UNIVERSITÄT HOHENHEIM

Institute of Plant Breeding, Seed Science and Population Genetics (350)
Fruwirthstr. 21
D-70593 Stuttgart
PD Dr. sc. agr. Jochen C. Reif

Genetic Diversity of worldwide Bread Wheat Landrace Accessions

Diploma-Thesis

Marlen Hübner
Altenhainerstrasse 16
D-65719 Hofheim am Taunus

TABLE OF CONTENTS

ABBREVIATIONS

CAAS	Chinese Academy of Agricultural Sciences
CIMMYT	International Maize and Wheat Improvement Center
GCP	Generation Challenge Program
ICARDA	International Center for Agricultural Research in the Dry Areas
INRA	National Agriculture Research Institute
LCs	landrace cultivars
MAS	marker assisted selection
MRD	modified Rogers´ distance
PCoA	principal coordinate analysis
QTL	quantitative trait locus
SSR	simple sequence repeat

Countries

AFG	Afghanistan
AZB	Azerbaijan/Armenia/Georgia
CHE	Switzerland
CHN	China
DEU	Germany/Netherlands
EGY	Egypt
ESP	Spain/Portugal
ETH	Ethiopia
FRA	France
GRC	Greece
HUN	Hungary/Romania
IND	India
IRN	Iran/Iraq
ISR	Israel

ITA	Italia
PAK	Pakistan
RF	Russian Federation
TUN	Tunisia
TUR	Turkey
UZB	Uzbekistan/Kyrgyzstan
YUG	Yugoslavia

ABSTRACT

Many wheat (*Triticum aestivum* L.) landrace cultivars (LCs) conserved in seed banks represent a valuable genetic resource for breeding. Patterns of genetic variation of wheat LCs are usually not sufficiently characterized. For adapting on current global developments on wheat market and improving food security the entire spectrum of available wheat germplasm must be used for increased wheat production. The objectives in our study were to (i) investigate the genetic diversity among hexaploid wheat landraces using simple sequence repeat (SSR) markers, and (ii) compare genetic diversity of different geographic data. A worldwide set of 232 wheat LCs were fingerprinted with 32 simple SSRs. For analyzing the genetic variation two subsets of LC accession were created: Set 1 included all 232 LCs and Set 2 included 118 LCs accessions with complete data of latitude, longitude and elevation. Among the accessions of Set 1 we detected a total number of 1107 alleles. In average we found gene diversity from 0.45 in Africa to 0.69 in Europe. In the STRUCTURE analysis for Set 1 as well in the principal coordinate analysis based on modified Rogers´ distance (MRD) in Set 2 we found no evidence of a clear grouping among the LCs. For Set 1, a dendrogram was constructed using UPGMA cluster analyses method to assess a pattern of the LCs. Two major branches derived from the whole set of individuals. Countries from Asia grouped direct together in one major branch. Countries from Africa, Middle East and Europe are distributed in both major branches. We conclude that SSRs can be used to assess the genetic variation as well as a tool for seed bank management. A drawback for assessing historical relationships of LCs is the lack of of complete datasets.

INTRODUCTION

Wheat contributes to around one quarter of the global human consumption of calories, for which there are no easy substitutes in many major wheat consuming countries. As the most internationally traded food crop, wheat is the single largest food import in developing countries and a major portion of emergency food aid (Dixon et al., 2008). For adapting on current global developments on wheat market and improving food security the entire spectrum of available wheat germplasm in conjunction with the newest technologies in plant science must be used for increased wheat production.

Genetic drift and selection have been reduced genetic variations of wheat germplasm across its evolution (Reif et al., 2005). Major circumstances for the reduction in variability are assumed to be bottleneck effects during the evolution of wheat which is based on interspecific hybridization, natural and especially intense human selection during the last centuries, as well as genetic erosion due to increasing agricultural systems. Crop improvement is currently based on a narrow gene pool of elite materials in the various wheat breeding programs. Historical wheat materials such as wild species or landraces have become isolated from the mainstream gene pools. Major risk of a narrow genetic base in germplasm cultivation is genetic vulnerability (Reif et al., 2005).

Specific pre-breeding programs need to be established to recover the loss of genetic variation. Instead of exploiting wild species, a more direct way to exploit novel allelic diversity is to cross elite material with genetic resources of the same genome, e.g. accessions of landrace cultivars (LC). LCs represent an elementary and tremendous source of genetic variation in wheat which is a basic requirement for the effort to transfer desired genes from mostly unadapted material into advanced breeding lines (Skovmand and Rajaram, 1990). Major genes, such as the classic dwarfing genes (Kihara, 1982), but also several disease resistance genes (Yu et al., 2007, Nazari et al., 2008), genes for improved

quality (Li et al., 2008) and abiotic stress tolerance (Moragues et al., 2008) could be observed in LC sources and leading to genetic gain in wheat improvement. The evaluation of the genetic variation present in LCs is therefore of utmost importance to enlarge the genetic base in breeding (Stodart et al., 2005). In addition, the evaluation of LCs is useful for their future conservation in seed banks.

Thousands of wheat LCs are already stored in seed banks worldwide but are mostly inadequately described especially for an efficient practical application in plant breeding (Dreisigacker et al., 2005). Besides field evaluations, marker techniques such as Microsatellites are very useful for detailed analyses of genetic variation. Molecular markers provide a direct measure of genetic diversity and go beyond indirect diversity measures based on agronomic traits or geographic regions. The SSR technique achieved fast acceptability because of its co-dominant nature, reproducibility, and high information content (De Loose et al., 1995). SSRs show a much higher level of polymorphism and are more informative in hexaploid wheat than any other marker (Röder et al., 1998). They are plentiful, independent of tissue or environmental effects, and allow cultivar identification early in plant development (Manifesto et al., 2001). The marker-based assessment of novel genes eliminates the need of extensive test crossing (Raman et al., 2003).

To our knowledge no study analyzed worldwide wheat LCs regarding to their geographical data like latitude, longitude and elevation. Therefore in this present study we survey genetic structure of 232 bread wheat LCs by SSRs generally and concerning to their passport data. The objectives in our study were to (i) investigate the genetic diversity among hexaploid wheat landraces using SSR marker, and (ii) compare genetic diversity of different geographic data.

MATERIALS AND METHODS

Plant Material

The LCs included in this study represented an almost complete spectrum of bread wheat LCs in the world excluding the American continent and Oceania. The in total 232 LCs have been part of a composite collection of 3000 accessions representing a high range of diversity in wheat and its wild relatives. The development of the composite collection was commissioned by the Generation Challenge Program (GCP) and accomplished by a collaboration of four research institutions: The International Maize and Wheat Improvement Center (CIMMYT), Mexico; the National Agriculture Research Institute (INRA), France; the Chinese Academy of Agricultural Science (CAAS), China, and the International Center for Agricultural Research in Dry Areas (ICARDA), Syria. Various LC accessions from four different continents or geographical areas (Asia, Europe, Africa, and Middle East) were chosen by the four collaborating institutions based on seed bank availability. Information on accessions was limited but included country of origin and geographical data such as elevation, latitude and longitude. The countries of origin included five countries in Asia (AFG, CHN, IND, PAK, RF), three countries in Africa (EGY, ETH, TUN), eight countries in Europe (CHE, DEU, ESP, FRA, GRC, HUN, ITA, YUG) and five countries in the Middle East (TUR, UZB, IRN, ISR, AZB) (Table 2).

For analyzing the genetic variation two subsets of LC accession where created: Set 1 included all 232 LCs and Set 2 included 118 LCs accessions with complete data of latitude, longitude and elevation. Elevation ranged from four to 3230 m above sea level.

SSR Analyses

DNA of accessions within the composite collection was extracted by each providing institute and was subsequently exchanged across institutions to avoid the need of subsequent data integration which often provides errors.

For DNA extraction at CIMMYT and similar at the three other institutes, leave tissue was harvested from five to 10 plants per accession. Genomic DNA extraction was performed with a modified CTAB (cetyltrimethylammonium bromide) method according to CIMMYT Applied Biotechnology Center´s Manual of Laboratory Protocols (http://www.cimmyt.org/english/wps/publs/Catalogdb/). Quality and quantity of the isolated DNA was determined on 1% (w/v) Agarose gels by comparing bands to known concentrations of λ DNA.

Marker analysis was also completed at the four research institutions participating in the commissioned GCP project. Information of the applied SSRs and final genotypic data were collated and stored online within the GCP central registry (http://gcpcr.grinfo.net/). At CIMMYT 20 SSRs, at CAAS five SSRs, at INRA eight SSRs and at ICARDA three SSRs were applied and scored across the entire set of accessions.

Polymerase change reactions (PCR) at CIMMYT were performed in a model MJ Research DNA Engine Tetrad™ System Thermocycler (MJ Research, Inc., Waltham, MA). Each reaction mixture contained 50 ng template DNA, 0.24 µM of each primer (forward, reverse), 200 µM dNTPs, 1.5 mM $MgCl_2$, 1x PCR buffer and 1U of *Taq*-polymerase. PCR was performed with the following temperature profile: 30 cycles consisting of 1 min denaturation at 94°C, 2 min annealing using temperatures between 54 and 62°C (depending on primer combination), and 2 min extension at 72°C. Primers were labeled with TET, HEX and FAM fluorescent dyes.

Amplification products were separated using an ABI PRISM™377 Sequencer (Perkin Elmer/ Applied Biosystems, Foster City, CA). Regarding to CIMMYT Applied Biotechnology Center´s Manual of Laboratory Protocols we used 4.5% (w/v) polyacrylamide denaturing gels (acrylamide:bisacrylamide 29:1). Running

conditions were 2400 V, 40mA, 120 W electrophoresis power and 40 mW laser power. Products from up to five SSRs could be distinguished simultaneously because of three different fluorescent dyes and migration distance differences. Fragment size were calculated semiautomatically by the computer software GeneScan 3.1 (Perkin Elmer/ Applied Biosystems) by comparing fragments with an internal size standard (GenScan 350 or 500) labeled with N,N,N,N,-tetramethyl-6-carboxyrhodamine. GeneScan fragments were assigned to alleles by the category function of the software Genotyper 2.1 (Perkin Elmer/Applied Biosystems) (Dreisigacker et al., 2004). For loading the amplification products we mixed 0.5- 1.0 µl of pooled PCR products with 8 µl of formamide:size standard mix. For preparing the formamide:size standard mix we used 1000 µl of Hi-Di™ formamide (Applied Biosystems) and 30 µl GS 350 or GS 500 labeled with ROX.
At the other three institutions diverse genotyping systems were applied. At ICARDA and CAAS amplification products were separated using an ABI PRISM®3100 sequencer. At INRA electrophoresis was performed with a Licor IR@ sequencer**.**

Statistical Analyses

Summary statistics

For calculating gene diversity, alleles per locus and group specific alleles in Set 1 for every continent, the software package PowerMarker (Liu, 2005) was used. Gene diversity was calculated at each locus as:

$$2n\,(1-\sum_{u}p_u^2)/(2n-1-f)$$

where p_u is the frequency of the *u*th allele, *n* is the sample size, and *f* is the inbreeding coefficient estimated from genotype frequencies (Weir, 1996). To calculate the correlations between elevation with latitude and longitude in Set 2, we used a Mantel test (Mantel, 1967) by setting the permutation number to 1000. The Modified Rogers' distance (MRD) was calculated between individuals, countries and continents in Set 1 according to the following equation (Wright, 1978):

$$MRD = \sqrt{\frac{1}{2m}\sum_{i=1}^{m}\sum_{k=1}^{a_i}(p_{ij} - q_{ij})^2}$$

whereas p_{ij} and q_{ij} are the allele frequencies of the *j*th allele at the *i*th marker; a_i refers to the number of alleles at the *i*th marker; and *m* is the number of SSRs (Dreisigacker et al., 2005). Standard errors of the MRD estimates were obtained by a bootstrap procedure with resampling 1000 times across marker and individuals (Weir, 1996).

Based on MRD values, principal coordinate analysis (PCoA) were performed to visualize the dispersion of genotypes in Set 2 (Gower, 1966). Clustering the LCs with the UPGMA method for Set 1, PCoA and Mantel test for Set 2 were performed with the statistical software Plabsoft (Maurer et al. 2008).

Population structure analysis

For the analysis of population structure and genetic relationships of Set 1, LC accessions were subdivided into genetic clusters using a model- based approach implemented in the software package STRUCTURE (Pritchard et al., 2000). A number of population *k* (cluster) was assumed to be present and to contribute to the genotypes of the sample individuals.

At least five runs of STRUCTURE were done by *k* and the number of *k* cluster was varied from 3 to 10. For each run, burn- in time was set to 50,000 and replication number to 100,000. The run with the maximum likelihood was used to assign LCs in clusters. Therefore, we used the *ad hoc* criterion described by Evanno et al. (2005) to estimate the number of subpopulations, as it promises to reliably detect the true number of subpopulations also in complex genetic situations (Stich et al., 2007). Our main modeling assumptions were Hardy-Weinberg equilibrium within populations and complete linkage equilibrium between loci within our LC population. Under these assumptions each allele at each locus in each genotype is an independent draw from the appropriate frequency distribution (Pritchard et al., 2000).

RESULTS

All 32 marker loci analyzed were polymorphic across the 232 individuals in Set 1. Among the accessions we detected a total number of 1107 alleles. Less than 10% were missing data, true null alleles, or failed PCR amplifications. The average number of alleles over continents was 8.4, with a range over continents from 7.2 (Africa) to 10.5 (Asia). On average of 23.5 group specific alleles was observed whereby one in Africa, eight in Europe, 10 in Middle East, and 75 in Asia. Gene diversity ranged from 0.45 in Africa to 0.69 in Europe (Table 1).

Table 1. Sample size, alleles per locus, group specific alleles and gene diversity for each of the four continents in Set 1.

Statistic	Middle East	Asia	Europe	Africa
Sample size	34	140	40	18
Alleles per locus	8.2	10.5	7.6	7.2
Group specific alleles	10	75	8	1
Gene diversity	0.67	0.66	0.69	0.45

We found correlations significantly different from zero ($P < 0.001$) between distances of (i) elevation and latitude (r = 0.2362) and (ii) elevation and longitude (r = 0.08942) in Set 2 using the Mantel test.

The first two principle coordinates (PC) performing PCoA based on MRD estimates explained a total of 12.81% (Appendix), for the countries a total of 40.17% and for the continents 77.45% of the molecular variance (data not shown). PCoA did not reveal a clear grouping among the LCs. Just a few Asian LCs reside outside of the scatter plot.

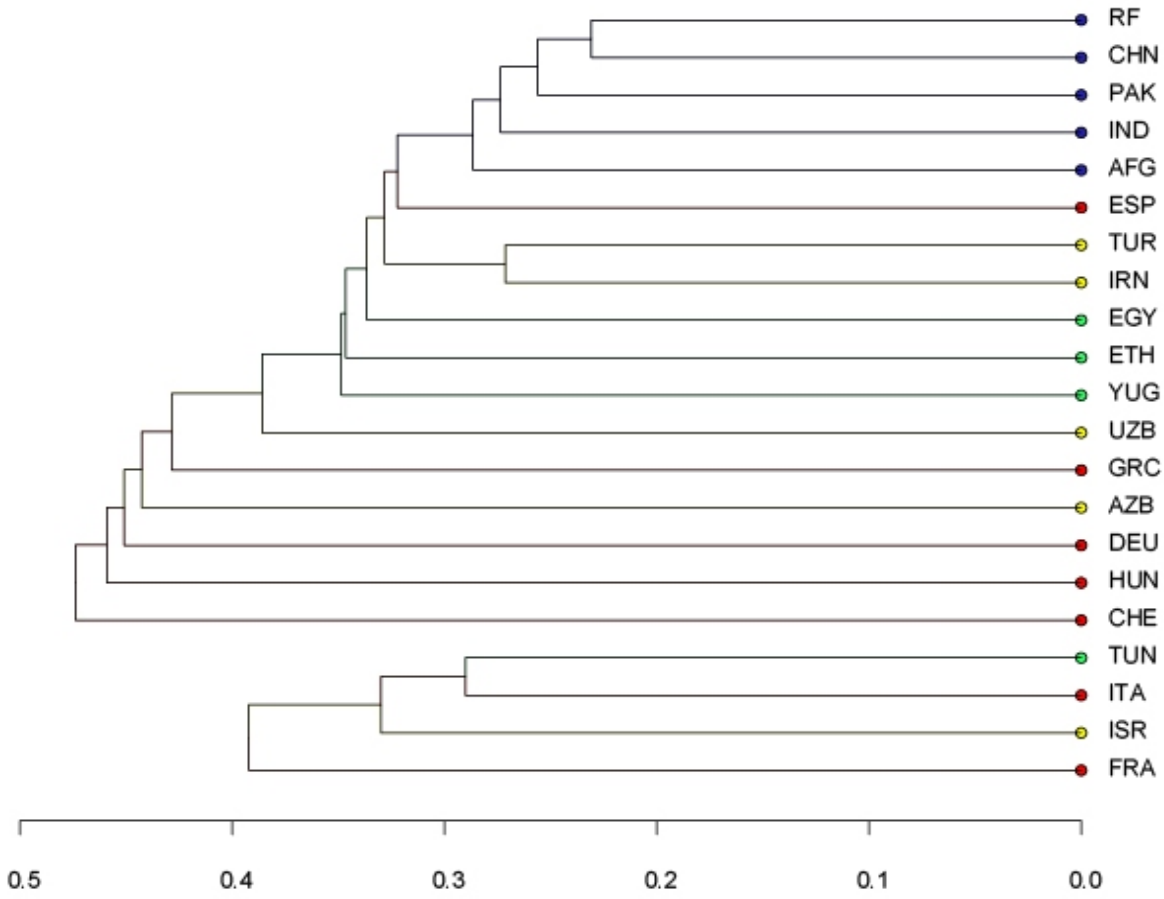

Fig. 1 UPGMA cluster of Set 1 (all 232 LCs) based on Modified Rogers´ distances between the countries; blue: Asia, red: Europe, yellow: Middle East, green: Africa.

Most closely related to each other are LCs from neighboring countries CHN and RF, TUR and IRN, and TUN and ITA. In this tree, geographic patterning reveals of the LCs. Two major branches were derived from the whole set of individuals. The smaller branch is showing a separation of four countries, FRA, ISR, ITA and TUN. The second major branch consists of the LC accessions in the residual countries. Within this branch geographic patterning of all countries from Asia (RF, CHN, PAK, IND and AFG) is seen. LC accessions from EGY and ETH are grouped in one clade whereas TUN, the third African country, is grouped close to ITA in the

other small major branch of the tree. LC accessions from DEU, HUN and CHE are showing a close grouping whereas ESP, YUG, UZB, GRC and AZB are grouped also in this big major branch but not in a clear geographic pattern. Within each continent or geographical regions countries group just for Asia direct together in one major branch. Countries from Africa, Middle East and Europe are divided in both major branches.

By setting k = 10 in the model based clustering we detected eight subpopulations as the true number in Set 1. For every continent individuals from the same country are connected one after another but the individuals of the whole LCs are not showing a clear cluster (Fig. 2)

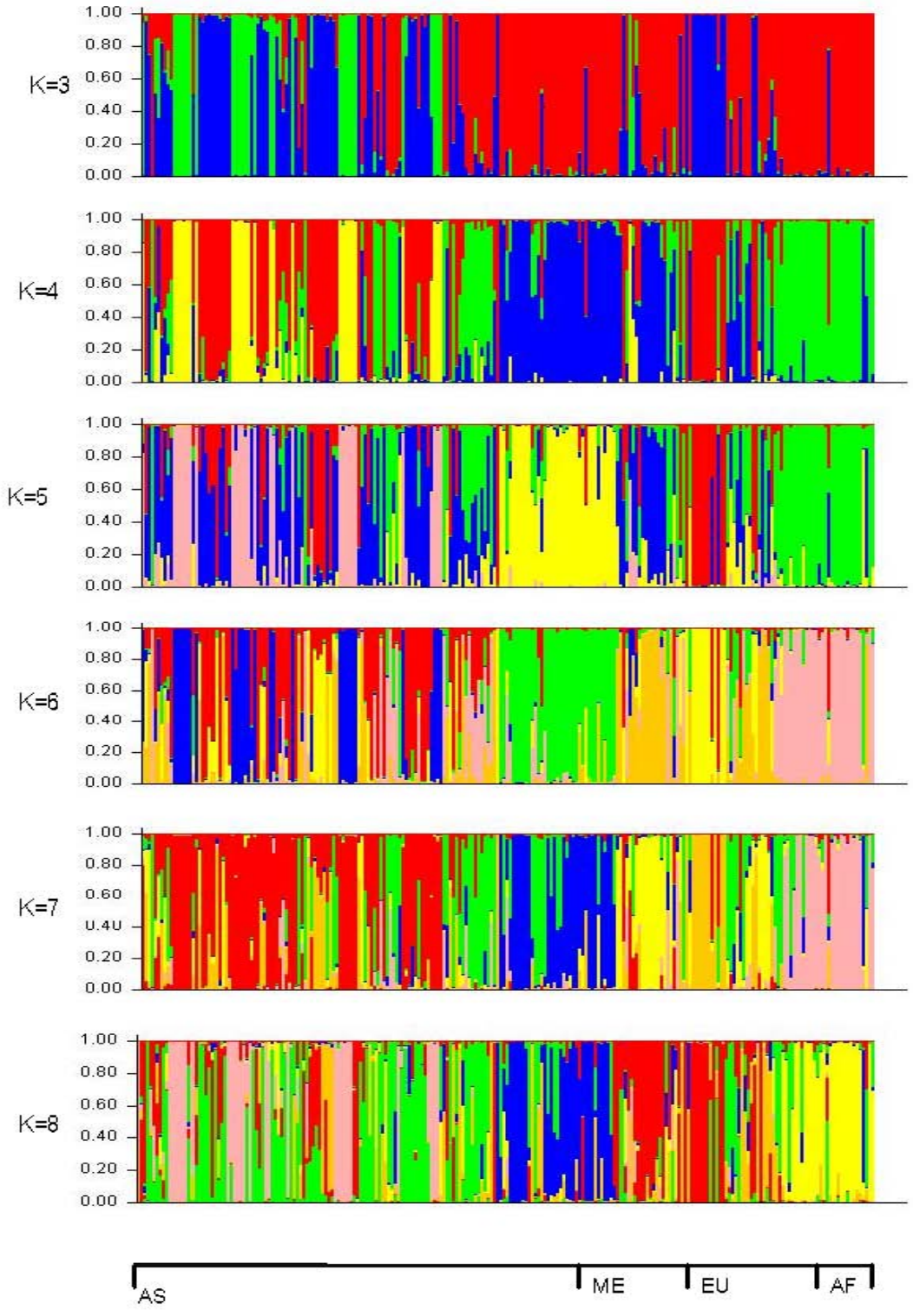

Fig. 2 Estimated population structure of Set 1 (232 individuals) in which continents with proper countries are grouped together. Each individual is represented by a thin vertical line.

DISCUSSION

Facing future problems in wheat production, the use of genetic diversity of different genetic sources is necessary. LCs are demonstrating an elementary and tremendous source of genetic variation in wheat. The importance of wheat LCs has been determined and shown in many projects (Mercado et al., 1996). In self-fertile crops, outcrossing and mutations create intra-population diversity, whereas an increase in inter-population diversity is generated by directed natural or human selection (Ennos, 1983).

In our study, genetic diversity was revealed worldwide between LCs from Europe, Asia, Africa and Middle East. Surprisingly, similar gene diversity was found in Europe, Asia and Middle East in comparison to low gene diversity observed in Africa (Table 1). We speculate that LCs from Middle East offer a high gene diversity because Middle East exhibits the center of origin of wheat. Dreisigacker et. al (2005) proposed an analysis of genetic diversity between Mexican and Turkish LCs. They observed a higher diversity within Turkish than within the Mexican LCs, substantiate with the longer evolutionary history of wheat in Turkey and the limited population size while transferring LCs from Europe to the New World. In contrast to Dreisigacker our diversity data show no trend regarding to the center of origin.

Europe and Asia were the first routes of wheat distribution and are geographic areas were wheat is grown as a dominant crop. Seed movement from the Middle East to Europe and Asia might have been very vital, introducing allelic diversity into those countries. Many LCs analyzed in our study could be transferred very late in history from the Middle East to Europe and Asia, environmental adaptation changed their genetic composition only a little.

Alleles per locus and group specific alleles were most frequent in Asia and least frequent in Africa (Table 1). We expected a much higher level of number of alleles in the Middle East and Europe because of being the center of origin. We also expected a much higher amount of group specific alleles in the Middle East because environmental adaptation in these countries lasted longer compared to the other geographical areas. However, results are most likely biased due to a high variation in sample size across the four groups (Table 1). The lack of historical records implicate a larger number of accessions per geographical area or the same amount accessions within groups for fingerprinting before making conclusions of evolutionary relationships.

The correlation (r) between distances of elevation and latitude across all data points was low ($P < 0.001$) as well as between distances of elevation and longitude. We expected a much higher correlation because we assumed a unity of LCs at same elevations. This assumption could not occupy with these data. One reason could be that the number of SSRs applied might not be enough to separate the LCs due to their ability of adaptation to the various Microclimates. The low correlation might also be explained again by the same assumptions made before. LCs could be spread out late in history from the center of origin and therefore changed their genetic composition just a little. Also different times in which LCs have been collected are factors which influence these data. Propagation and domestication periods are regional completely different between countries and continents and also human selection and their selection pressure can be completely different.

The individuals in Set 2 did not reveal a clear grouping according to their continents and countries of origin in the PCoA (Appendix). As justified for the analysis before, LCs in our study could be transferred from the Near East to other parts of the world in different times or late in history. LCs had not enough time to adapt on environmental factors and changed their genetic composition according to their environment. The Asian LCs, which formed a close relation in the phylogenetic tree are not showing any peculiar grouping (Fig. 2). Just a few Asian LCs reside outside of the scatter plot. Outcrossing can explain why some LCs are

positioned separately from the scatter plot in the PCoA. Furthermore LCs could be intercrossed with wild species, or these few LCs adapted or have been selected for specific niche.

The structure of the UPGMA tree is showing most closely related LCs from CHN and RF, TUR and IRN, and TUN and ITA (Fig. 1). We expected a much better grouping of accessions in closely related countries. RF and CHN as well as TUR and IRN are bordering states. Distribution ways from the Middle East to RF and from RF to CHN or from TUR to IRN are possible. TUN and ITA are not sharing one border but are separated by the Mediterranean See. Distribution by crossing the see by ships is a common propagation route, due to the small distance between ITA and TUN on the seaway.

Regarding to our hypothesis, we expected a grouping of four branches within the tree, however two major branches were derived from the whole set of individuals (Fig. 1). This finding could result from the late distribution of wheat LCs from the center of origin to other parts of the world and therefore no clear patterning appeared. Within the big branch it is peculiar that all countries from Asia are grouped close together. African countries are distributed over both branches. EGY and ETH are grouped together in one clade whereas TUN is grouped close to ITA in the other small major branch of the tree (Fig. 1). Beside our assumptions of a clear grouping of continents we have to consider that several introductions of wheat into Africa could have appeared. Distribution of wheat might have occurred via Europe, the Middle East and Asia, overland or oversee. This different distribution or domestication ways enable this disposition of the African LCs over both branches. So the disposition of the African countries could support the timetable of distribution and domestication of wheat in Africa. Other countries from Middle East and Europe are also distributed over both branches what can also resulted from this cause.

In our STRUCTURE analyses no groupings of accessions were revealing. Countries belonging to the same continents are stringed together (Fig. 2). By setting k = 10 in the model based clustering we detected eight subpopulations and

we assumed initially *k*=3 cluster in the STRUCTURE runs (Fig. 2). As in the analysis before we assumed a clear grouping of countries or continents or a grouping of elevation but none of these expected assumptions could be verified. This analysis shows again the difficulty of assessing environmental impact on genetic structure, particularly if information about collecting dates are restricted. In this case there is no indication and therefore assumptions cannot well documented.

Warburton et al. (2008) proposed a STRUCTURE analyses were maize LCs showed a clear cluster. Each individual is represented by a thin vertical line, which is clearly partitioned in colored segments that represent the individual estimated membership. Here we see a good separation of historical germplasm. LCs in our set were initially not chosen concerning to their role in domestication. Probably more work has to be done in advance when studies investigating domestication history are performed. Either LCs with known passport data and their role in history are chosen or intensive search of passport data have to be realized in advance. However, this is time consuming and in this study and thesis not subject. This way only present diversity can be explained, historical conclusions are difficult.

Limiting factors in this study were missing or incomplete passport data which is not productive in precise characterization of LCs for diversity studies (Dreisigacker, 2005). Passport data were not available for about the half of the dataset. Many passport data such as the correct collection sites were not available. Using molecular markers is a common tool for receiving more information about LCs and wild relatives with less information content in data bases. A big challenge for gene banks is to fulfill information for LCs and other accessions during the moment of admitting new seeds into gene banks and update public web pages regularly. Probably this challenge was not seen in the past and collections were just made based on the fear of losing diversity. Not having the right tools for documenting and analyzing germplasm as precisely as today could also be a cause for incomplete databases. Today modern tools like GPS systems have to be used for

detecting the precise location even with corresponding climate data during collection and data can be stored easily in institutional databases. Complete datasets can help research scientists and breeding programs to introduce new alleles easier and probably more targeted because they have the choice of using accessions collected in environments already facing specific biotic or abiotic stresses for which they are breeding for. For the improvement for tolerance to drought e.g. landraces grown in drought prone environments might be chosen. The selection of accessions could this way be specified based on the available geographic information.

Another problem in conservation of germplasm in genbanks are duplicated entries within and between collections. Hintum (2000) estimated a method to quantify probable duplication within and between gene banks which forms also a major problem in conservation ex situ. This includes accessions with the same variety name, or with the same collection number and same collection site. Accessions were also scored as duplicates if they both had the same well known number (such as PI number or LAC number). A tremendous level of duplication of germplasm accessions exist in gene banks and counteracting efficient work in diversity studies.

High costs for storing seeds in this complexity and performing lab work in this dimension force systems to circumvent these problems. Describing stored germplasm requires a field and lab evaluations of large dimensions which can probably not easily be handled by each gene bank due to capacity and financial constraints. Establishing core collections is an excellent tool for efficient characterization and the accomplishment of diversity studies. The size of germplasm collections is often a barrier for application and proper management. Choosing accessions for utilization is very difficult because material is generally poorly described particularly in combination with the large numbers of accessions (Hintum et al., 1995). Finding core collections which are representative for a whole set of individuals and including a maximum of the genetic variation contained in the whole collection, with a minimum of repetitiveness (Brown et al., 1989a) have to be made for proper description. For the conservation of genetic resources by the development of core collections, as well as for breeding purposes, it is

necessary to estimate the magnitude of genetic variation (Liu et al., 2001). Core collections (Hao et al., 2008; Balfourier et al., 2007) and sampling strategies (Franco et al., 2006) for many different cultivars are already revealed

Diverse cluster methods have been used for demonstrating the same dataset in this thesis. General intentions were the use of different existing algorithm and methods for grouping objects of similar kind into respective categories. Cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Each cluster method has its advantage and disadvantage showing accessions in slightly different perspectives (Reif et al., 2005).

Several marker systems are now available for genotypic population analyses. The choice of marker systems depends on the research objective and on the type of population. Characterizing genetic diversity in wheat with SSRs has often be used by scientists, but new marker systems are already under development and used. Besides evolving marker systems, new tools of genetic mapping such as association mapping (Flint-Garcia et al., 2003) have a potential to reshape our understanding of genetic populations. They will play an increasing role in unraveling the genetic basis of quantitative variation (Wu et al., 2001) in sets of diverse germplasm such as sets of LCs and will further facilitate their use.

CONCLUSION

The present study, including worldwide bread wheat LCs, lead to several conclusions. SSR markers serve as a very effective tool for evaluating large collections of genetic resources and provide important information about the genetic variation of wheat LCs. For drawing conclusions in historical relationships detailed and complete information are necessary and have to be present. Probably more work has to be done in advance when particular studies are performed detecting domestication ways. Either you choose LCs with known passport data or intensive search of passport data has to be realized in advance. Approaches used in this work also provide important information for modern plant breeding concerning the exploitation of the diversity of cereal crops. In the case of complete datasets, introduction of new alleles in breeding programs can be conducted with the help of choosing LCs adapted to specific environmental conditions for which you are breeding for. Another consequence is that it is not possible to transfer results in diversity studies from LCs or wild relatives from other crops to bread wheat LCs. Regarding to the conservation of crop genetic resources describing stored germplasm requires huge capacity and financial constraints. Establishing core collections is an excellent tool for efficient diversity studies. All existing cluster methods are using different algorithms for sorting objects into groups. These calculations have to be considered in choosing the right cluster method.

REFERENCES

Balfourier, F., V. Roussel, P. Strelchenko, F. Exbrayat-Vinson, P. Sourdille, G. Boutet, J. Koenig, C. Ravel, O. Mitrofanova, M. Beckert, G. Charmet. 2007. A worldwide bread wheat core collection arrayed in a 384-well plate. Theoretical and Applied Genetics 114:1265-1275.

Brown, A.H.D., 1989a. A case for core collections. p.136-156. In A.H.D. Brown, O.H. Frankel, D.R. Marshall and J.T. Williams (eds.) The Use of Plant Genetic Resources. Cambridge University Press, Cambridge.

Canela, M., M.N. Aleman, R. Villegas i Tort, D. Sorrells, M.E. Royo i Calpe, Conxita. Association mapping of candidate genes for drought tolerance in Mediterranean durum wheat landraces. From Seed to Pasta: the Durum Wheat Chain. International Durum Wheat Symposium. Bologna, Italy, June 30 - July 3, 2008.

CIMMYT. 2001. The applied Biotechnology Center´s Manual of Laboratory Protocols. 1st ed. Verified March 2009 (http://www.cimmyt.org/ABC/Protocols/manualABC.html), CIMMYT, Mexico, D.F, Mexico.

De Loose M., G. Gheysen. 1995. Identification methods based on molecular techniques. UPOV. (International Union for the Protection of New Varieties of Plants). Working Group on Biochemical and Molecular Techniques and DNA-Profiling in Particular, Wageningen, the Netherlands.

Dixon, J., H.-J. Braun, J.H. Crouch. 2008. Overview: Transitioning wheat research to serve the future needs of the developing world. In: R. Ortiz (ed.), CIMMYT Science Week. El Batan, Mexico, March 3 - 8, 2008.

Dreisigacker, S., P. Zhang, M.L. Warburton, B. Skovmand, D. Hoisington, A.E. Melchinger. 2005. Genetic diversity among and within CIMMYT wheat landraces accessions investigated with SSRs and implications for plant genetic resources management. Crop Science 45:653-661.

Dreisigacker S., P. Zhang, M.L. Warburton, M. van Ginkel, D. Hoisington, M. Bohn, M.E. Melchinger. 2004. SSR and pedigree analyses of genetic

diversity among CIMMYT wheat lines targeted to different megaenvironments. Crop Science 44:381-388.

Ennos, R.A. 1983. Maintenance of genetic variation in plant populations. p. 129-155. *In* M.K. Hecht et al. (ed.) Evolutionary biology. Vol. 16. Plenum Press, New York.

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

Flint-Garcia, A., J.M. Thornsberry, E.S. Buckler. 2003. Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54:357-74.

Franco, J., J. Crossa, M. L. Warburton, S. Taba. 2006. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. Crop Science 46:854-864.

Gower, J.C.1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325-388.

Hao C.Y., Y.C. Dong, L.F. Wang, G.X. You, H.N. Zhang, H.M. Ge, J.Z. Jia, X.Y. Zhang. 2008. Genetic diversity and construction of core collection in Chinese wheat genetic resources. Chinese Science Bulletin 53:1518-1526.

Hintum, T.J.L. van, R. von Bothmer, D.L. Visser. 1995. Sampling strategies for composing a core collection of cultivated barley (*Hordeum vulgare* s. lat.) collected in China. Hereditas 122:7-17.

Hintum, T.J.L. van. 2000. Duplication within and between germplasm collections. III. A quantitative model. Genetic Resources and Crop Evolution 47:507-513.

Kihara, H. 1982. Wheat studies: Retrospect and prospects. Elsevier Science Ltd., St. Louis, MO.

Li, L., Z. He, M. Lillemo, Q. Sun, X. Xia. 2008. Molecular characterization of allelic variations at *Pina* and *Pinb* loci in Shandong wheat landraces, historical and current cultivars. Journal of Cereal Science 47:510-517.

Liu, K., S.V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.

Liu, F., G.L. Sun, B. Salomon, R. von Bothmer. 2001. Distribution of allozymic alleles and genetic diversity in the American Barley Core Collection. Theoretical and Applied Genetics 102:606-615.

Manifesto, M. M., A. R. Schlatter, H. E. Hopp, E. Y. Suárez, J. Dubcovsky Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Science 41:682-690.

Mantel, N.1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Maurer, H.P., A.E. Melchinger, M. Frisch. 2008. Population genetic simulation and data analysis with Plabsoft. Euphytica 161:133-139.

Mercado, L.A., E. Souza, K.D. Kephart. 1996. Origin and diversity of North American hard spring wheats. Theoretical and Applied Genetics 93: 593-599.

Nazari, K, A. Amini, A. Yahyaoui, R. Sing. Characterization of seedling and adult-plant resistance to stem rust race Ug99 in Iranian bread wheat landraces. The 11th International Wheat Genetics Symposium proceedings. Brisbane, Australia, August 24 - 29, 2008.

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

Raman, H., A. Karakousis, J.S. Moroni, R. Raman, B.J. Read, D.F. Garvin, L.V. Kochian, M.E. Sorrells. 2003. Development and allele diversity of microsatellites markers linked to the aluminium tolerance gene *alp* in barley. Australian Journal of Agricultural Research 54:1315-1321.

Reif, J.C., P. Zhang, S. Dreisigacker, M.L. Warburton, M. van Ginkel, D. Hoisington, M. Bohn, A.E. Melchinger. 2005. Wheat genetic diversity trends during domestication and breeding. Theoretical and Applied Genetics 110:859-864.

Röder, M.S., V. Korzun, K. Wendehake, J. Plaschke, M.H. Tixier, P. Leroy, M.W. Ganal. 1998. A microsatellite map of wheat. Genetics 149:2007-2023.

Skovmand, B., and S. Rajaram. 1990. Utilization of genetic resources in the improvement of hexaploid wheat. *In* J.P. Srivastava and A.B. Damania (ed.) Wheat genetic resources: Meeting diverse needs. J. Wiley & Sons, Chichester, UK

Stich, B., J. Möhring, H.-P. Piepho, M. Heckenberger, E.S. Buckler, A.E. Melchinger. 2007. Comparison of mixed-model approaches for association mapping. Genetics 178:1745-1754.

Stodart, B.J., M. Mackay, H. Raman. 2005. AFLP and SSR analysis of genetic

diversity among landraces of bread wheat (*Triticum aestivum* L. em. Thell) from different geographic regions. Australien Journal of Agricultural Research 56:691-697.

Weir, B.S. 1996. Genetic data analysis II. Sinauer Associates, Sunderland, MA.

Wright, S. 1978. Evolution and genetics of populations. University Chicago Press, Chicago.

Wu, R., C.X. Ma, G. Casella. 2001. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. Genetics 160:779-792.

Yu, J., G. Bai, W. Zhou, Y. Dong, F. Kolb. 2007. Quantitative trait loci for fusarium head blight resistance in Chinese wheat landrace Wangshuibai. Phytopathology 98:87-94.

APPENDIX

Table 2. Continent, country, CIMMYT accession number, elevation, latitude and longitude of 232 LC accessions of wheat.

Continent	Country	CIMMYT accession	Elevation m	Latitude	Longitude
Africa					
	Egypt (EGY)				
		CWI12577	-	27°00´N	30°00´E
		CWI12585	-	29°46´N	31°18´E
		CWI13215	-	-	-
		CWI12571	-	-	-
		CWI12589	-	-	-
	Ethiopia (ETH)				
		CWI14129	-	8°00´N	38°00´E
		CWI11084	-	9°41´N	39°32´E
		CWI21844	2470	10°27´N	37°34´E
		CWI21358	2363	11°08´N	39°38´E
		CWI21365	2363	11°08´N	39°38´E
	Tunisia/Morocco (TUN)				
		CWI21848	-	-	-
		CWI21855	-	-	-
		CWI21857	-	-	-
		CWI21859	-	-	-
		CWI21852	-	-	-
		CWI21853	-	-	-
		CWI21856	-	-	-
		CWI21858	-	-	-
Asia					
	Afghanistan (AFG)				
		CWI12657	914	32°22´N	62°02´E
		CWI12740	2499	33°55´N	64°55´E
		CWI12919	-	34°00´N	66°00´E
		CWI6960	-	34°00´N	66°00´E
		CWI12718	2408	34°06´N	64°21´E
		CWI12725	2499	34°06´N	64°25´E
		CWI12713	2621	34°08´N	64°25´E
		CWI12695	2195	34°09´N	64°03´E
		CWI12743	2652	34°13´N	64°49´E
		CWI15622	-	34°15´N	64°52´E
		CWI12705	2408	34°19´N	64°25´E
		CWI12681	1676	34°21´N	63°51´E

	CWI12753	2652	34°22′N	65°01′E
	CWI12820	2926	34°23′N	66°55′E
	CWI12649	1311	34°29′N	62°33′E
	CWI15651	3150	34°30′N	66°17′E
	CWI12761	2195	34°31′N	65°09′E
	CWI12773	2256	34°31′N	65°14′E
	CWI12785	2134	34°31′N	65°26′E
	CWI15654	2650	34°32′N	65°27′E
	CWI12801	2469	34°33′N	65°47′E
	CWI15652	2800	34°33′N	65°47′E
	CWI12611	1676	34°33′N	62°56′E
	CWI12789	2438	34°33′N	65°33′E
	CWI12736	2652	34°33′N	64°49′E
	CWI12811	2774	34°38′N	65°45′E
	CWI12626	2042	34°39′N	63°07′E
	CWI12836	2713	34°45′N	67°02′E
	CWI12780	2103	34°50′N	65°18′E
	CWI12856	792	35°55′N	68°44′E
	CWI12864	945	36°41′N	67°50′E
	CWI14608	458	36°42′N	67°06′E
China (CHN)				
	CWI13547	8	23°07′N	113°15′E
	CWI13323	-	31°14′N	121°28′E
	CWI14621	-	31°45′N	120°15′E
	CWI14127	20	32°04′N	118°46′E
	CWI14644	-	33°00′N	120°00′E
	CWI13158	-	35°00′N	105°00′E
	CWI16247	-	35°00′N	103°11′E
	CWI13329	1342	38°42′N	103°11′E
	CWI13330	1342	38°42′N	103°11′E
	CWI14029	6	39°19′N	117°48′E
	CWI13318	-	42°00′N	86°00′E
	CWI13311	1050	43°32′N	87°54′E
	CWI13310	1050	43°32′N	87°54′E
	CWI13319	857	43°48′N	87°35′E
	CWI13321	857	43°48′N	87°35′E
	CWI13806	-	44°00′N	112°00′E
	CWI14017	257	44°56′N	130°32′E
	CWI14525	-	45°00′N	125°00′E
	CWI14563	-	45°00′N	125°00′E
	CWI14417	-	45°00′N	125°00′E
	CWI14514	137	45°11′N	124°49′E
	CWI14136	238	45°44′N	127°27′E
	CWI14201	151	46°13′N	125°16′E
	CWI14159	147	47°24′N	130°22′E
	CWI14144	-	48°00′N	128°00′E
	CWI14499	279	48°01′N	126°15′E
	CWI14150	316	48°12′N	125°24′E

	CWI14215	226	48°29´N	124°50´E
	CWI14031	-	-	-
	CWI13259	-	-	-
	CWI13260	-	-	-
India (IND)				
	CWI21878	510	18°00´N	75°18´E
	CWI13335	592	18°32´N	73°52´E
	CWI13462	-	19°30´N	75°00´E
	CWI13579	-	25°00´N	86°00´E
	CWI13582	-	25°00´N	86°00´E
	CWI13589	-	25°00´N	86°00´E
	CWI13575	-	25°00´N	86°00´E
	CWI13491	130	27°00´N	80°00´E
	CWI13492	130	27°00´N	80°00´E
	CWI13833	-	31°06´N	77°10´E
	CWI13834	1570	34°14´N	74°47´E
	CWI11128	-	-	-
	CWI21895	-	-	-
	CWI21896	-	-	-
	CWI21899	-	-	-
Pakistan (PAK)				
	CWI15458	-	30°00´N	70°00´E
	CWI15462	-	30°00´N	70°00´E
	CWI10314	-	30°00´N	70°00´E
	CWI13441	-	31°00´N	72°00´E
	CWI13115	2590	34°53´N	76°11´E
	CWI13108	2680	34°57´N	76°33´E
	CWI13114	2410	34°58´N	76°09´E
	CWI13113	2380	35°02´N	76°05´E
	CWI13109	2590	35°03´N	76°29´E
	CWI13112	2260	35°07´N	75°58´E
	CWI13111	2990	35°08´N	75°23´E
	CWI13107	2500	35°09´N	76°19´E
	CWI13110	2590	35°10´N	75°25´E
	CWI13104	2440	35°10´N	76°05´E
	CWI13105	2500	35°11´N	76°11´E
	CWI13106	2530	35°12´N	76°15´E
	CWI13103	2320	35°15´N	75°51´E
	CWI13117	2230	35°18´N	75°46´E
	CWI13118	2290	35°19´N	75°38´E
	CWI13098	2160	35°22´N	75°30´E
	CWI13099	2230	35°25´N	75°44´E
	CWI13097	2190	35°26´N	75°25´E
	CWI13096	2190	35°33´N	75°24´E
	CWI13101	2320	35°34´N	75°36´E
	CWI13095	2040	35°35´N	75°20´E
	CWI13102	2350	35°36´N	75°33´E
	CWI13094	2040	35°37´N	75°08´E

		CWI13089	1710	35°43´N	74°34´E
		CWI13093	1980	35°43´N	74°49´E
		CWI13090	1830	35°46´N	74°32´E
		CWI13091	3050	36°08´N	73°14´E
		CWI13085	2930	36°09´N	72°54´E
		CWI13083	3230	36°10´N	72°41´E
		CWI13084	3140	36°10´N	72°46´E
		CWI13082	2770	36°10´N	72°58´E
		CWI13081	2560	36°10´N	73°06´E
		CWI13080	2210	36°14´N	73°27´E
		CWI13074	2230	36°16´N	74°36´E
		CWI13077	2440	36°20´N	73°21´E
		CWI13078	2470	36°26´N	73°23´E
		CWI13086	2380	36°31´N	73°54´E
	Russian Federation (RF)				
		CWI13932	-	44°00´N	133°00´E
		CWI14123	97	55°00´N	73°24´E
		CWI13386	4	59°53´N	30°16´E
		CWI13377	4	59°53´N	30°16´E
		CWI13385	4	59°53´N	30°16´E
		CWI13391	4	59°53´N	30°16´E
		CWI13349	4	59°53´N	30°16´E
		CWI13378	4	59°53´N	30°17´E
		CWI13390	4	59°53´N	30°18´E
		CWI13379	4	59°53´N	30°19´E
		CWI13903	-	60°00´N	47°00´E
		CWI14034	-	60°00´N	47°00´E
		CWI12313	-	-	-
		CWI12366	-	-	-
		CWI13535	-	-	-
		CWI13071	-	-	-
		CWI13931	-	-	-
		CWI13344	-	-	-
		CWI21907	-	-	-
		CWI21961	-	-	-
		CWI21977	-	-	-
Europe					
	Switzerland (CHE)				
		CWI12261	-	-	-
		CWI12317	-	-	-
		CWI12420	-	-	-
		CWI12375	-	-	-
	Germany/Netherland (DEU)				
		CWI13488	2000	48°30´N	11°30´E
		CWI12291	-	-	-
		CWI13415	-	-	-
	Spain/Portugal (ESP)				
		CWI13698	150	28°19´N	16°34´W

		CWI13699	150	28°19´N	16°34´W
		CWI13798	-	39°00´N	09°08´W
		CWI21957	-	-	-
		CWI13794	-	-	-
		CWI13797	-	-	-
		CWI13791	-	-	-
		CWI13792	-	-	-
		CWI13796	-	-	-
	France (FRA)				
		CWI21863	-	-	-
		CWI21972	-	-	-
	Greece (GRC)				
		CWI22029	-	39°00´N	22°00´E
		CWI14652	-	39°00´N	22°00´E
		CWI11155	-	-	-
	Hungary/Romania (HUN)				
		CWI9769	-	47°25´N	19°20´E
		CWI21949	-	-	-
		CWI22032	-	-	-
	Italia (ITA)				
		CWI21929	-	-	-
		CWI19875	-	-	-
		CWI21991	-	-	-
		CWI22003	-	-	-
		CWI21944	-	-	-
		CWI22019	-	-	-
	Yugoslavia (YUG)				
		CWI12509	1190	41°26´N	21°44´E
		CWI12518	434	41°59´N	21°13´E
		CWI12521	-	42°36´N	20°00´E
		CWI19348	-	-	-
		CWI19377	-	-	-
		CWI19346	-	-	-
		CWI19386	-	-	-
		CWI19373	-	-	-
		CWI19228	-	-	-
		CWI19378	-	-	-
Middle East					
	Turkey (TUR)				
		CWI19945	353	36°42´N	38°56´E
		CWI19790	90	36°53´N	30°42´E
		CWI19942	502	36°58´N	38°25´E
		CWI19909	1010	37°11´N	33°14´E
		CWI19896	570	37°51´N	40°39´E
		CWI19941	974	38°42´N	31°02´E
		CWI8612	-	39°00´N	35°00´E
		CWI19904	1135	39°11´N	35°14´E
		CWI20049	760	39°45´N	30°57´E

	CWI19898	857	41°01´N	34°02´E
Uzbekistan/Kyrgyzstan (UZB)				
	CWI13456	-	41°00´N	75°00´E
	CWI13453	608	39°35´N	66°35´E
	CWI13394	-	41°00´N	64°00´E
	CWI13343	452	41°20´N	69°18´E
Iran/Iraq (IRN)				
	CWI13423	-	29°00´N	53°00´E
	CWI13420	1630	30°08´N	52°36´E
	CWI13292	-	32°00´N	53°00´E
	CWI13708	-	32°00´N	53°00´E
	CWI12449	-	32°00´N	53°00´E
	CWI13419	-	32°00´N	53°00´E
	CWI8558	-	32°00´N	53°00´E
	CWI13650	1297	37°31´N	45°14´E
	CWI14447	-	33°00´N	44°00´E
Israel (ISR)				
	CWI21572	-	31°30´N	34°45´E
	CWI22055	-	-	-
	CWI22056	-	-	-
	CWI22058	-	-	-
	CWI21832	-	-	-
	CWI22053	-	-	-
	CWI22057	-	-	-
Azerbaijan/Armenia/Georgia (AZB)				
	CWI13341	-	40°11´N	44°30´E
	CWI14535	-	40°30´N	45°00´E
	CWI14534	-	40°30´N	47°00´E
	CWI13799	-	42°00´N	43°30´E

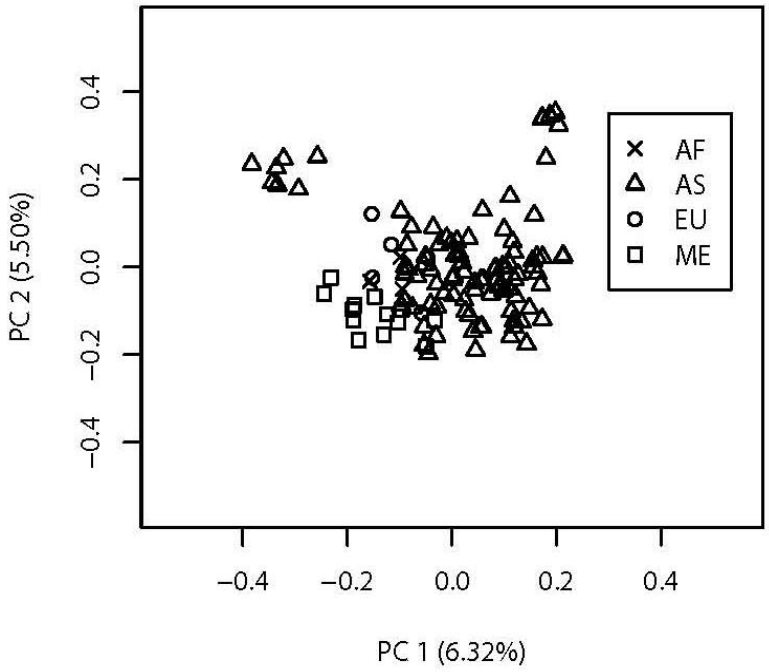

Fig. 2 PCoA for the 118 individuals in Set 2; AF (Africa), AS (Asia), EU (Europe) and ME (Middle East). Geographic origin is designated by symbols (see legend).

ACKNOWLEDGEMENTS

I would like to thank PD Dr. Jochen C. Reif for the offer to realize my diploma thesis, for his advice and support.

I am deeply indebted to Dr. Susanne Dreisigacker, contributing suggestions and encouragement during writing my thesis.

I also acknowledge Sandra Fischer for her help and encouragement.

I also want to thank Maria Zaharieva for her support and advices professionally and personally.

Especially, I am very grateful to my parents and my friends for their continuous support in countless situations.

ERKLÄRUNG

Hiermit versichere ich, Marlen Hübner, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen, als die im Literaturverzeichnis angegebenen Quellen, verwendet habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht. Die Zeichnungen oder Abbildungen dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechendem Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Stuttgart, den 13.03.2009