Themenkreis E: Wildsammlung, Inkulturnahme, Züchtung

ESL 19 Next generation breeding tools for chamomile: Evaluating genetic diversity, ploidy variation, and identifying marker-trait associations

Züchtungsmethoden der nächsten Generation für Kamille: Bewertung der genetischen Diversität, der Ploidievariation und Identifizierung von Marker-Merkmal Assoziationen

Lars-Gernot Otto[1*], Jonathan Brassac[1], Prodyut Mondal[2], Marlis Sonnenschein[3], Bartolome Plocharski[3], Wolfram Junghanns[4], Susanne Preiss[2], Jörg Degenhardt[2], Mariateresa Lazzaro[5], Marika Bocchini[5], Emidio Albertini[5], Andreas Plescher[3], Beate Fraust[1], Sang He[1], Jochen Reif[1], Timothy Sharbel[6]

* corresponding author: ottol@ipk-gatersleben.de

[1] Quantitative Genetics Research Group, Department Plant Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, D-06466 Seeland OT Gatersleben, Germany.

[2] Research Group of Pharmaceutical Biotechnology, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany.

[3] Pharmaplant GmbH, Artern, Germany.

[4] Dr. Junghanns GmbH, Aschersleben, Germany.

[5] Department of Agricultural, Food and Environmental Sciences, University of Perugia, Perugia, Italy.

[6] Global Institute for Food Security, University of Saskatchewan, Saskatoon, Canada.

DOI 10.5073/jka.2018.460.019

Zusammenfassung

Die Nutzung von Kamille (*Matricaria recutita* L.) als Arzneipflanze hat eine lange Tradition und umfasst einen weiten Anwendungsbereich. Die Blütenköpfe von Kamille enthalten eine Vielzahl an medizinisch wirksamen Inhaltsstoffen. Next-Generation Sequenzierungsmethoden (NGS) werden bei den Hauptkulturpflanzen verwendet, um genetische Ressourcen zu erschließen und die Züchtung zu unterstützen. Genotypisierung-durch-Sequenzierung (GBS) wurde bei der Nicht-Modellpflanze Kamille zur Charakterisierung der genetischen Diversität angewandt. Unter Nutzung von den erhaltenen 6495 hochqualitativen SNP-Markern wurden mittels einer genomweiten Assoziationsstudie (GWAS) DNA-Sequenzen identifiziert, die signifikant mit dem pharmazeutisch wichtigen Alpha-Bisabolol-Gehalt assoziiert sind. Die Ploidievariation in der Art Echte Kamille wurde mittels Hochdurchsatz-Durchflusszytometrie untersucht. Di-, tri- und tetraploide Pflanzen wurden identifiziert und in Feldversuchen charakterisiert. Da für das Ernteprodukt bei Kamille keine Samen benötigt werden, könnte Triploidie ein Weg sein, eine sterile Kamillensorte zu erzeugen. Mit einer sterilen Sorte könnte so das Problem gelöst werden, dass Kamillensamen im Boden bis zu 15 Jahre lang nach dem Anbau auskeimen, was den Fruchtwechsel auf den Ackerböden erheblich erschwert und u.a. zur Akkumulation von Kamillenkrankheiten führt.

Abstract

Chamomile (*Matricaria recutita* L.) has a long history of use in herbal medicine with various applications, and the flower heads contain numerous medicinally active compounds. Next generation sequencing (NGS) approaches are applied to exploit genetic resources in the major crop plants to develop genomic resources, and to enhance breeding. Genotyping-by-sequencing (GBS) has been used to evaluate the genetic structure of cultivated populations in the non-model crop chamomile using 6495 SNP markers, and to perform a genome wide association study (GWAS) identifying sequences significantly associated with the medicinally important alpha-bisabolol content. Ploidy variation in chamomile was investigated by high-throughput flow-cytometry. Di-, tri- and tetraploid plants were identified, and in field trials characterized. Since seeds are not needed in the harvested product of chamomile, triploidy could be a way to obtain a sterile chamomile variety, omitting the

problems of chamomile seeds lying up to 15 years dormant in the soil and facilitating crop rotation in the fields.

Keywords: *Matricaria recutita*, chamomile, genetic diversity, GBS, AFLP

Stichwörter: *Matricaria recutita*, Kamille, genetische Diversität, GBS, AFLP

Introduction

German chamomile (*Matricaria recutita* L. syn. *Chamomilla recutita* (L.) Rauschert) with a long history of use in herbal medicine is one of the most important medicinal plants. However, unlike in major crops, few breeding efforts are done, and rarely "next generation" breeding methods are applied. Here one of the next-generation sequencing (NGS) approaches, Genotyping-by-Sequencing (GBS), which is widely used in many crop plants, has been applied to chamomile of various origins. Together with AFLP-markers, numerous SNP markers rendered by GBS (Otto et al., 2017) allowed the evaluation of genetic diversity within this species in addition to a genome-wide association study (GWAS) for two important traits, flowering time and the content of alpha-bisabolol. The high density of the SNP-markers provided a high chance of identifying closely linked or functional (located within the gene of interest) markers associated with important agronomical traits.

Additionally, with the intention of studying the approaches to develop a sterile triploid chamomile, for several chamomile origins ploidy level variation was determined and some traits were characterized for plants within each ploidy class (diploid, triploid, tetraploid). The methods used here might be applied to other MAPs (medicinal and aromatic plants), with chamomile serving as a model.

Material and methods

Plant Material

Different origins (varieties, wild and cultivated populations, accessions) of chamomile from various geographic regions were included in the analyses, as for the evaluation of ploidy variation described in Otto et al. (2015), and for the analysis of genetic diversity by GBS described in Otto et al. (2017). For the analysis of genetic diversity with AFLP markers, 2 to 8 plants each from 46 chamomile origins and from 11 origins out of 8 related species (*Artemisia absinthium* L.; *Artemisia dracunculus* L., *Artemisia vulgaris* L., *Matricaria discoidea* DC., *Matricaria nigellaefolia* DC, *Matricaria trichophylla* (Boiss) Boiss., *Matricaria perforata* Mérat, *Achillea millefolium* L.) were analysed. The focus was placed on cultivated chamomile.

The field trials for the phenotyping of di-, tri- and tetraploid plants were done with blocks of clonal plants for each genotype in the years 2014 and 2016 in Artern, Germany, encompassing 7.5 m^2 for each genotype. To reduce the influence of the individual genetic background, the diploid, triploid and tetraploid plants compared were originating from the same or similar population.

Ploidy analysis

Flow-cytometric analysis was done according to Otto et al. (2015) using *Vicia faba* L. as internal standard and the 2-component buffer CyStain UV Precise P (Partec, Münster, Germany) with additional 0.5 % β-mercaptoethanol and 1 % polyvinylpyrrolidone (PVP) to stabilize the DNA from oxidation.

AFLP-marker analysis

AFLP-analysis (Amplified Fragment Length Polymorphism) was done modified according to Oberprieler et al. (2009) and Vos et al. (1995) with 6 primer combinations (EcoRI+ACA – MseI+CAC,

EcoRI+ACA – MseI+CGA, EcoRI+ACC – MseI+CAT, EcoRI+ACC – MseI+CAG, EcoRI+AGG – MseI+CTT, EcoRI+AGG – MseI+CGA) using a fragment analyzer (CEQ 8000, Beckman-Coulter, Krefeld, Germany). A first AFLP analysis was done with a subset of 47 samples (38 chamomile and 9 related species) using 79 markers.

Phenotyping and Genotyping-by-Sequencing (GBS) for genome-wide association study

The analysis of chemical compounds was done by gas chromatography - mass spectrometry (GC-MS) with modifications according to KÖLLNER et al. (2004) and IRMISCH et al. (2012). DNA was extracted with the Chloropure kit (Agencourt Bioscience Corp., Beverly, Massachusetts, USA). The sample preparation and the GBS was accomplished at the Biotechnology Resource Center (BRC, Cornell, USA) according to a modified protocol from ELSHIRE et al. (2011) on the Illumina HiSeq 2000/2500 (100 bp, single-end reads) using the enzyme ApeKI for digestion. In the absence of a reference genome, the assembly was done *de novo* (OTTO et al., 2017). SNPs were detected using the software pipeline pyRAD v. 3.0 (Eaton, 2014). The genetic diversity and structure of cultivated chamomile was analysed with 6,495 filtered, high-quality SNPs using the admixture model of STRUCTURE 2.3.4 (PRITCHARD et al, 2000; FALUSH et al, 2003; HUBISZ et al., 2009). The genome-wide association study was performed with the R package "rrBLUP" (ENDELMAN, 2011).

Results and Discussion

Analysis of genetic diversity and population structure

The analysis of the genetic diversity in chamomile by AFLPs and SNP markers (generated by Genotyping-by-Sequencing) revealed that *Matricaria recutita* L., from here on chamomile, is clearly distinct from related species like *M. discoidea* (Fig. 1). Within *M. recutita*, a group of 14 tetraploid origins was genetically highly homogenous as revealed by STRUCTURE (PRITCHARD et al., 2000), principle coordinate analysis and neighbor-joining analysis performed with 6495 high-quality SNPs, mined with the pyRAD-pipeline (EATON, 2014). In contrast, several tetraploid and most of the diploid origins displayed a higher genetic diversity. The lower genetic diversity in the tetraploid origins possibly reflects their breeding history, namely, their development from a limited set of diploid populations. These data could be used to exploit available genetic diversity in chamomile breeding, e.g. to profit from the heterosis effect. For further analysis, it is planned to investigate wild chamomile populations from the geographic origin, i.e. the Near East and South-Southeast Europe, to deeper unravel the genetic variation.

Genome wide association study for alpha-bisabolol content and flowering time

18 chemical compounds were measured in flower heads by GC-MS. The highest genomic heritability, estimating how much of the observed phenotypic variation can be explained genetically, was calculated for alpha-bisabolol (0.427) according to VISSCHER et al. (2008) and DE LOS CAMPOS et al. (2015). For all other compounds the genomic heritability was lower than 0.25. Thus, the further analysis was focused on alpha-bisabolol as one of the medicinally most important compounds of chamomile.

Varieties with genotypes possessing high alpha-bisabolol content (e.g. 'Manzana') could be genetically discriminated. By the genome wide association study, SNP markers significantly associated with flowering time and alpha-bisabolol, arguably the most important medicinal compound of chamomile, could be identified (OTTO et al., 2017). Moreover, four sequences carrying alpha-bisabolol-associated SNPs are involved in plant biotic and abiotic stress response in different plants species (BLAST alignments).These markers could pave the way for the application of marker assisted selection (MAS) in chamomile to select the desired plant at a young stage

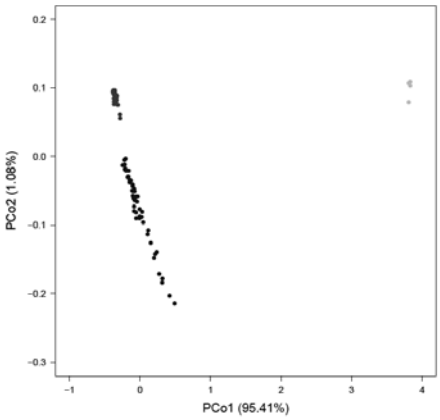
without extensive phenotyping of all the plants in the breeding population, e.g. for the analysis of the chemical compounds.

Ploidy variation in chamomile

By high-throughput flow-cytometry, the ploidy variation in 15 chamomile origins was investigated in seeds and plants (OTTO et al., 2015). Tetraploid origins often contained a portion di-, tri- and clearly aneuploid individuals. Triploids plants were selected in these tetraploid varieties and populations, but were also generated by crossing tetraploid and diploid parents. The triploid plants showed good agricultural performance, and were highly sterile.

A high individual variation in phenotypic traits could be observed between plants, independent of their ploidy level. However, altogether differences were detected between chamomiles of different ploidy levels. Diploid chamomile flowered significantly earlier then tetraploid one (OTTO et al., 2017). The flower heads from tetraploid plants were larger than these from di- and triploid ones (Fig. 2). No clear differences in average flower heads size were present between the investigated di- and triploid plants. In the field trials, the yield (harvested mass of flower heads) of di-, tri- and tetraploid plants was comparable, with a strong variation observed between single genotypes and with single triploid genotypes reaching the highest yield. The flowering period and the plant height were also comparable between plants of the various ploidy levels.

In chamomile, the flower heads are harvested before most of them generate seeds. .As in many fruit and ornamental plants, in which seeds are not desired, a triploid seedless chamomile variety could be very valuable due to the absence of self-sowing. Seeds from chamomile can lay dormant for up to 15 years in the soil before germination, and chamomile is rather resistant to herbicides. Thus, once a field is contaminated with chamomile, crop rotation is hampered. Also, successive cultivation of chamomile leads to the accumulation of chamomile specific diseases. A triploid sterile variety could solve this problem.



Colours: **green**: outgroup from *M. discoidea*; **red**: homogenous group of 14 tetraploid origins; **blue**: rather diverse group consisting of all remaining samples

Fig. 1 (OTTO et al. 2017): Principle coordinate analysis of the 95 samples reveals the outgroup *M. discoidea* to be clearly distinct from *M. recutita*.

Acknowledgements

The majority of this work was financed based upon a resolution of the Federal Parliament of Germany (Deutscher Bundestag) by the Bundesministerium für Ernährung und Landwirtschaft

(BMEL, support code 11NR389 and 14NR063) via the Fachagentur Nachwachsende Rohstoffe e.V. (FNR) as project executing organisation for the funding programme "Renewable Resources". The funding body did not execute any influence on the study nor the manuscript.

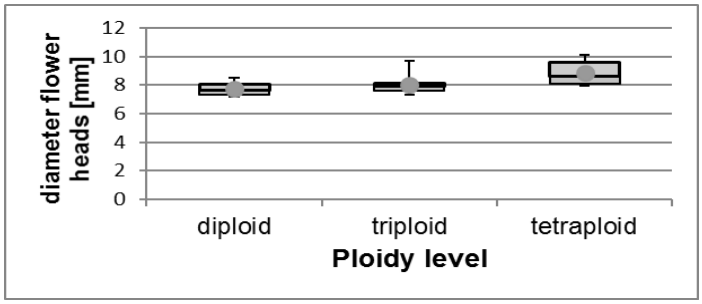


Fig. 2: Average diameter for the flower heads from diploid, triploid and tetraploid plants.

Box-whisker-plots of the average flower head diameter for the different ploidy levels from plants in field trial (diploid n=4, triploid n=12, tetraploid n=7): The quartiles 2 and 3 are drawn with the median in between them. The whiskers indicate the maximum and minimum values, whereas the dot in the middle represents the arithmetic mean.

References

DE LOS CAMPOS, G., D. SORENSEN, D. GIANOLA, 2015: Genomic heritability: what is it? PLoS Genet **11(5)**: e1005048.

EATON D.A.R., 2014*: PyRAD:assembly of de novo RADseq loci for phylogenetic analyses.* Bioinformatics **30(13),** 1844-1849. doi: 10.1093/bioinformatics/btu121.

ELSHIRE, R. J., J. C. GLAUBITZ, Q. SUN, J. A. POLAND, K. KAWAMOTO, E. S. BUCKLER, S. E. MITCHELL, 2011: A. Robust Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE **6(5)**: e19379. doi:10.1371/journal.pone.0019379.

ENDELMAN, J. B., 2011: Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome **4**, 250-255. doi: 10.3835/plantgenome2011.08.0024.

FALUSH, D., M. STEPHENS, J. K. PRITCHARD, 2003: Inference of population structure: Extensions to linked loci and correlated allele frequencies. Genetics **164,** 1567–1587.

HE J., X. ZHAO, A. LAROCHE, Z.-X. LU, H. LIU, Z. LI, 2014: Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. **5,** 484. http://doi.org/10.3389/fpls.2014.00484.

HUBISZ M., D. FALUSH, M. STEPHENS, J. PRITCHARD, 2009: Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. **9(5),** 1322–1332.

Irmisch S., S. T. Krause, G. Kunert, J. Gershenzon, J. Degenhardt, T. G. Köllner, 2012: The organ-specific expression of terpene synthase genes contributes to the terpene hydrocarbon composition of chamomile essential oils. BMC Plant Biology **12**: 84. doi:10.1186/1471-2229-12-84.

Köllner T.G., C. Schnee, J. Gershenzon, J. Degenhardt, 2004: The sesquiterpene hydrocarbons of maize (*Zea mays*) form five groups with distinct developmental and organ-specific distributions. Phytochemistry **65**, 1895-1902.

OBERPRIELER, Ch., J. MEISTER, Ch. SCHNEIDER, N. KILIAN, 2009: Genetic structure of *Anogeissus dhofarica* (Combretaceae) populations endemic to the monsoonal fog oases of the southern Arabian Peninsula. Biological Journal of the Linnean Society **97**, 40–51.

OTTO, L.-G., W. R. JUNGHANNS, A. PLESCHER, M. SONNENSCHEIN, T. F. SHARBEL,2015: Towards breeding of triploid chamomile (*Matricaria recutita* L.) - Ploidy variation within German chamomile of various origins. Plant Breeding J. **134(4)**, 485-493. DOI: 10.1111/pbr.12285.

OTTO, L-G., P .MONDAL, J. BRASSAC, S. PREISS, J. DEGENHARDT, S. HE, J. C. REIF, T. F. SHARBEL, 2017: Use of genotyping-by-sequencing to determine the genetic structure in the medicinal plant chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by genome-wide association mapping. BMC Genomics **18**:599. https://doi.org/10.1186/s12864-017-3991-0.

PRITCHARD J. K., M. STEPHENS, P. DONNELLY, 2000: Inference of population structure using multilocus genotype data. Genetics **155,** 945–959.

PRITCHARD, J. K., M. STEPHENS, P. DONNELLY, 2000: Inference of population structure using multilocus genotype data. Genetics **155,** 945–959.

VISSCHER P.M., W. G. HILL, N. R. WRAY, 2008: Heritability in the genomics era-concepts and misconceptions. Nat Rev Genet **9**, 255–266. doi: 10.1038/nrg2322. pmid:18319743.

VOS P., R. HOGERS, M. BLEEKER, M. van de REIJANS, T. LEE, M. HORNES, A. FRIJTERS, J. POT, J. PELEMANN, M. KUIPER, M. ZABEAU, 1995: AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res **23**, 4407-4414.