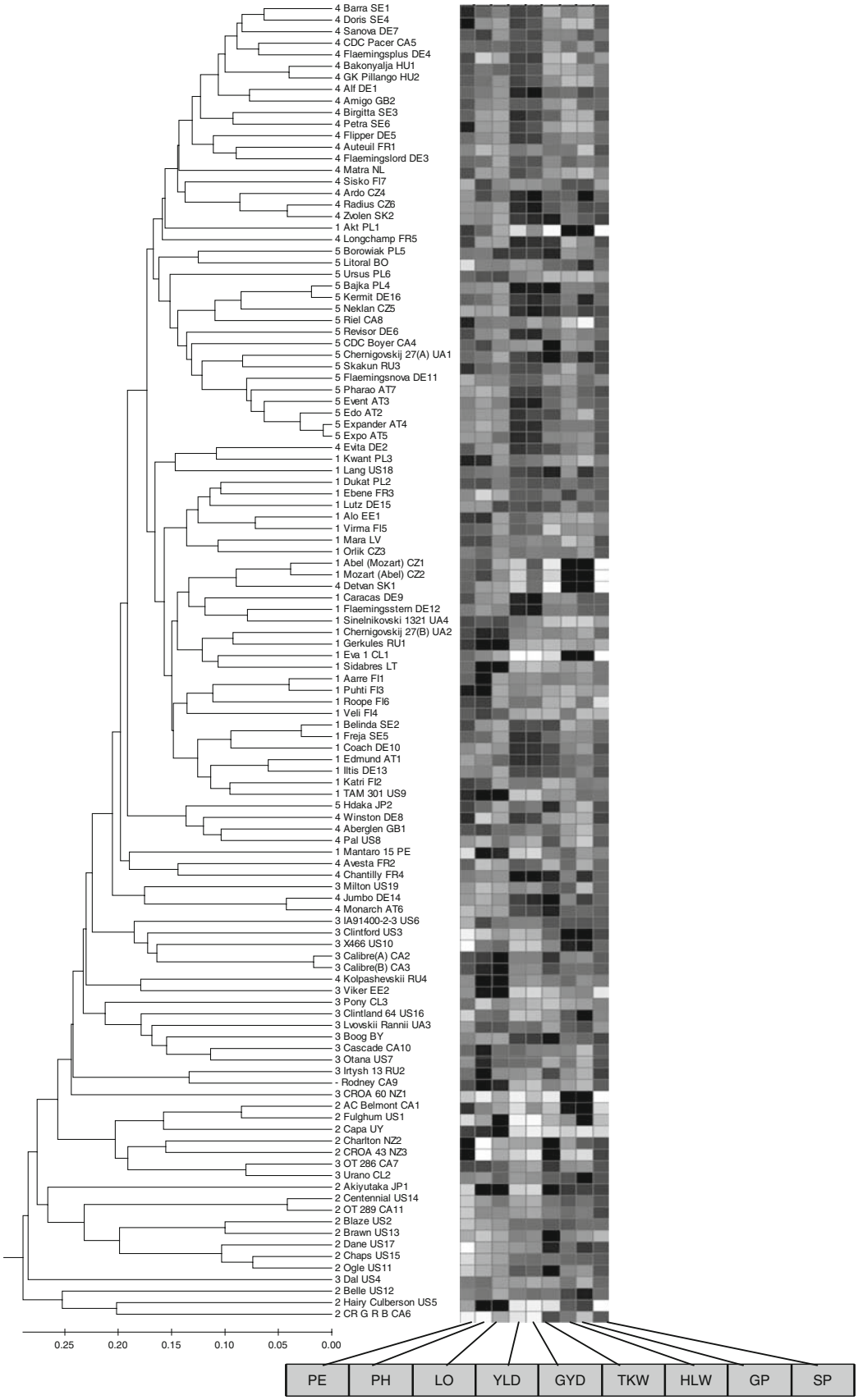
The current study was conducted to examine genetic diversity and population structure in oat varieties of worldwide origin, to compare levels of variability in European oats versus those from other regions, and to test for preliminary evidence of genetic associations between AFLP markers and quantitative traits.

Materials and methods

Plant Material

One hundred and fourteen spring oat (*Avena sativa* L.) varieties[1] of worldwide origin were chosen for this study. All varieties were evaluated for agronomic and physical grain quality traits in replicated field experiments in previously reported work (Buerstmayr et al. 2007). Briefly, the oat varieties were sown in three replicated field experiments in Austria and one in Germany. Panicle emergence, plant height and lodging severity were evaluated under field conditions, and grain yield, 1,000 kernel weight, hectolitre weight, screenings percentage >2 mm, and groat percentage were measured after harvesting (see Buerstmayr et al. 2007 for details). Mean trait values were scaled to unit variance, centred to mean zero, and depicted as a heat-map in Fig. 1. Among the 114 varieties, 74 originated from different parts of Europe and 40 from outside Europe: North America (30), South America (5), Asia (2) and Oceania (3). The names, countries of origin and hull colours of the varieties are listed in Table 1. For many of these varieties, addi-

[1] Although most oat genotypes in this study can be considered as cultivars (cultivated varieties) several numbered breeding lines were also investigated. We will use the term 'variety' to describe all types of germplasm in this study. All varieties were considered to be nearly homozygous pure lines.


4 Barra SE1
4 Doris SE4
4 Sanova DE7
4 CDC Pacer CA5
4 Flaemingsplus DE4
4 Bakonyalja HU1
4 GK Pillango HU2
4 Alf DE1
4 Amigo GB2
4 Birgitta SE3
4 Petra SE6
4 Flipper DE5
4 Auteuil FR1
4 Flaemingslord DE3
4 Matra NL
4 Sisko FI7
4 Ardo CZ4
4 Radius CZ6
4 Zvolen SK2
1 Akt PL1
4 Longchamp FR5
5 Borowiak PL5
5 Litoral BO
5 Ursus PL6
5 Bajka PL4
5 Kermit DE16
5 Neklan CZ5
5 Riel CA8
5 Revisor DE6
5 CDC Boyer CA4
5 Chernigovskij 27(A) UA1
5 Skakun RU3
5 Flaemingsnova DE11
5 Pharao AT7
5 Event AT3
5 Edo AT2
5 Expander AT4
5 Expo AT5
4 Evita DE2
1 Kwant PL3
1 Lang US18
1 Dukat PL2
1 Ebene FR3
1 Lutz DE15
1 Alo EE1
1 Virma FI5
1 Mara LV
1 Orlik CZ3
1 Abel (Mozart) CZ1
1 Mozart (Abel) CZ2
4 Detvan SK1
1 Caracas DE9
1 Flaemingsstern DE12
1 Sinelnikovski 1321 UA4
1 Chernigovskij 27(B) UA2
1 Gerkules RU1
1 Eva 1 CL1
1 Sidabres LT
1 Aarre FI1
1 Puhti FI3
1 Roope FI6
1 Veli FI4
1 Belinda SE2
1 Freja SE5
1 Coach DE10
1 Edmund AT1
1 Iltis DE13
1 Katri FI2
1 TAM 301 US9
5 Hdaka JP2
4 Winston DE8
4 Aberglen GB1
4 Pal US8
1 Mantaro 15 PE
4 Avesta FR2
4 Chantilly FR4
3 Milton US19
4 Jumbo DE14
4 Monarch AT6
3 IA91400-2-3 US6
3 Clintford US3
3 X466 US10
3 Calibre(A) CA2
3 Calibre(B) CA3
4 Kolpashevskii RU4
3 Viker EE2
3 Pony CL3
3 Clintland 64 US16
3 Lvovskii Rannii UA3
3 Boog BY
3 Cascade CA10
3 Otana US7
3 Irtysh 13 RU2
- Rodney CA9
3 CROA 60 NZ1
2 AC Belmont CA1
2 Fulghum US1
2 Capa UY
2 Charlton NZ2
2 CROA 43 NZ3
3 OT 286 CA7
3 Urano CL2
2 Akiyutaka JP1
2 Centennial US14
2 OT 289 CA11
2 Blaze US2
2 Brawn US13
2 Dane US17
2 Chaps US15
2 Ogle US11
3 Dal US4
2 Belle US12
2 Hairy Culberson US5
2 CR G R B CA6
0.25
0.20
0.15
0.10
0.05
0.00
PE
PH
LO
YLD
GYD
TKW
HLW
GP
SP

◀ **Fig. 1** UPGMA dendrogram of 114 oat varieties based on DICE distance for 77 AFLPs. The numbers left of the variety names indicate the Structure Coefficients based on the STRUCTURE analysis (see Table 1). The last digits behind the variety names indicate the variety identities (see Table 1). The panel on the right illustrates the phenotypes for nine traits in greyscales. *Dark colours* indicate high numerical values for the trait, and *light colours* indicate low values

tional information can be found in oat databases such as the National Plant Germplasm System of the USA (http://www.ars-grin.gov/npgs/), the European Avena Database (http://eadb.bafz.de/) and the Canadian Pedigrees of Oat Lines (POOL) database (Tinker and Deyl 2005: http://avena.agr.gc.ca/OGIS/). Three pairs of varieties were identified as putative sister-lines, and are identified with extensions A or B as needed: (1) The Czech variety Abel is registered and marketed in Germany with the variety name Mozart, (2) the variety Chernigovskij 27 was received from two different sources, but originally spelled differently, and (3) two independent sources of Calibre were included.

Molecular marker analysis

Five panicles were harvested from each variety from field plots grown in 2002 at the plant breeding station Edelhof in Austria (Buerstmayr et al. 2007). Five seeds (one per panicle) were sown in small pots in the greenhouse. Young seedlings at the three-leaf stage were harvested and lyophilized. Dried leaf tissue from five plants was pooled, ground in a mixer mill, and DNA was extracted using a CTAB method in 1.5 ml tubes (modified from Saghai-Maroof et al. 1984). DNA quality and quantity were measured in a UV photometer. The AFLP analysis (Vos et al. 1995) was conducted as described by Hartl et al. (1999) and Buerstmayr et al. (2002) using *Mse*I and *Sse*8387I restriction enzymes. In total, 8 AFLP primer combinations with two selective nucleotides on the 3′ end of either primer were performed. AFLP fragments in 25 cm 6% acrylamide gels were detected using a LI-COR 4200 IR^2 dual-dye sequencing system. Gel images were scored visually and polymorphic bands were recorded as present or absent. Any data points that were not clearly scorable were treated as missing values. Monomorphic AFLP bands were not included in the statistical analysis. AFLPs with rare alleles (less than 5 dominant scores) were scored, but were excluded from further statistical analysis. The standard list for AFLP primer nomenclature (http://wheat.pw.usda.gov/ggpages/keygeneAFLPs.html) was applied. The polymorphic AFLP markers were named based on the primer combination followed by an arbitrary number which referred to the specific polymorphism. Marker scores were recorded in a matrix of 1's and 0's to represent presence or absence of dominant bands.

Cluster analysis and principal components analysis

Genetic similarities were calculated from the scoring matrix using the DICE coefficient (Dice 1945) in NTSYS 2.11 (Rohlf 2000). From the similarity data, genetic distance data were calculated for each pair of varieties (distance = 1 – similarity) and used for UPGMA clustering in MEGA 3 (Kumar et al. 2004). The AFLP marker data were subjected to a principal components analysis, with scores 1 versus 0 scaled to unit variance for each marker. Computation was performed by singular value decomposition of the data matrix using the 'prcomp' procedure in the statistics module (version 2.6.2) of the *R* statistical package (http://www.r-project.org/).

Population structure

The program STRUCTURE, version 2.2 (Pritchard et al. 2000) was used to identify K discrete subpopulations based on models characterized by non-correlated gene frequencies with no admixture. Essentially, populations were subdivided into K subpopulations, each with unique allele frequency profiles, and population membership was adjusted until the goodness of fit (measured by $Pr(X|K)$) was maximized. A complete scoring matrix was used to examine population structure. Marker alleles were coded as '1' or '0' and individuals were treated as haploids to avoid any assumptions about dominance or heterozygotes. Two additional data subsets were produced by removing (at random) one marker from each pair of markers found to have identical or reciprocal genotypes at more than 90% or 95% of the tested varieties. Values of $K = 1$ through $K = 9$ were tested for each data set by finding at least three numerical solutions for each combination. Each numerical solution was optimized using 1,000,000 iterations (including 500,000 'burn-in' iterations). A value of K was selected as the minimum K at which $Pr(X|K)$ no longer increased with increasing values of K in most numerical solutions. At this level, membership in subpopulations was compared for consistency among replications and data subsets. A representative solution was selected, and the membership of each variety in one of the K subpopulations was designated using a numerical index, described hereafter as the 'Structure Coefficient'.

Association analysis

The mean phenotypic values based on four field environments reported by Buerstmayr et al. (2007) were used for association analysis. Four traits, including groat yield (GYLD) lodging severity (LO), panicle emergence (PE), plant height (PH) were analysed across all varieties for which complete marker data were available. Five additional traits,

Table 1 List of 114 spring oat varieties used for genetic diversity analysis

Variety	ID[a]	Country of origin	Hull	SC[b]
Abel (Mozart)	CZ1	Czech Republic	N	1
Mozart (Abel)	CZ2	Czech Republic	N	1
Akt	PL1	Poland	N	1
Detvan	SK1	Slovak Republic	N	4
Auteuil	FR1	France	B	4
Avesta	FR2	France	B	4
Ebene	FR3	France	B	1
Edmund	AT1	Austria	W	1
Orlik	CZ3	Czech Republic	W	1
Alf	DE1	Germany	W	4
Evita	DE2	Germany	W	4
Flaemingslord	DE3	Germany	W	4
Flaemingsplus	DE4	Germany	W	4
Flipper	DE5	Germany	W	4
Revisor	DE6	Germany	W	5
Sanova	DE7	Germany	W	4
Winston	DE8	Germany	W	4
Alo	EE1	Estonia	W	1
Viker	EE2	Estonia	W	3
Aarre	FI1	Finland	W	1
Katri	FI2	Finland	W	1
Puhti	FI3	Finland	W	1
Veli	FI4	Finland	W	1
Virma	FI5	Finland	W	1
Chantilly	FR4	France	W	4
Longchamp	FR5	France	W	4
Aberglen	GB1	Great Britain (UK)	W	4
Amigo	GB2	Great Britain (UK)	W	4
Bakonyalja	HU1	Hungary	W	4
GK Pillango	HU2	Hungary	W	4
Sidabres	LT	Lithuania	W	1
Mara	LV	Latvia	W	1
Matra	NL	Netherlands	W	4
Dukat	PL2	Poland	W	1
Kwant	PL3	Poland	W	1
Gerkules	RU1	Russian Federation	W	1
Irtysh 13	RU2	Russian Federation	W	3
Skakun	RU3	Russian Federation	W	5
Barra	SE1	Sweden	W	4
Belinda	SE2	Sweden	W	1
Birgitta	SE3	Sweden	W	4
Doris	SE4	Sweden	W	4
Freja	SE5	Sweden	W	1
Petra	SE6	Sweden	W	4
Chernigovskij 27(A)	UA1	Ukraine	W	5
Chernigovskij 27(B)	UA2	Ukraine	W	1
Edo	AT2	Austria	Y	5
Event	AT3	Austria	Y	5
Expander	AT4	Austria	Y	5
Expo	AT5	Austria	Y	5
Monarch	AT6	Austria	Y	4
Pharao	AT7	Austria	Y	5
Boog	BY	Belarus	Y	3
Ardo	CZ4	Czech Republic	Y	4
Neklan	CZ5	Czech Republic	Y	5
Radius	CZ6	Czech Republic	Y	4
Caracas	DE9	Germany	Y	1
Coach	DE10	Germany	Y	1
Flaemingsnova	DE11	Germany	Y	5
Flaemingsstern	DE12	Germany	Y	1
Iltis	DE13	Germany	Y	1
Jumbo	DE14	Germany	Y	4
Lutz	DE15	Germany	Y	1
Roope	FI6	Finland	Y	1
Sisko	FI7	Finland	Y	4
Bajka	PL4	Poland	Y	5
Borowiak	PL5	Poland	Y	5
Ursus	PL6	Poland	Y	5
Kolpashevskii	RU4	Russian Federation	Y	4
Zvolen	SK2	Slovak Republic	Y	4
Lvovskii Rannii	UA3	Ukraine	Y	3
Sinelnikovski 1321	UA4	Ukraine	Y	1
Kermit	DE16	Germany	Y	5
AC Belmont	CA1	Canada	N	2
Eva 1	CL1	Chile	N	1
CROA 60	NZ1	New Zealand (Aotearoa)	N	3
Fulghum	US1	United States	T	2
Litoral	BO	Bolivia	W	5
Calibre(A)	CA2	Canada	W	3
Calibre(B)	CA3	Canada	W	3
CDC Boyer	CA4	Canada	W	5
CDC Pacer	CA5	Canada	W	4
CR-G-R-B[c]	CA6	Canada	W	2
OT 286	CA7	Canada	W	3
Riel	CA8	Canada	W	5
Rodney	CA9	Canada	W	–
Cascade	CA10	Canada	W	3
Urano	CL2	Chile	W	3
Akiyutaka	JP1	Japan	W	2
Hairy Culberson	US5	United States	W	2
Mantaro 15	PE	Peru	W	1
Blaze	US2	United States	W	2
Clintford	US3	United States	W	3
Dal	US4	United States	W	3
Hdaka	JP2	Japan	W	5

Table 1 continued

Variety	ID[a]	Country of origin	Hull	SC[b]
IA91400-2-3	US6	United States	W	3
Otana	US7	United States	W	3
Pal	US8	United States	W	4
TAM 301	US9	United States	W	1
X466	US10	United States	W	3
Capa	UY	Uruguay	W	2
OT 289	CA11	Canada	Y	2
Pony	CL3	Chile	Y	3
Charlton	NZ2	New Zealand (Aotearoa)	Y	2
CROA 43	NZ3	New Zealand (Aotearoa)	Y	2
Belle	US12	United States	Y	2
Brawn	US13	United States	Y	2
Centennial	US14	United States	Y	2
Chaps	US15	United States	Y	2
Clintland 64	US16	United States	Y	3
Dane	US17	United States	Y	2
Lang	US18	United States	Y	1
Milton	US19	United States	Y	3
Ogle	US11	United States	Y	2

Hull colour: *W* white, *Y* yellow, *B* black, *T* tan, *N* hulless oat

[a] Identity indicated by the ISO3166 country codes and for countries with more than one variety an arbitrarily chosen unique number within each country of origin

[b] Structure Coefficient based on STRUCTURE analysis of the data set excluding markers >95% similarity

[c] Full name: CR Gon Regoregox-Brorrn8025

including grain yield (YLD), hectolitre weight (HLW), thousand kernel weight (TKW), groat percentage (GP) and screenings percentage (SP) were analysed across all covered-seeded varieties, omitting seven hulless varieties.

Six methods were used to test for associations between AFLP markers and quantitative traits. All methods were performed using the software TASSEL, version 2.0.1 (Bradbury et al. 2007). First, a general linear model (GLM) was tested to identify single marker effects on quantitative traits. This is referred to as 'naïve GLM' because it does not account for population structure as a potential cause of the genotype–phenotype relationship. A second GLM was tested where the Structure Coefficient was used as a cofactor. Two additional GLM models were tested in which PCA axis 1 (PCA1) or PCA axes 1 through 4 (PCA1-4) were used as quantitative covariates. All of the GLM procedures tested fixed-effect models in which mean phenotypes of a given trait were predicted by the independent variables. A fifth and sixth model were tested using a unified mixed linear model (MLM) following Yu et al. (2006). One contained the matrix of kinship coefficients estimated among all varieties using the molecular marker data, and the second contained the kinship matrix plus the Structure Coefficient. All six models were tested for each of the 77 AFLP markers. Partial R^2 values were computed for the fixed marker effects, and tests for significance were applied using F statistic associated with the marker.

Table 2 Selective *Sse*8387I/*Mse*I AFLP primer combinations used for genotyping 114 oat varieties showing the number of polymorphic bands per primer combination

Primer combination	Selective bases	Number of polymorphic bands
S24/M20	TC/GC	15
S12/M17	AC/CG	6
S13/M14	AG/AT	4
S12/M19	AC/GA	8
S23/M20	TA/GC	11
S18/M17	CT/CG	15
S13/M20	AG/GC	10
S13/M25	AG/TG	8

Results

A total of 87 polymorphic AFLP fragments were scored across 114 oat varieties based on 8 AFLP primer combinations. Ten of these were designated as loci with rare alleles, because one allelic state (present or absent) was present in fewer than six oat varieties. For the remaining 77 loci, the average number of polymorphic fragments per primer combination was 9.6, ranging from 4 to 15 (Table 2). The DICE similarity between varieties ranged from 0.98 (difference in only one AFLP band) between Expo and Expander to 0.19 between Rodney and OT289 (difference in 24 AFLPs), with an overall mean over all variety pairs of 0.57. The two versions of Calibre, and the varieties Abel and Mozart, each differed by only two bands resulting in genetic similarities of 0.96 and 0.92, respectively. On the other hand, the two versions of Chernigovskij 27 differed at 12 AFLP markers (similarity = 0.78).

Cluster analysis

The UPGMA cluster tree is shown in Fig. 1. The upper part of the dendrogram consisted mainly of oat varieties from different parts of Europe, while the bottom part of the dendrogram contained varieties that were primarily from other parts of the world. In several cases, varieties from the same breeding program grouped in the same cluster branches, e.g. Event, Edo, Expo, and Expander from Saatzucht Edelhof in Austria. Similarly, the varieties Doris and Barra from Sweden clustered together, as did Aare, Puhti, Roope and Veli from Finland. Another example is the group of US varieties Brawn, Blaze, Dane and Chaps, and Ogle. The relative expression of the oat varieties for nine phenotypic

traits is shown as a heat map in greyscales in Fig. 1. Dark colours indicate high-numerical values for the trait, and light colours indicate low values. For example, the hulless varieties can be identified by their low values for screening percentage (white in the last column) and high values for groat percentage (dark in the second last column).

Principal component analysis

In order to gain further insight into the genetic diversity in this set of varieties, and to provide an alternate method for identifying structure, principal components analysis (PCA) was applied. The first four principal components (PC) explained 11.5, 6.8, 5.8 and 5.1% of the AFLP variation, respectively. Figure 2a shows the scatter-plot of PC1 versus PC2 including all 113 oat varieties. The plot illustrates that oat varieties from outside Europe were a great deal more diverse than the European oat varieties, which clustered mainly in the middle right section of the plot. Figure 2b shows a magnification the middle right section of Fig. 2a. Here, the majority of the European oat varieties, although from different parts of Europe, grouped together in relatively close proximity. No obvious clustering related to hull colour was evident, since white, yellow, and black oats all formed overlapping clouds, and even the hulless oat varieties did not appear as a separate group.

Discrete population structure

A model-based clustering method was used to infer population structure and to assign individuals to discrete populations. The variety Rodney was eliminated from this analysis due to missing scores, and three data sets were tested containing 113 varieties and 77, 73, or 60 loci (the later two with highly correlated markers removed, as described in the methods). The value of $Pr(X|K)$ was optimized at $K = 5$ for most numerical solutions in all three data sets (Figure S1). The membership of varieties in $K = 5$ subpopulations was highly consistent within each of the three data sets across multiple solutions (results not shown) and typical solutions produced memberships that were highly consistent among the three data sets of varying completeness (Table S2). The incomplete marker subsets were intended to observe the effect of removing pairs of markers that were potentially linked, which could bias the estimates of population structure. However, since no linkage data from segregating progeny were available, there is also a risk that some markers were highly correlated due to population structure, and that their removal may bias the estimate of this structure in a different way. Since group membership was not strongly dependent on removal of highly correlated markers, we chose to continue working with a set of groups formed based on the set of 113 varieties and 73 markers (column 2 in Table S2) as a compromise between these two potential biases. The resulting numeric Structure Coefficients are shown in Table 1 and in Fig. 1. These coefficients are in close agreement with the major clusters in Fig. 1.

Association analysis

Significant associations between markers and quantitative traits for three of the tested models are reported in Table 3. All of the markers reported in this table were significant at a comparison wise error rate of 0.0005 for at least one trait using at least one of the six detection methods, thus the significance criteria is protected using a Bonferroni correction factor for an experiment-wide error rate of 0.038. While this correction is overly conservative in accounting for multiple comparisons, it does not account for the presence of associations caused by population structure. In Table 3, the presence of significant tests in the second position (GLM + Structure Coefficient) or third position (MLM with kinship coefficients) indicates that the marker term in the model remains significant when these corrections are made. For example, the highly significant effect ($P < 0.001$) of marker S13/M25-85 on grain yield is eliminated ($P > 0.01$) in the two models that account for population structure, the effect of S13/M14-25 on lodging remains highly significant with a decreased R^2, and the effect of marker S18/M17-60 on screening percentage is relatively unaffected. In some cases (e.g. the effect of S24/M20-13 on lodging) the models that accounted for structure caused the marker to be significant when it was not significant in the naïve GLM model.

A more informative illustration of the effect of applying various corrections for population structure is shown by the R^2 plots in Fig. 3 and in supplementary Figure S3. Figure 3 shows R^2 values for all six models that were tested for two exemplary markers, while the supplementary figure extends this to additional markers using colour illustrations. These plots clearly show a reduction in R^2 in most models that account for population structure, and that this reduction is sometimes (but not always) strongest in the models that contain PCA effects (models 3 and 4) or kinship (models 5 and 6). For example, the effect of S12/M17-23 on both grain yield and groat yield (Fig. 3a) seems to fall but equilibrate at a consistent level in models 3 through 6 whereas the effect of marker S13/M20-66 on plant height (Fig. 3b) fluctuates, but is lowest in model 6.

Discussion

Genetic similarity

The high genetic similarity between the duplicated varieties Abel and Mozart, and Calibre(A) and -(B) indicated

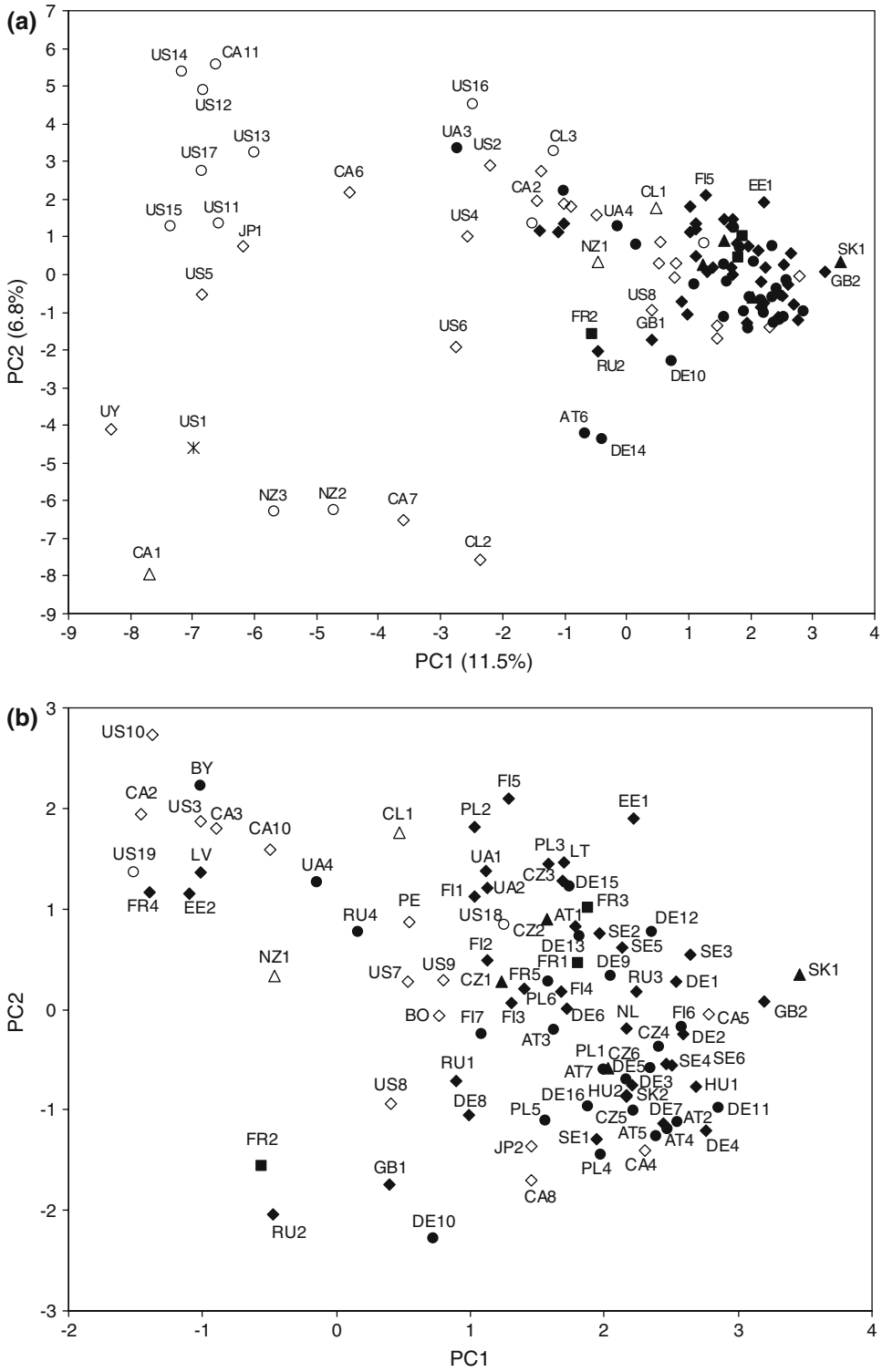


Fig. 2 Scatterplot of principal components 1 and 2 of 113 oat varieties based on genotypes of 77 AFLP markers. **a** All varieties, **b** magnification of the middle right part of **a**. Symbols: *circle* yellow hulls, *diamond* white hulls, *square* black hulls, *star* tan hulls, *triangle* hulless oat; European oat accessions are indicated by *filled symbols*, non-European oats by *empty symbols*. For each variety its identity is shown (see Table 1)

that these pairs are indeed genetically almost identical. Minor genetic differences between samples from different seed sources are not uncommon, since they can arise due to residual heterozygosity in the founding generation of the breeding material. However, the two versions of Chernigovskij 27 showed a greater number of genetic differences than did many pairs of distinct varieties, therefore these two entries probably represent varieties of different origin or breeding history. These results are in good agreement with the results obtained for agronomic and quality traits on the same plant material (Buerstmayr et al. 2007).

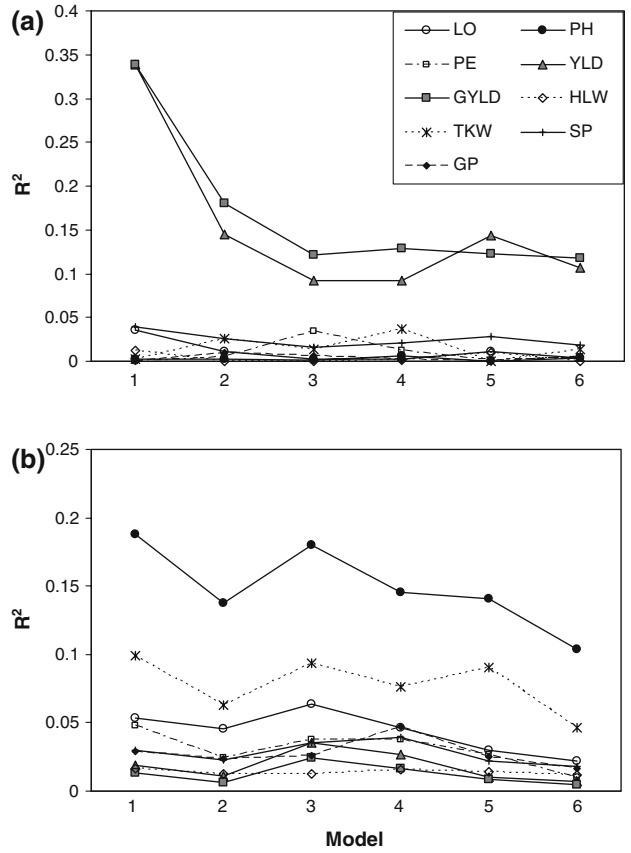


Fig. 3 Two R^2 plots: **a** for marker S12/M17-23 and **b** for marker S13/M20-66. These plots show the estimated proportions of phenotypic variance (partial $R^2 \times 100\%$) explained by a given marker for nine quantitative traits identified in six models of association analysis. The nine traits are shown by different line and symbol combinations. Six models of association mapping are shown from left to right: *1* Naïve general linear model (GLM) with no correction for population structure; *2* GLM with Structure Coefficient as cofactor; *3* GLM with PCA1 as covariate; *4* GLM with PCA1–PCA4 as covariates; *5* Mixed Linear Model (*MLM*) as described by Yu et al. (2006); *6* MLM with Structure Coefficient. Supplementary Figure S3 shows coloured graphs for additional markers

Cluster analysis

The UPGMA dendrogram grouped the oat varieties mainly according to geographical origin. As expected, frequently varieties from the same breeding program grouped in the same cluster branches, e.g. Event, Edo, Expo, and Expander from Saatzucht Edelhof in Austria, which were all selected from a Flaemingsnova cross. Likewise, the group of the US varieties Brawn, Blaze, Dane and Chaps appeared in the same cluster branches with their common ancestor Ogle. The varieties Jumbo from Germany and Monarch from Austria were also closely related. Monarch most likely has Jumbo or a sister line from Jumbo in its pedigree (Anton Neumayer, personal communication). However, the clustering also revealed some genetic similarities that were not previously suspected, such as the similarity between Baijka from Poland and Kermit from Germany. The heat map of the traits facilitates visual inspection of whether phenotypes relate to structure, as measured by the cluster analysis. For example, three of the hulless lines (Abel, Mozart, and Devtan) are clearly related, while the remainder (CROA 60, AC Belmont, Eva 1, and Akt) appear unrelated. Unlike in the study in *Arabidopsis thaliana* presented by Zhao et al. (2007), no strong tendency toward clustering of phenotypes was evident in this analysis.

Principal component analysis

Principal component analysis showed that the AFLP diversity within the European oat germplasm is comparatively narrow compared to the diversity in the germplasm from outside Europe. Within the European oat germplasm a few cultivars from Eastern Europe were somewhat more diverse. No clustering related to hull colour was evident. These findings are in good agreement with Fu et al. (2005) who analysed a world collection of 670 oat accession with AFLP markers. They found that oat accessions from Russia and the USA were most diverse and that only red oats formed a distinct group, but no clear delineation between the distributions of common and hulless oats was evident.

Discrete population structure

Model based clustering suggests that a large amount of the allelic diversity can be described by subdividing the varieties into 5 discrete subpopulations, where each subpopulation has a unique set of allele frequencies. This is clearly a simplification of the observed data, and obviously not representative of all exchanges and crosses that led to the development of these varieties. However, this apparent structure can be used to compare with other methods of clustering, and to test models of association analysis that would account for genetic associations arising from its presence. The five discrete subpopulations were in close agreement with the major clusters based on DICE similarity (Fig. 1), indicating that the structure in this population is revealed fairly consistently based on a variety of clustering methods.

Implications of clustering and diversity

UPGMA cluster analysis, principal coordinate analysis and STRUCTURE analysis separated the oat varieties into groups that were related to geographical origin.

Due to the narrow separation among varieties of European origin, it is concluded that a rather small proportion of the available genetic variation in this species is currently used for oat improvement within the majority of the European oat

Table 3 Percentage of phenotypic variance (partial $R^2 \times 100\%$) for over-all means of nine quantitative traits explained by markers identified in significant models based on three methods of association analysis: GLM (first numeral), GLM with PCA1-PCA4 as covariates (second numeral) or mixed linear model (MLM: third numeral)

Locus	PE			PH			LO			YLD			GYLD			TKW			HLW			GP			SP		
S13/M14-25				8	**14**	6	**34**	**25**	**23**	**16**	–	**6**	**13**	–	**4**										9	**10**	7
S13/M14-27							**15**	**9**	**9**	**21**	**7**	**7**	**18**	5	**6**	**10**	**13**	9							**17**	**14**	**15**
S24/M20-3										–	6	–	–	**7**	–												
S24/M20-4	7	–	7							9	–	–							**11**	–	**11**						
S24/M20-6										9	–	–															
S24/M20-9	**27**	**11**	**23**	9	**11**	6	–	5	–	9	–	–	7	–	–												
S24/M20-10	6	–	6							**12**	–	4	9	–	–										–	6	–
S24/M20-13							–	**10**	**5**							7	–	**10**							–	8	–
S12/M17-23										**34**	**9**	**14**	**34**	**13**	**12**												
S23/M20-39										**11**	–	–	8	–	–												
S23/M20-40							9	–	4	**16**	–	4	**10**	–	–												
S23/M20-41	–	–	4							10	–	3	7	–	–												
S23/M20-45	6	–	–				**13**	–	**7**	**31**	6	**12**	**25**	5	**6**												
S13/M25-80										7	–	–	**11**	–	–												
S13/M25-82										7	–	3	–	–	**3**												
S13/M25-85										**13**	–	–	**12**	–	–												
S12/M19-34	6	–	–													9	**13**	8									
S13/M20-66				**19**	**15**	**14**										**10**	8	9									
S13/M20-69				9	–	**7**																–	**11**	–			
S13/M20-74	**10**	–	7																								
S18/M17-50							7	–	–	**14**	–	4	**10**	–	–										8	–	6
S18/M17-59	**11**	–	7																**12**	–	**12**						
S18/M17-60										8	–	–				–	**9**	–				8	8	7	**12**	**11**	**11**

Mean values of quantitative traits from Buerstmayr et al. (2007)

Markers shown in the table were significant at a comparison wise rate of 0.0005 for at least one trait using at least one of the six detection methods (see Fig. 3 and methods section). The R^2 values shown correspond to the significance of AFLP at $P < 0.01$ (regular font) or $P < 0.001$ (in bold). A pair of dashes (–) indicates that the model represented by the numeral in this position did not meet the $P < 0.01$ criterion

Trait abbreviations: *PE* panicle emergence, *PH* plant height, *LO* lodging severity, *YLD* grain yield, *GYLD* groat yield, *TKW* thousand kernel weight, *HLW* hectolitre weight, *GP* groat percentage, *SP* screenings percentage

germplasm. The reasons for this may be that (1) European oat breeders have performed crosses mainly within European germplasm and/or (2) adaptation to European conditions is conditioned by many loci, such that selection often excluded the incorporation of new alleles when wider crosses were made. Since the AFLP loci surveyed in this study are likely to represent a broad and random cross section of the oat genome, it seems more likely that diversity has been restricted by choice of crossing parents, and that increased diversity in European material could be achieved without loss of regional adaptation. Fu et al. (2004) found a similar lack of diversity within a set of Canadian oat varieties, and identified an urgent need to broaden the genetic variation for sustainable oat improvement in Canada. Similar results were reported by Baohong et al. (2003) for Chinese oat accessions.

Improved varieties from other parts of the world showed much wider genetic variation based on AFLP fingerprints in this study. For instance, varieties from North and South America showed greater variation than European varieties, an observation that is consistent with that of Fu et al. (2005). The superior genetic diversity in North American spring oats may result from a more frequent use of exotic varieties (other species or ecotypes) in breeding, as reported by Rodgers et al. (1983). Buerstmayr et al. (2007) identified several oat varieties from North America that showed characteristics such as earliness and physical grain quality that would be valuable in European varieties. Importantly, these characters were measured under a variety of replicated European environments, and they showed reasonable agronomic adaptation to those environments. The increased use of such parents in European breeding programs could simultaneously increase diversity and improve levels of valuable traits such as grain quality. From crosses of genetically divergent parents (e.g. high a yielding European variety crossed with an introduced parent showing

superior grain quality), novel varieties with improved physical grain quality and high yield potential may be selected.

Association analysis

The molecular marker data from the present study, in combination with data reported by Buerstmayr et al. (2007), provided an opportunity to examine preliminary evidence for linkage-related marker-trait associations. Since mapping data for the AFLP markers were not available, it was not possible to examine the structure of disequilibrium among linked markers. Therefore, we cannot yet speculate on the degree to which disequilibrium extends between linked markers. However, with the exception that estimates of population structure may be partially biased by clustered markers, the methods applied in this study are independent of linkage analysis.

Separating the role of population structure *vs*. the role of genetic linkage as causes for marker-trait association remains the greatest challenge in association analysis. The six models used in this study can generally be described as accounting for 'Q' (population structure that results from the existence of sub-populations) or for 'K' (general similarity in genetic background arising from shared kinship). Thus, the model containing the Structure Coefficient is primarily a 'Q' model, while the MLM model is a 'K' model. Both factors may be important in causing associations between markers and phenotype that are not related to genetic linkage between markers and QTLs, and both factors are partially related. For this reason we have also tested a model containing both factors, as well as models containing PCA covariates, which may account for some proportion of both 'Q' and 'K'.

Although there is no definitive test for whether an association is due to population structure or genetic linkage, a co-examination of different models and different traits using R^2 plots (Fig. 3 and supplementary Figure S3) can provide an informative summary of the major trends affecting an analysis. Differences in results among the models may illustrate the relative importance of different parts of the population structure accounted for by different models. In some cases, the combination of both 'Q' and 'K' in the sixth model provided the strongest reduction in R^2, and presumably, the best correction for population structure. However, this was not always the case, as illustrated in Fig. 3a. Surprisingly, the addition of three PCA axes (model 4 vs. 3) often did not change the R^2 substantially. This suggests that the first PCA axis accounts for most of the population structure related to phenotypic variance. The inclusion of multiple traits in these plots provides a built in control for significance. Although there may be occasions when all traits are affected by the same marker, there were generally several traits that were unaffected, and these provide a visual 'baseline' for examining the relative importance of these effects. Thus, the R^2 plot provides a useful supplement to formal significance tests which merely identify the presence of non-random associations.

While further validation is required, the markers showing strongest effects in this study provide ideal candidates for further study or future inclusion in strategies of marker assisted selection. Future studies in mapping populations or in expanded sets of oat varieties will help to corroborate whether the QTL effects implicit in these marker-trait associations are robust and useful in a practical breeding program, and further efforts in comparative mapping may help to identify whether any of these effects are caused by orthologs of known genetic factors. However, as there will likely be crosses made among some of the varieties investigated in this study, it is worthwhile to speculate on some of the possibilities for marker assisted selection revealed by these experiments.

Based on the traits affected, and the size of the corrected R^2 values, there are seven markers that we consider to be the most interesting candidates for further work. These are S13/M14-25, S13/M14-27, S24/M20-9, S12/M17-23, S23/M20-45, S13/M20-66, and S18/M17-59. For example, the strong effect of S13/M14-25 on lodging could provide a useful target for marker assisted selection, because the lodging phenotype is not reliably expressed in a breeding program. Effects on plant emergence and plant height may be less useful for marker-based selection because they are traits that are easily measured. However, the validation of QTL for these traits can assist with functional genomic studies or can facilitate allele mining for new sources of variance. Marker S12/M17-23, which affects grain yield, would provide another obvious target to follow. Interestingly, this marker effect is highly consistent with the effect on groat yield, despite the fact that the groat yield estimates include an additional set of hulless lines. Numerically, varieties that have the null allele at this locus yielded an average of 12 dt ha^{-1} more seed or 8 dt ha^{-1} more groats than those with the 'plus' allele. Since most of the lines containing the plus allele are easily identified as poor yielding, the role of a marker such as this would be most relevant in a strategy that involved introgression of other traits from these lines, and the yield-suppressing allele would be subjected to negative selection. Due to the high value of yield, and the difficulty of selecting this trait in early generations, further validation of the markers affecting yield would be very useful.

As with any type of QTL analysis, the estimates of QTL for multiple traits can provide evidence for genetic causes of correlations among traits, and may allow undesirable correlations to be partially broken. The grain yield and groat yield were most highly correlated, with much commonality among QTL. However, some potentially expected

correlations were not present. Notably, there were no common QTL-related markers between plant height and lodging except for S13/M14-25. Although grain yield and groat percentage were positively correlated with $r = 0.32$ (Buerstmayr et al. 2007), no common significant markers were detected for these two traits.

One of the issues that will need to be determined is whether the linkage distances between the markers and the QTL are small enough that the markers can be used reliably for selection. Other factors could also affect the utility of these markers, including the possibility that the associated QTL are epistatic, or that they are highly dependent on one environment. This would seem unlikely because association genetics ought to be most effective in detecting QTL that show additive effects, and because the phenotypes used in this study are based on mean values across four environments. While the tested environments are not representative of all oat growing areas, we do speculate that the grain quality traits are least dependent on the production environment.

Conclusion

We now know a substantial amount about the phenotypes, genotypes, genetic diversity, and population structure for this set of oat lines from diverse global origins, and we are now aware of the limited amount of diversity in European germplasm compared to what is available from other origins. This study also provides one of the first reported investigations of association analysis in a diverse population of oat, and thus, it will provide a useful benchmark for comparison with future results and with results from other species. Although population structure is an important cause of marker trait associations in oat, we now have good evidence that linkage-related associations are also important, and we have some useful estimates of which markers may be important to follow when attempting to introgress genes from more exotic germplasm to broaden the genetic diversity of oat.

Acknowledgments We gratefully acknowledge all colleagues and institutions that contributed germplasm to this study: Keith Armstrong (Crop & Food Research, NZ), S. Bous (Pflanzenzucht Oberlimpburg, DE), Erik van Baer, (Semillas Baer, CL), Ken Daniels (Power Seeds, Ireland), Heidi F. Kaepler (Univ. Wisconsin, US), Fredrik L. Kolb (Univ. Illinois, US), Tatsuo Kuwabara (Hokaido, JP), Bengt Mattson (Svalöf Weibull, SE), Jennifer Mitchell Fetch (Winnipeg, CA), James B. Holland (Ames, Iowa, US), Hywel Roderick (Aberysthwyth, UK), Brian Rossnagel (Univ. Saskatchewan, CA), Marketta Saastamoinen (Boreal Plant Breeding, FI), Mark E. Sorrells (Cornell Univ., US), Maria H. Starzyk (Roslin, PL), Deon Stuthman (Univ. Minnesota, US), Horst Vasel (Wentzelsche Saatzucht, DE), Cebeco Zaden (NL), Institute of Applied Genetics and Breeding (Piestany, SK), Lochow-Petkus GmbH (Bergen, DE), Selgen Plant Breeding (Praha, CZ), Serasem (FR), Centre for Plant Genetic Resources of Ukraine (UA). We apologise for any omissions in the above list. We gratefully acknowledge the support of the research work of IFA-Tulln by the Government of Lower Austria.

References

Aufhammer G, Fischbeck G (1973) Getreide – Produktionstechnik und Verwertung. DLG Verlags GmbH, Frankfurt

Baohong G, Zhou X, Murphy JP (2003) Genetic variation within Chinese and Western cultivated oat accessions. Cereal Res Comm 31:339–346

Beer SC, Siripoonwiwat W, O'Donoughue LS, Souza E, Matthews D, Sorrells ME (1997) Associations between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages. J Ag Genomics (http://wheat.pw.usda.gov/jag/) 3:item 1. Cited 8 Jan 2008

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Buerstmayr H, Krenn N, Stephan U, Grausgruber H, Zechner E (2007) Agronomic performance and quality of oat (*Avena sativa* L.) genotypes of worldwide origin produced under Central European growing conditions. Field Crops Res 101:343–351

Buerstmayr H, Lemmens M, Hartl L, Doldi L, Steiner B, Stierschneider M, Ruckenbauer P (2002) Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I. Resistance to fungal spread (Type II resistance). Theor Appl Genet 104:84–91

Dice LR (1945) Measures for the amount of ecologic association between species. Ecology 26:297–302

Food and Drug Administration (1997) Food labeling: health claims; oats and coronary heart disease; final rule. Fed Regist 62:3583–3601

Forsberg RA, Reeves DL (1992) Breeding oat varieties for improved grain quality. In: Marshall HG, Sorrells ME (eds) Oat science and technology. Am Soc Agron, Madison, pp 751–775

Fu YB, Peterson GW, Scoles G, Rossnagel B, Schoen DJ, Richards KW (2003) Allelic diversity changes in 96 Canadian oat varieties released from 1886 to 2001. Crop Sci 43:1989–1995

Fu YB, Kibite S, Richards KW (2004) Amplified fragment length polymorphism analysis of 96 Canadian oat varieties released between 1886 and 2001. Can J Plant Sci 84:23–30

Fu YB, Peterson GW, Williams D, Richards KW, Mitchell Fetch J (2005) Patterns of AFLP variation in a core subset of cultivated hexaploid oat germplasm. Theor Appl Genet 111:530–539

Gupta PK, Sachin R, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Hartl L, Mohler V, Zeller FJ, Hsam SLK, Schweizer G (1999) Identification of AFLP markers closely linked to the powdery mildew resistance genes *Pm1c* and *Pm4a* in common wheat (*Triticum aestivum* L.). Genome 42:322–329

Hoffmann LA (1995) World production and use of oats. In: Welch RW (ed) The oat crop-production and utilization. Chapman and Hall, London, pp 34–61

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Li CD, Rossnagel BG, Scoles GJ (2000) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat varieties. Theor Appl Genet 101:1259–1268

Mohammadi SA, Parasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci 43:1235–1248

O'Donoughue LS, Souza E, Tanksley SD, Sorrells ME (1994) Relationships among North American oat varieties based on restriction fragment length polymorphisms. Crop Sci 34:1251–1258

Paczos-Grzeda E (2004) Pedigree, RAPD and simplified AFLP-based assessment of genetic relationships among *Avena sativa* L. varieties. Euphytica 138:13–22

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

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Rodgers DM, Murphy JP, Frey KJ (1983) Impact of plant breeding on the grain yield and genetic diversity of spring oats. Crop Sci 23:737–740

Rohlf FJ (2000) NTSYS-pc 2.11: numerical taxonomy and multivariate analysis system. Exeter Publishing Ltd., Setauket

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Souza E, Sorrell ME (1989) Pedigree analysis of North American oat varieties released from 1951–1985. Crop Sci 29:595–601

Souza E, Sorrells ME (1991a) Relationships among 70 North American oat germplasm: I. cluster analysis using quantitative characters. Crop Sci 31:599–605

Souza E, Sorrells ME (1991b) Relationships among 70 North American oat germplasm: II. cluster analysis using qualitative characters. Crop Sci 31:605–612

Tinker NA, Deyl JK (2005) A curated internet database of oat pedigrees. Crop Sci 45:2269–2272

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4