RESEARCH ARTICLE

Assessment of the diversity of comfrey (*Symphytum officinale* L. and *S. × uplandicum* Nyman)

Joana Ruzicka · Karin Berger-Büter · Nils Esslinger · Johannes Novak

Received: 8 July 2020 / Accepted: 1 March 2021 / Published online: 16 March 2021
© The Author(s) 2021

Abstract Comfrey *Symphytum officinale* L. (true comfrey) and *S. × uplandicum* Nyman (a hybrid between *S. asperum* Lepech × *S. officinale* L., Russian comfrey) are used externally for the treatment of pain, inflammation and swelling of muscles and joints in degenerative arthritis, in acute myalgia, sprains, contusions and strains after accidents. Besides plant secondary compounds associated with beneficial activities (e.g. rosmarinic acid and allantoin) comfrey forms also toxic pyrrolizidine alkaloids (PA). To improve further breeding and study the genetic relationships of a comfrey collection, a sample set of 219 *S. officinale* and 5 hybrid plants of *S. × uplandicum* were analysed with 34 SNP markers by KASP (Kompetitive Allele Specific PCR), developed

from a next generation sequencing approach of three different individuals of *S. officinale*. In parallel, the plants were analysed for the polyphenol rosmarinic acid, the purine derivative allantoin, and the toxic pyrrolizidine alkaloids. Besides the two beneficial compounds, further 13 polyphenols and 10 purine derivatives were determined. The plants were grouped into six distinct genetic clusters. Rosmarinic acid was not linked to any of the clusters, while one cluster was distinctively different for some compounds, amongst them allantoin and globoidnan A. Also linked to a certain genetic cluster was a low content of PA, which could become a valuable gene pool for minimizing PA content by breeding. A subset of the samples was analysed in a second year again where rosmarinic acid and allantoin showed a medium stability, while globoidnan A was completely unstable.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10722-021-01156-x>.

J. Ruzicka · J. Novak (✉)
Institute for Animal Nutrition and Functional Plant Compounds, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210 Wien, Austria
e-mail: Johannes.Novak@vetmeduni.ac.at

K. Berger-Büter
Breeding Botanicals International, Emeligarten 6,
8592 Uttwil, Switzerland

N. Esslinger
Alpinamed AG, Alte Landstrasse 11, 9306 Freidorf,
Switzerland

Keywords *Symphytum* · Molecular markers · SNP · KASP · Phenolic compounds · Rosmarinic acid · Allantoin · Purine derivatives · Pyrrolizidine alkaloids

Introduction

The Euro-Siberian and Mediterranean genus *Symphytum* (comfrey, Boraginaceae) comprises 40 perennial species (Hacıoğlu 2011) of which two taxa, *S. officinale* L. (true comfrey) and *S. × uplandicum*

Nyman (*S. asperum* Lepech $\times$ *S. officinale* L., Russian comfrey), are widely used in the Austrian traditional medicine internally or externally for various disorders (musculoskeletal and gastrointestinal disorders, rheumatism and gout) (Vogl et al. 2013). Today, multiple randomized clinical trials have demonstrated the positive effects of topical comfrey treatment against pain, inflammation and swelling of muscles and joints in degenerative arthritis, acute myalgia, and sprains, contusions and strains after accidents (Staiger 2012). *S. officinale* is still amongst the five most frequently documented species in ethnoveterinary contemporary knowledge of farmers in pre-alpine and alpine regions of two Swiss cantons mostly used against injuries of the musculoskeletal system and—together with *Arnica montana* L.—against mastitis (Stucki et al. 2019). Also in British Columbia, Canada, comfrey plays an important ethnoveterinary role used against abscesses, wound and injuries, diarrhoea and scours, udder oedema and ketosis (Lans et al. 2007).

Symphytum officinale contains five groups of major compounds: carbohydrates (mucilage 29%), purine derivatives, triterpenes, polyphenols, and pyrrolizidine alkaloids. Purine derivatives (PD) are degradation products from purines, purine bases and purine nucleosides. In this group, allantoin (Fig. 1), a diureide of glyoxylic acid, is the major compound in a range of 0.6–4.7%. In wound healing, allantoin regulates the inflammatory response and stimulates the fibroblastic proliferation and the extracellular matrix synthesis resulting in a faster reestablishment

of the normal skin. In the group of polyphenolics, two compounds could be identified by HPLC/UV, rosmarinic acid and globoidnan A. Rosmarinic acid (Fig. 1), an ester of caffeic acid, exhibits antioxidant (Lin et al. 2019) and anti-inflammatory effects (Luo et al. 2020). Globoidnan A (Fig. 1), a lignan, was only recently detected in *S. officinale* (Trifan et al. 2020). In total, 13 different pyrrolizidine alkaloids (PA) were identified in *Symphytum* spp. with intermedine, lycopsamine and their corresponding N-oxides as well as echimidine-N-oxide as main PA compounds (Mädge et al. 2020; Trifan et al. 2020). Comfrey is mutagenic in liver and the PA appear to be responsible for comfrey-induced toxicity and tumour induction (Mei et al. 2010; Trifan et al. 2020).

Since comfrey possesses plant secondary compounds responsible for its medicinal activities as well as as problematic compounds in a wide variability, breeding may contribute significantly to efficacy and safety of this valuable plant. Collection of accessions and their first evaluation is the logical first step. For most medicinal plants, however, well-documented accessions, organized and maintained in gene banks, are not available. Often, breeding starts—as also happened here—with a collection where no prior information is available. In this case, studying genetic relationships with molecular markers contributes significantly to a knowledge-based breeding strategy. However, this approach is for many non-model species like comfrey, still not straightforward due to non-existent molecular markers.

Fig. 1 Structure formulae of some of the compounds analyzed (Source: PubChem—<https://pubchem.ncbi.nlm.nih.gov>)

Therefore, we first developed markers based on next-generation sequencing (NGS) approach and then determined the relationships of different accessions by single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP), a fluorescence-based genotyping technology that evolved into a global benchmark technology (Rasheed et al. 2017; Semagn et al. 2014).

Material and methods

Sample material

The sample set consisted of 219 different accessions of *S. officinale* and 5 accessions of *S. × uplandicum*. The accessions were obtained either from commercial seed or plant suppliers or via root collection from *S. officinale* plants at their wild sites. The geographical distribution of the accessions comprised entire Europe, including UK and Lithuania. All accessions were cultivated at one location in Switzerland (Uttwil). Seeds were pre-sown in May 2017 and seedlings and root propagation materials transplanted to the field in June 2017. Planting of *Symphytum* accessions was in rows with 90 cm distance between rows, 60 cm distance between plants and the first root harvest was performed in January 2018. A subset of the samples was harvested a second time in 2019.

DNA extraction

Genomic DNA for NGS-sequencing was extracted from air-dried specimens using a modified CTAB-protocol (Schmiderer et al. 2013). DNA concentrations of the extracts were determined using a NanoDropTM 2000c (ThermoFisher Scientific, USA). For KASP-analysis, seven punches of dried leaf material were taken using a plant sampling kit (LGC, Germany). DNA for complete plant genotyping was extracted by LGC (Germany).

NGS-sequencing and SNP detection

Three morphologically and phytochemically different individuals were selected to develop SNP markers by next generation sequencing approach performed by LGC (Germany). The three individuals were sequenced with Illumina NextSeq after a

normalisation step to reduce the genome size below 1 Gb. In total, 3.8 Gb were obtained from the three individuals, 61, 61 and 45 Million read pairs, respectively. The reads of one individual were de novo assembled with a kmer length of 40 using Tadpole (Bushnell) as implemented in Geneious 11.1.5 (Biomatters Ltd., New Zealand). The reads of the two other individuals were mapped to the contigs. Contigs with a good mapping coverage (> 50) were verified by a BLAST search (NCBI) and only those of nuclear origin were used as template to map individual reads for SNP detection. The ‘SNP finder’ as implemented in Geneious 11.1.5 (Biomatters Ltd, New Zealand) detected SNPs.

Primer development and KASP-analysis

The allele-specific primers for the marker candidates were designed carrying the standard FAM (5′ GAAGGTGACCAAGTTCATGCT 3′) and HEX (5′ GAAGGTCGGAGTCAACGGATT 3′) tails and with the targeted SNP at the 3′ end (Supplementary Table 1). A common primer was designed for an amplicon length of less than 120 bp. The samples were analysed at LGC (Germany) using the Kompetitive Allele Specific PCR (KASPTM) genotyping technology and raw data was checked with SNPViewer2 Version 4.0.0 (LGC, Germany).

Morphology

All plants were included in the evaluation of the morphological traits. Plant height was measured in cm. The leaf type, i.e., broad (1) or narrow (2) leaves, was determined based on the basal leaves. Growth performance was classified from 1 = vigorous to 3 = weak plants. Furthermore, the health status was classified from 1 = very healthy to 3 many disease symptoms or feeding damage on leaves.

Phytochemical analysis

Phytochemical analysis was performed with all samples in the first year. In the second year, only a subset of samples was analysed for polyphenols and purine derivatives to gain a first insight of environmental influence on some important plant secondary compounds of comfrey.

All reagents used for the HPLC (acetonitrile, formic acid and phosphoric acid) were purchased in MS-gradient-grade or reagent-grade (Sigma-Aldrich AG, Steinheim, Germany). All reference substances were purchased as certified standards (PhytoLab GmbH & Co. KG, Germany).

Polyphenols (PP) and purine derivatives (PD) on HPLC/UV

The sampled fresh roots were crushed in ethanol 70%v/v with a Heavy Duty Blender 38BL3 (Waring Products Division Dynamics Corporation of America, USA). This mixture was sucked off through a filter paper grade MN 85/70 (Machery-Nagel GmbH & Co. KG, Germany). This sample solution was stored in an amber bottle at 15–25 °C.

PD analyses were performed using hydrophilic interaction liquid chromatography (HILIC) with UV-detection using a Shimadzu Nexera X2 HPLC system (Shimadzu Cooperation, Japan) equipped with a degasser DGU-20A5R, low-pressure gradient pump LC-30AD, autosampler SIL-30AC, column oven CTO-20AC and a diode-array-detector SPD-M30A. Purine derivatives separation was performed at 25 °C by using a 150/4.6 Nucleodur 100-3 NH₂-RP column (Machery-Nagel, Germany) with a mobile phase of phosphoric acid pH 3.0 (solvent A) and acetonitrile (solvent B), using 10 µL injections. The flow rate was 1 mL min⁻¹ and the gradient was 0–6 min, 85% B; 6–7 min, 85–50% B; 7–9 min, 50% B; 9–10 min, 50–85% B; 10–15 min, 85% B (equilibration). The peaks were detected at 192 nm. All data were acquired and processed with LabSolution version 5.73 software (Shimadzu Cooperation, Japan).

Analysis of polyphenols was performed using a Kontron 500 HPLC system (Kontron Instruments, Germany) equipped with a degasser Biotech Model 2003, binary gradient solvent pump Kontron System 565, autosampler Kontron HPLC 565, column oven Column Thermostat Jetstream 2Plus and a diode-array-detector Kontron HPLC 540. Polyphenols were separated at 25 °C by using a 125/4 Nucleodur 100-5 C18 ec column (Machery-Nagel, Germany) with a mobile phase of phosphoric acid pH 2.0 (solvent A) and acetonitrile (solvent B), using 20 µL injections. The flow rate was 1 mL min⁻¹ and the gradient was 0–3 min, 12% B; 3–12 min, 12–35% B; 12–14 min, 35–80% B; 14–19 min, 80% B; 19–20 min, 80–12%

B; 20–25 min, 12% B (equilibration). The peaks were detected at 330 nm. All data were acquired and processed with Geminix version 1.91 software (Goebel Instrumentelle Analytik GmbH, Germany). Areas in PP and PD analysis were standardized to the total of the respective compound group to compare compounds and describe the profile and to the maximum response of the respective compound to check and present results of individual compounds.

Pyrrolizidine alkaloids on LC/MS

Pyrrolizidine-alkaloid analysis was performed with a modified method based on the method described by Trifan et al. (2018), using an Agilent 1260 HPLC system (Agilent Technologies, USA) equipped with a binary gradient solvent pump, degasser, autosampler and column oven, hyphenated with an AB Sciex 4500 QTRAP mass spectrometer (AB Sciex, USA) equipped with an electrospray ionisation source (ESI) and a triple quadrupole mass analyser with a trapping function. Pyrrolizidine alkaloids separation was carried out at 40 °C on an InfinityLab Poroshell 120 EC-C18 column (4.6 × 50 mm, 2.7 µm particles; Agilent Technologies, USA) with a mobile phase of 0.1% formic acid in water (solvent A) and acetonitrile (solvent B), using 5 µL injections. The flow rate was 0.8 mL min⁻¹ and the gradient was 0–1.5 min, 5% B; 1.5–2 min, 5–10% B; 2–3 min, 10% B; 3–5 min, 10–20% B; 5–6 min, 20–90% B; 6–8 min, 90% B; 8–9 min, 90–5% B; 9–13 min, 5% B (equilibration). ESI was operated in the positive ion mode: capillary temperature 400 °C, curtain gas 30 psi, nebulizer gas 40 psi, ionspray voltage 5500 V. Nitrogen was used as curtain and collision gas. MRM was used for quantitative analysis of pyrrolizidine alkaloids. The data were acquired and processed using Analyst 1.6.2 software (AB Sciex, USA). Double injections were made for each standard and for the sample. Analytes were identified by comparing retention time and m/z values obtained by MS and MS₂ with those of standards obtained under the same conditions. The calibration curves obtained in MRM mode (multiple reaction monitoring) were used for quantitation; #areas were compared with calibration curves generated by three repeated injections of known standards at six concentrations (0.1–25 ng/mL).

Statistical analysis

Data analysis was performed in R (R Core Team 2019). The marker data were analysed using the package DAPC (Discriminant Analysis of Principal Components) (Jombart et al. 2010) as implemented in poppr (Kamvar et al. 2014, 2015). Morphological and phytochemical data were calculated with FactoMinR (Lê et al. 2008) and agricolae (Mendiburu 2015).

Results and discussion

Molecular markers, morphological and phytochemical traits were analysed in 224 plants of *Symphytum* (219 *S. officinale* and 5 hybrids *S. × uplandicum*) from an ongoing breeding project to learn more about their genetic grouping and relatedness within this collection, the phenotypic variation and its relation to the genetic clusters.

Genetic analysis

In total, 34 SNP-markers were applied for the molecular part of this study (Table 1, Supplementary Table 1). Since hardly any sequence data was available for this plant, the marker candidates were developed on basis of an NGS-approach with three phenotypically very different individuals. Developing genetic markers with NGS data of at least three individuals is an efficient approach to distinguish marker candidates from sequencing errors. The normalisation step to narrow the plant DNA down to 1 Gb was necessary because the genome size of *S. officinale* is big ($2C = 3.8$ Gb, Šmarda et al. 2019). All markers cross-amplified successfully in the hybrid *S. × uplandicum*.

The markers had relatively low average diversity and expected heterozygosity of 0.21 in both parameters and a quite high Evenness of 0.67 (Table 1). Both Simpson index and expected heterozygosity also had the same range, namely from 0.0049 to 0.5, while Evenness was between 0.28 and 1.0. The highest contribution of a marker to a linear discriminant function was by marker SYM00069 with 17.4% to LD2. Only 13 markers from that random set of markers had a low impact (i.e., a relative contribution of below 5% in any of the LDA). A few markers even had a high impact on three of the four discriminant

functions (SYM00069, SYM00368 and SYM00491). It was interesting to note that a marker with low heterozygosity and Evenness (SYM00073) still had a very high impact of 12% to LD3.

In a discriminant analysis approach as proposed by Jombart et al. (2010), six clusters could be identified (Figs. 2, 3). The linear discriminant function LD1 separated very well the group of clusters 1 to 4 from the group of clusters 5 and 6. LD2 separated cluster 1 from all other clusters (Fig. 2). LD3 separated well some clusters that were intermixed in LD1 and LD2, namely the cluster pairs 3 and 4 as well as the cluster pairs 5 and 6 (Fig. 3). Three species hybrids (*S. × uplandicum*) belonged to cluster 3, while the other two were grouped into cluster 5, indicating at least two independent hybridisation events.

S. officinale and its hybrid, *S. × uplandicum* are quite interesting from the cytological point of view. In *S. officinale*, four cytotypes ($2n = 24$ (diploid, white flowering), 36, 40, 48 (tetraploid, white and purple flowering)) were detected, where $2n = 36$ could only be found in experimental hybridisations between diploid and tetraploid *S. officinale* (Gadella et al. 1974). The fertile hybrid *S. × uplandicum* is characterised by $2n = 36$. Since this hybrid is rather frequent in the natural distribution area of *S. officinale*, hybridisation (at least with *S. asperum*) seems to be common in nature, while crosses between diploid and tetraploid *S. officinale* are difficult and in most cases without success (Gadella et al. 1974). Unfortunately, the exact ploidy level of our analysis materials was unknown. However, most of the accessions were tetraploid, because a white flower colour was very rare (data not shown). In the case of a missing reference genome (as it is the case for comfrey) and polyploidy, the intensity of genotyping errors in KASP assays increases (Rasheed et al. 2017). To overcome this problem to a certain degree, the DNA was normalised to increase sequencing depth and the marker candidates were strictly preselected with only markers with clear differences in fluorescence passing.

Phenotypic analysis

Morphology

The plant height was between 50 and 125 cm, with a mean value of 90 ± 13.6 cm. *S. officinale* was on average 90.2 cm high, *S. × uplandicum* on average

Table 1 Characteristics of the SNP markers (1D ... Simpson index, H_{exp} ... expected heterozygosity, Evenness, Contr_LD1 to Contr_LD4 ... relative contribution of the marker to the respective linear discriminant function, values higher than 5% in bold)

Locus	1-D	H_{exp}	Evenness	Contr_LD1 (%)	Contr_LD2 (%)	Contr_LD3 (%)	Contr_LD4 (%)
SYM00069	0.0349	0.0350	0.3875	9.8	17.4	0.0	7.8
SYM00073	0.0049	0.0049	0.2837	0.5	0.1	11.8	0.3
SYM00080	0.1848	0.1853	0.5764	0.0	1.8	0.0	3.9
SYM00086	0.0284	0.0285	0.3734	0.1	0.4	2.1	0.1
SYM00126	0.2637	0.2643	0.6601	0.0	14.3	0.4	0.4
SYM00147	0.2241	0.2247	0.6175	0.0	0.2	0.2	6.0
SYM00148	0.3737	0.3747	0.7932	14.2	9.3	2.8	3.6
SYM00171	0.0123	0.0123	0.3246	0.0	3.6	0.2	5.5
SYM00221	0.1357	0.1361	0.5245	0.0	2.3	15.5	0.1
SYM00223	0.1569	0.1572	0.5471	0.0	0.1	0.1	1.0
SYM00226	0.3623	0.3632	0.7780	4.1	0.3	14.7	0.3
SYM00256	0.1387	0.1391	0.5278	0.0	8.5	0.0	2.0
SYM00263	0.4213	0.4229	0.8619	0.7	0.0	0.9	0.0
SYM00280	0.1816	0.1821	0.5731	0.0	0.1	9.6	4.1
SYM00323	0.1898	0.1903	0.5816	0.0	0.3	1.1	0.0
SYM00338	0.4285	0.4296	0.8732	0.0	9.8	0.1	9.7
SYM00341	0.0525	0.0526	0.4194	0.0	0.0	0.0	0.1
SYM00368	0.4754	0.4767	0.9530	13.3	0.0	5.8	5.0
SYM00408	0.2507	0.2513	0.6459	4.1	2.7	0.6	1.0
SYM00418	0.0535	0.0536	0.4210	0.0	5.1	0.0	0.5
SYM00432	0.0663	0.0664	0.4404	0.0	0.0	0.0	0.7
SYM00481	0.4673	0.4686	0.9383	0.0	12.4	0.0	0.3
SYM00491	0.4984	0.5002	0.9968	14.0	0.0	8.1	6.6
SYM00504	0.4367	0.4379	0.8864	0.1	0.0	0.0	1.7
SYM00532	0.4768	0.4780	0.9555	2.7	0.1	0.1	1.2
SYM00533	0.2685	0.2692	0.6654	0.5	1.3	0.3	0.6
SYM00557	0.4625	0.4638	0.9299	14.4	0.1	0.0	0.6
SYM00567	0.2060	0.2065	0.5985	5.5	0.4	12.8	0.6
SYM00586	0.3795	0.3806	0.8010	0.0	0.0	0.0	8.7
SYM00595	0.2571	0.2578	0.6529	0.0	0.8	4.5	0.0
SYM00610	0.0401	0.0402	0.3978	5.5	5.6	0.3	11.6
SYM00624	0.4134	0.4144	0.8499	1.8	0.2	0.0	0.0
SYM00645	0.4720	0.4735	0.9466	8.5	2.1	0.7	3.1
SYM00648	0.4444	0.4457	0.8990	0.3	0.0	7.3	12.9
Mean	0.2061	0.2067	0.6671				

90.3 cm (t test; $p = 0.97$). *S. officinale* is usually between 50 and 100 cm high, while *S. asperum* and *S. × uplandicum* are between 100 and 200 cm high (Adler et al. 1994). In our study, the five hybrids were smaller than expected. Since *S. officinale* was in the upper range of its natural variation regarding plant

height, the reason for the relatively small hybrids is not due to environmental conditions, but rather in the genetical outfit of these plants. The leaf shape (broad to narrow) was classified as follows: Most plants (40%) had quite exactly intermediary forms of leaves, while 10% of the plants had distinctively broad and

Fig. 2 Discriminant analysis of principal components (DAPC) of SNP markers (x-axis: linear discriminant function 1 (LD1), y-axis: linear discriminant function 2 (LD2)). The colours represent the six genetic clusters identified by discriminant analysis

Fig. 3 Discriminant analysis of principal components (DAPC) of SNP markers (x-axis: linear discriminant function 1 (LD1), y-axis: linear discriminant function 3 (LD3)). The colours represent the six genetic clusters identified by discriminant analysis

18% had narrow leaves. Growth performance and health status did not show obvious variation (data not shown).

Phytochemistry

In comfrey, the polyphenol rosmarinic acid and the purine derivative allantoin are partly responsible for the beneficial activities of this plant. Due to their chemical differences, they need different chromatographic settings to separate them well. Therefore, the two compound groups were analysed with two different HPLC/UV settings. Furthermore, comfrey is characterized by the presence of toxic pyrrolizidine alkaloids. Today's 'gold standard' for analysis of PA is LC/MS (Trifan et al. 2018) applied for this work.

In chromatograms of the two HPLC/UV methods, further compounds were well separated visible under the UV wavelength optimal for rosmarinic acid and allantoin, respectively, that can be classified according to their UV-spectra as 'polyphenolic compounds' (PP) in the rosmarinic acid method and 'purine derivatives' (PD) in the allantoin method (Supplementary Table 2). Because the UV wavelength selected was not optimum for each unknown compound, they may be underrepresented in the results. Even if underestimated, samples can still be compared because the generally exact quantification of UV-detectors.

In total, 20 PP and 14 PD were found in the chromatograms, but only 13 PP and 10 PD were included in statistical analyses. The other compounds were too small for a reliable quantification. Amongst the PP, the two biggest peaks could be identified as rosmarinic acid and globoidnan A, a lignan only recently identified in comfrey (Trifan et al. 2020). Globoidnan A was first time identified in *Eucalyptus globoidea* as a potent novel inhibitor of HIV integrase (Ovenden et al. 2004). In the PD profile, only allantoin could be identified. All other compounds are in the following referred to as either 'unknown PP#x' or 'unknown PD#x', where x denotes the number of the peak. The PP profile was dominated by rosmarinic acid and globoidnan A with means and ranges of 47% (20–82%) and 46% (0–70%) of total PP, respectively. All other compounds visible in the chromatogram were below 2%. Allantoin was dominating the PD profile with a mean of 67% (range 42% to 100%) of total PD, followed by PD#6 with a mean of 18% (range

0% to 42%). All other peaks were with less than 5% much lower.

Environmental influence on phytochemical composition, comparing two years

A subset of the samples was also analysed in a second year. Only a very few compounds showed a medium correlation between the two years and medium stability in the ranking order, so a certain degree of stability (Fig. 4). The most stable compounds were an unknown PD#8 and rosmarinic acid with correlation coefficients between the two years of 0.61 and 0.59, respectively. Allantoin showed medium stability with $r = 0.44$, while the lignan globoidnan A was completely unstable ($r = 0.06$).

Comparison between genotype and phenotype

In the next step, a possible reflection of the genetic grouping in phenotypic differences were analysed. From the morphological observations, plant height and leaf type showed some significant differences between the genetic clusters (Fig. 5). In plant height, clusters 1 and 6 were significantly lower by 13% and 3% than the overall mean, respectively, while plants of cluster 3 were on average 4% higher than the mean. The types of leaves were different from the dominating intermediary type in cluster 3 (significantly narrower leaves) and cluster 6 (significantly broader leaves).

The PA boxplot is characterised by a few extreme values (Fig. 5). Therefore, most of the samples are below 50% of the total range after standardisation. In the sum of the toxic PA, cluster 1 (16%) is very interesting due to the significantly lower PA contents compared to the overall mean of 29%. In economic plants, the occurrence of toxic compounds besides positive traits is of concern. PA are very dangerous toxic compounds restricting the use of *S. officinale* to external applications (Staiger 2012). For this reason, but also to minimize any risk of PA cross-contamination in the production process, reduction or complete elimination of PA is a desired trait. Fortunately, PA contents in *S. officinale* are not correlated to beneficial active compounds (data not shown). Therefore, a reduction by classical disruptive selection strategies favouring beneficial compounds is possible. Kruse et al. (2019) have demonstrated that a PA reduction by

Fig. 4 Paired boxplot graphically indicating the stability of the ranking order of some phytochemical compounds in two years (rosmarinic acid ($r = 0.59$, top left), globoidnan A ($r = 0.06$, top

right), allantoin ($r = 0.44$, bottom left) and an unknown PD 8 ($r = 0.61$, bottom right)

Fig. 5 Boxplots of significant morphological parameters plant height (top), plant habit (centre) and leaf type (bottom). Significant deviations from the overall mean are denoted on top of the respective boxplots

about 60% in *S. officinale* hairy roots by RNAi would be achievable. On the other side, the natural PA variability in *S. officinale* is so enormous that classical breeding may improve safety of products already significantly after a few selection generations.

Some PP and PD also differed significantly between clusters, namely globoidnan A and allantoin and the minor compounds PP#9, PP#15, PP#16 and PD#10 (Fig. 6). Unfortunately, plants of cluster 1 could not be analysed for their PP and PD profiles and are missing in the following presentation. Rosmarinic acid contents were not linked to any cluster, while cluster 6 had a much lower content in globoidnan A (21%) and allantoin (31%) compared to the overall means of 38% (globoidnan A) and 57% (allantoin). Furthermore, the variability was—apart from one plant—very low in this cluster. From the minor compounds (compounds with less than 5% of the total PP or PD content, respectively) many samples fell below the limit of detection. However, four of them (unknown compounds PP#9, PP#15, PP#16 and PD#10) still showed a correlation to genetic clustering, all linked—as before with globoidnan A and

allantoin—to cluster 6. In PP#9 and PD#10 cluster 6 differed from all other clusters with all plants below the limit of detection. However, PP#15 and PP#16 were in cluster 6 significantly higher, all above the limit of detection. The genetic cluster 6 is regarding its phytochemical profile outstanding and easy to differentiate. For cosmetic/medicinal use, however, this cluster is not favourable for further breeding activity due to its low allantoin content and the low variability in rosmarinic acid. The recently in *S. officinale* detected globoidnan A (Trifan et al. 2020) is a highly bioactive lignan (Ovenden et al. 2004). Possible beneficial or adverse effects in comfrey need to be elucidated. In the case, this lignan would be undesired, cluster 6 could be a perfect gene pool to decrease globoidnan A. Besides the interesting correlation between genetic clusters and phytochemical composition, the study shows the enormously high variation of plant secondary compounds, making comfrey an interesting target for efficient selection for rosmarinic acid and allantoin. Cluster 1 could become a very interesting donor for low PA content in a classical

Fig. 6 Boxplot of PP and PD compounds with significant deviation from the overall mean, rosmarinic acid (a), globoidnan A (b), unknown PP#9 (c), unknown PP#15 (d), unknown PP#16 (E), allantoin (F) and unknown PD#10 n(G)

crossbreeding approach. Some markers developed here could be used to guarantee outcrossing.

Author contributions All authors contributed to the study conception and design. Maintaining collection, sample collection and morphological analysis: KBB. All phytochemical analysis was performed by NE. Genetic Analysis was performed by JR. Data analysis of joint data sets and the first draft of the manuscript were performed by JR and JN and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by University of Veterinary Medicine Vienna. This study was funded by Alpinamed.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adler W, Oswald K, Fischer R, Fischer MA (eds) (1994) Exkursionsflora von Österreich: Bestimmungsbuch für alle in Österreich wildwachsenden sowie die wichtigsten kultivierten Gefäßpflanzen (Farnpflanzen und Samenpflanzen) mit Angaben über ihre Ökologie und Verbreitung. Verlag Eugen Ulmer, Stuttgart
- Bushnell B BBMap. sourceforge.net/projects/bbmap/. Accessed 30 March 2020
- Gadella TWJ, Kliphuis E, Perring FH (1974) Cytotaxonomic studies in the genus *Symphytum*. VI. Some notes on *Symphytum* in Britain. *Acta Bot Neerl* 23:433–437. <https://doi.org/10.1111/j.1438-8677.1974.tb00960.x>
- Hacıoğlu (2011) Phylogeny of *Symphytum* L. (Boraginaceae) with special emphasis on Turkish species. *Afr J Biotechnol*. <https://doi.org/10.5897/AJB11.1094>
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94. <https://doi.org/10.1186/1471-2156-11-94>
- Kamvar ZN, Tabima JF, Grunwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2013:1–14. <https://doi.org/10.7717/peerj.281>
- Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet*. <https://doi.org/10.3389/fgene.2015.00208>
- Kruse LH, Stegemann T, Jensen-Kroll J, Engelhardt A, Wesseling A-M, Lippert A, Ludwig-Müller J, Ober D (2019) Reduction of pyrrolizidine alkaloid levels in comfrey (*Symphytum officinale*) hairy roots by RNAi silencing of homospermidine synthase. *Planta Med* 85:1177–1186. <https://doi.org/10.1055/a-0998-5125>
- Lans C, Turner N, Khan T, Brauer G, Boepple W (2007) Ethnoveterinary medicines used for ruminants in British Columbia. *Can J Ethnobiol Ethnomed* 3:11. <https://doi.org/10.1186/1746-4269-3-11>
- Lê S, Josse J, Husson F (2008) FactoMineR: an R Package for multivariate analysis. *J Stat Soft*. <https://doi.org/10.18637/jss.v025.i01>
- Lin C, Xiao J, Xi Y, Zhang X, Zhong Q, Zheng H, Cao Y, Chen Y (2019) Rosmarinic acid improved antioxidant properties and healthspan via the IIS and MAPK pathways in *Caenorhabditis elegans*. *BioFactors* 45:774–787. <https://doi.org/10.1002/biof.1536>
- Luo C, Zou L, Sun H, Peng J, Gao C, Bao L, Ji R, Jin Y, Sun S (2020) A Review of the anti-inflammatory effects of rosmarinic acid on inflammatory diseases. *Front Pharmacol* 11:153. <https://doi.org/10.3389/fphar.2020.00153>
- Mädge I, Gehling M, Schöne C, Winterhalter P, These A (2020) Pyrrolizidine alkaloid profiling of four Boraginaceae species from Northern Germany and implications for the analytical scope proposed for monitoring of maximum levels. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* 37:1339–1358. <https://doi.org/10.1080/19440049.2020.1757166>
- Mei N, Guo L, Fu PP, Fuscoe JC, Luan Y, Chen T (2010) Metabolism, genotoxicity, and carcinogenicity of comfrey. *J Toxicol Environ Health B Crit Rev* 13:509–526. <https://doi.org/10.1080/10937404.2010.509013>
- Mendiburu FD (2015) agricolae: Statistical procedures for agricultural research. R Package Version 1.2-3. CRAN.R-project.org/package=agricolae
- Ovenden SPB, Yu J, Wan SS, Sberna G, Tait RM, Rhodes D, Cox S, Coates J, Walsh NG, Meurer-Grimes BM (2004) Globoidnan A: a lignan from *Eucalyptus globoides* inhibits HIV integrase. *Phytochemistry* 65:3255–3259. <https://doi.org/10.1016/j.phytochem.2004.10.006>
- R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol Plant* 10:1047–1064. <https://doi.org/10.1016/j.molp.2017.06.008>
- Schmiderer C, Lukas B, Novak J (2013) Effect of different DNA extraction methods and DNA dilutions on the amplification

- success in the PCR of different medicinal and aromatic plants. *Z Arnzei und Gewürzpfla* 18:65–72
- Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breed* 33:1–14. <https://doi.org/10.1007/s11032-013-9917-x>
- Šmarda P, Knápek O, Březinová A, Horová L, Grulich V, Danihelka J, Veselý P, Šmerda J, Rotreklová O, Bureš P (2019) Genome sizes and genomic guanine+cytosine (GC) contents of the Czech vascular flora with new estimates for 1700 species. *Preslia* 91:117–142. <https://doi.org/10.23855/preslia.2019.117>
- Staiger C (2012) Comfrey: a clinical overview. *Phytother Res* 26:1441–1448. <https://doi.org/10.1002/ptr.4612>
- Stucki K, Cero MD, Vogl CR, Ivemeyer S, Meier B, Maeschli A, Hamburger M, Walkenhorst M (2019) Ethnoveterinary contemporary knowledge of farmers in pre-alpine and alpine regions of the Swiss cantons of Bern and Lucerne compared to ancient and recent literature—is there a tradition? *J Ethnopharmacol* 234:225–244. <https://doi.org/10.1016/j.jep.2018.12.022>
- Trifan A, Opitz SEW, Josuran R, Grubelnik A, Esslinger N, Peter S, Bräm S, Meier N, Wolfram E (2018) Is comfrey root more than toxic pyrrolizidine alkaloids? Salvianolic acids among antioxidant polyphenols in comfrey (*Symphytum officinale* L.) roots. *Food Chem Toxicol* 112:178–187. <https://doi.org/10.1016/j.fct.2017.12.051>
- Trifan A, Wolfram E, Esslinger N, Grubelnik A, Skalicka-Woźniak K, Minceva M, Luca SV (2020) Globoidnan A, radosiin and globoidnan B as new phenolic markers in European-sourced comfrey (*Symphytum officinale* L.) root samples. *Phytochem Anal* 1:21. <https://doi.org/10.1002/pca.2996>
- Vogl S, Picker P, Mihaly-Bison J, Fakhrudin N, Atanasov AG, Heiss EH, Wawrosch C, Reznicek G, Dirsch VM, Saukel J, Kopp B (2013) Ethnopharmacological in vitro studies on Austria's folk medicine—an unexplored lore in vitro anti-inflammatory activities of 71 Austrian traditional herbal drugs. *J Ethnopharmacol* 149:750–771. <https://doi.org/10.1016/j.jep.2013.06.007>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.