

# Marker-trait association analysis for agronomic and compositional traits in sainfoin (*Onobrychis viciifolia*)

K. Kempf<sup>1</sup>, C.S. Malisch<sup>2</sup>, C. Grieder<sup>3</sup>, F. Widmer<sup>1</sup> and R. Kölliker<sup>1,4</sup>

 <sup>1</sup>Molecular Ecology, Agroscope, Zurich, Switzerland
<sup>2</sup>Forage Production and Grassland Systems, Agroscope, Zurich, Switzerland
<sup>3</sup>Fodder Crop Breeding, Agroscope, Zurich, Switzerland
<sup>4</sup>Molecular Plant Breeding, Institute for Agricultural Sciences, ETH Zürich, Zurich, Switzerland

Corresponding author: R. Kölliker E-mail: roland.koelliker@usys.ethz.ch

Genet. Mol. Res. 16 (1): gmr16019483 Received October 27, 2016 Accepted December 12, 2016 Published January 23, 2017 DOI http://dx.doi.org/10.4238/gmr16019483

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Sainfoin (*Onobrychis viciifolia*) is a perennial forage legume with great potential for use in sustainable agriculture due to its low input requirements, good drought tolerance, and production of forage rich in polyphenolic compounds, which are beneficial for animal health. However, its distribution and cultivation are limited due to its moderate agronomic performance and a general lack of well adapted, highly yielding cultivars. Faster progress in breeding is imperative, but is often hampered by the complex inheritance of traits and limited knowledge on the genetic composition of this tetraploid, outbreeding species. Molecular genetic tools might aid phenotypic selection; however, to date no information on marker-trait associations is available for sainfoin. Hence, the goal of the present study was to detect marker-

Genetics and Molecular Research 16 (1): gmr16019483

trait associations in a biparental  $F_1$  population. Single plants were screened for recently developed genetic markers and phenotyped for important agronomic traits and concentrations of different polyphenolic compounds. Significant trait-associated markers (TAM) were detected for plant height (11), plant vigor (1), and seed yield (7). These three traits were positively correlated with each other and shared some TAMs. Correlations among markers suggested that two independent loci control these three vigor-related traits. One additional, independent TAM was detected for the share of prodelphinidins in total condensed tannins. Our results provide insight into the genetic control of important traits of sainfoin, and the TAMs reported here could assist selection in combination with phenotypic assessment.

**Key words:** Sainfoin; *Onobrychis viciifolia*; Marker-trait associations; Simple sequence repeats; Sequence related amplified polymorphisms

# **INTRODUCTION**

The perennial, tetraploid legume sainfoin (Onobrychis viciifolia) has a long tradition as a valued forage source for working horses and other farm animals in Europe. With the cultivation of more productive species, application of inorganic fertilizers and the use of automated harvest systems, its cultivation decreased during the 20th century. Consequently, sainfoin has almost disappeared from current agricultural landscapes. However, the recently increased interest in more sustainable agricultural production makes this crop a valuable alternative for various reasons. First, sainfoin shows high tolerance to drought and low nutrient availability (Doyle et al., 1984), making it a valuable option for less favorable sites, which are expected to increase due to larger climatic variability accompanying climate change. Second, it is capable of fixing atmospheric nitrogen via association with rhizobia, reducing the need for nitrogen fertilization and improving soil quality (Hill, 1980). Third, as its most prominent feature, sainfoin contains a high concentration of beneficial secondary metabolites such as condensed tannins (CT, proanthocyanidins) and other polyphenolic compounds, which have anthelmintic properties against gut parasites, increase protein utilization, prevent bloating, and reduce methane emissions (McMahon et al., 2000; Hoste et al., 2012). Hence, sainfoin has the potential to be a valuable ruminant feed.

However, despite its advantages, the wide cultivation of sainfoin is hampered by its often poor agronomic performance and the limited availability of varieties adapted to temperate climates (Goplen et al., 1991). Major agronomic weaknesses also include poor competitive ability against weeds or companion species, slow establishment, low seed set, and weak persistence. Persistence is a complex trait that depends on several factors, of which plant vigor (strong and healthy growth) is a major determinant. In addition to plant vigor, plant height, leaf area, disease resistance, and seed production are also indicators for strength and capacity to survive and, consequently, to persist. Breeding improved varieties with regard to these agronomic traits is indispensable to increase farmers' acceptance of sainfoin and to promote its cultivation. Furthermore, targeted breeding for desired composition and concentrations of CT and other polyphenolic compounds in sainfoin is also desirable, because nutraceutical properties are consistent with the demands for future plant breeding to assist animal welfare and health (Humphreys, 2005).

Genetics and Molecular Research 16 (1): gmr16019483

Sainfoin breeders have to contend with an allogamous (outbreeding) reproduction system, as is common in most forage species, and is often secured through self-incompatibility mechanisms (Humphreys, 2000; Posselt, 2010). Hence, under the absence of efficient methods to induce self-fertilization or perform controlled crosses between inbred individuals, varieties are classically bred as open pollinated varieties or synthetics. Starting from selection of superior plants within a base population, this process of developing a new variety takes at least ten years (Posselt, 2010). The resultant varieties exhibit high levels of genetic diversity, evoking a generally broad environmental adaptability. For sainfoin breeding, this large diversity presents an opportunity to select for superior genotypes improved for important traits such as persistence, seed yield and concentrations of CT, and other polyphenolic compounds (flavonols and their glycosides).

To date, information on the response of sainfoin to phenotypic selection and inheritance patterns of traits is scarce. Additionally, measuring complex traits, such as CT concentration and composition, is both complex and labor intensive (Engström et al., 2014). In addition, determination of seed yield and plant vigor is labor intensive and usually requires field experiments over several years. Indirect selection can accelerate breeding, and can be performed using traits that are easily determined and correlated with the target trait, or via genetic markers linked to the trait of interest [marker assisted selection (MAS)]. The basic prerequisite for MAS is the availability of markers that are tightly linked to genes or quantitative trait loci (QTL). Such markers could be used to select for traits that are difficult to measure or dependent on the developmental stage, to maintain recessive alleles, to pyramid monogenic traits, and to speed up the breeding progress in general. In the case of polyploid species and traits subject to polygenic inheritance, both factors impede the efficient selection of genotypes based on their phenotype, and application of MAS would be especially valuable (Barrett et al., 2009). Detection of potential QTLs in sainfoin would, therefore, be particularly promising to improve breeding progress.

In other forage plant species, several QTLs suitable for MAS have been identified, such as those for vigor and seed yield in white clover (Trifolium repens; Faville et al., 2003), seed yield in red clover (Trifolium pratense; Herrmann et al., 2006), or height and regrowth in alfalfa (Medicago sativa; Robins et al., 2007). In sainfoin, adequate numbers of markers are not yet available and, consequently, no studies have been performed to detect marker-trait associations or QTLs. Marker systems that do not rely on prior DNA sequence information offer a valuable approach for the initial characterization of minor species that generally lack available sequence information. Sequence-related amplified polymorphism (SRAP), which preferentially amplify in gene rich regions (Li and Quiros, 2001), have been successfully used for the molecular investigation of allogamy and autogamy in sainfoin (Kempf et al., 2015). Recently, RNA transcriptome sequencing, using next-generation sequencing, has been successfully used to develop comprehensive simple sequence repeats (SSR) marker resources that can be applied to assess genetic diversity in sainfoin (Kempf et al., 2016; Mora-Ortiz et al., 2016). In order to use these newly developed marker resources to identify possible marker trait associations in sainfoin, the objectives of the present study were to i) phenotype a biparental F, population of sainfoin for a wide range of traits under field conditions, ii) investigate correlations among agronomic and compositional traits, and iii) identify dominant SRAP and co-dominant SSR markers associated with phenotypic traits. Hence, the present study provides a first step towards MAS in sainfoin.

Genetics and Molecular Research 16 (1): gmr16019483

# **MATERIAL AND METHODS**

# **Plant material**

The plant material used in the present study consisted of a  $F_1$  population (N = 122) from a biparental cross between two plants of the varieties Brunner (parent 1; Agroscope, Zurich, Switzerland) and Perdix (parent 2, Agroscope, Nyon, Switzerland), and the first generation of plants resulting from self-fertilization (S<sub>1</sub> progenies) of the two parents (N = 85 for parent 1; N = 30 for parent 2). Crossings (F<sub>1</sub> progenies) and selfings (S<sub>1</sub> progenies) have previously been identified using SSR and SRAP markers (Kempf et al., 2015). The plant material was established in January 2012 by placing two sets of five clones of each parent in a greenhouse chamber for bumble bee pollinations (*Bombus terrestris*, "Bombus Maxi Hummeln", Andermatt Biocontrol, Switzerland). Seeds were harvested separately for each parent. Plants from pre-germinated seeds were then kept in the greenhouse for two months before being transplant as juvenile plants on the field site (Delley, Fribourg, Switzerland) in July 2012. All plants were randomly planted into rows of 10 plants, arranged in two rectangular blocks.

## Phenotyping

Plants were grown in the field from 2012 to 2014, chemical composition was determined in 2012, and phenotyping of agronomic traits was performed in 2013. Due to a large loss of plants during the winter following the 2013 season, no phenotyping could be performed in 2014. Plant height was determined as the distance from the soil to the last leaflets of stretched plants in spring (HSp) and summer (HSu). Plant vigor (Vigor) was scored visually on a scale from 1 (very weak) to 9 (very strong). Susceptibility to rust disease (caused by *Uromyces onobrychidis*) was also visually scored on a scale from 1 (non-susceptible, no disease occurrence) to 9 (highly susceptible, high disease occurrence). Flowering time (FT) was recorded as the number of days after May 1st, when a plant showed three or more open flowers. Seed yield was measured as the total number of seeds per plant in 2013 (SN), as described by Kempf et al. (2015).

In spring 2013, the chlorophyll content of leaves was measured non-destructively using a SPAD chlorophyll meter (Konica Minolta SPAD 502Plus, Tokyo, Japan). This device measures the transmittance of the leaf in the red and infrared regions reporting a unit-less index, which correlates with the relative quantity of chlorophyll in the leaf tissue (Castelli et al., 1996). SPAD values (SPAD) were recorded on the uppermost fully developed leaf of each single plant, calculating the average of three measurements taken on three different leaflets outside the middle lamella. After taking SPAD measurements, the respective leaves were cut and their area (LA, cm<sup>2</sup>) and length (LL, cm) were determined photographically: leaves were laid planar on a white paper and photographed with a full-frame (21 megapixel) digital single-lens reflex (DSLR) camera (EOS 5D MkII, Canon Inc., Tokyo, Japan), which was converted to a three-band vegetative stress camera (LDP LLD, Carlstadt, NJ, USA, www.maxmax. com), reporting normalized difference vegetation index (NDVI) values. On the basis of NDVI values, the image was separated into plant (high NDVI) and background (low NDVI) pixels, using a custom script within the R-environment (R Development Core Team, 2014).

For chemical analysis, plant material (10 fully developed leafs per plant) was sampled in 2012 and immediately freeze dried. The concentrations of prodelphinidins and procyanidins,

Genetics and Molecular Research 16 (1): gmr16019483

the flavonols quercetin, kaempferol, and myricetin (including their glycosides), as well as the polyphenolic acids quinic acid and gallic acid, were analyzed using ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS). The analysis was performed on an Acquity UPLC system (Waters Corp., Milford, MA, USA) coupled with a Xevo TQ triplequadrupole mass spectrometer (Waters Corp., Milford, MA, USA) according to the methods described by Engström et al. (2014). The concentration of total CTs was calculated as the sum of the concentrations of its sub-constituents, prodelphinidins and procyanidins. The share of prodelphinidins in the total CT (PD-share) was calculated by dividing the concentration of prodelphinidins by the concentration of total CT.

# Genotyping

A subsample of the plant material used for chemical analysis was taken for DNA analysis. The dried leaves were ground using a ball mill (Cell tissue Analyzer 2, Qiagen, Hilden, Germany). DNA was extracted with the Illustra<sup>™</sup> DNA Extraction Kit PHYTOPURE (GE Healthcare, Little Chalfont, Buckinghamshire, United Kingdom) according to the manufacturer instructions. DNA concentration was measured by gel electrophoresis with a mass standard (High DNA Mass Ladder, Invitrogen<sup>™</sup>, Life Technologies, Carlsbad, CA, USA). For genotyping, SRAP markers and SSR markers were employed. For the SRAP marker system, four fluorescently labeled forward (me1 to me4) and reverse (em1 to em4) primers were applied in 16 combinations in all plants, while sainfoin specific primers (Mora-Ortiz et al., 2016) were used for SSR analysis (Table S1). PCR and fragment analysis were performed as described in Kempf et al. (2015) using an iCycler (Bio-Rad, Hercules, CA, USA) and an ABI 3500 Genetic Analyzer (Applied Biosystems, Forster City, CA, USA). Allele counts for the SRAP markers were those used previously to analyze self-fertilization (Kempf et al., 2015), and the approach of counting alleles present in one parent and absent in the other was followed. SSR marker data was analyzed *de novo*.

#### Statistical analysis

Trait-associated markers (TAM) were detected by marker regression within crossings and selfings of parent 2 [due to the low number of plants (N = 30), selfings of parent 1 were omitted from the analysis]. For the analysis, all SRAP and SSR alleles were coded based on their presence (1) or absence (0). P values of marker-trait correlations were obtained with one degree of freedom score test as implemented in the mmscore function of R-package GenABEL (Aulchenko et al., 2007; Chen and Abecasis, 2007). To correct for multiple testing effects, Benjamini-Hochberg q-values were calculated based on the P values obtained using the function qvaluebh95 to obtain significance values for markers at a false discovery rate of 5%. Additionally, population structure within crossings and selfings was investigated. For this, pairwise genetic distances between individuals were calculated based on all marker alleles using the modified Rogers' distance D<sub>w</sub> (Wright, 1978), which shows the extent of genetic diversity between two plants (Lee et al., 1989), ranging from 0 to 1. Principal coordinate (PCO) analysis was then performed on D<sub>w</sub> values using the R function cmdscale (R Development Core Team, 2014). PCO analysis indicated the presence of certain unexpected population structures in crossings (Figure 1A), which demanded a correction during marker regression. For this, the kinship matrix was calculated (only for crosses) as the proportion of

Genetics and Molecular Research 16 (1): gmr16019483

shared alleles among plants (Eding and Meuwissen, 2001) using the R function ibs and was considered as a variance-covariance matrix to correct for random genotypic effects, using the maximum likelihood implementation in the polygenic function of the GenABEL package. No population structure was visible in neither selfing group (Figure 1B, C). Therefore, kinship was not corrected for during marker regression in selfings of parent 2.



**Figure 1.** Principle coordinate analysis for: **A.**  $122 \text{ F}_1$  progenies derived from a cross between parent 1 and parent 2; **B.**  $30 \text{ S}_1$  progenies of parent 1; and **C.**  $85 \text{ S}_1$  progenies of parent 2, based on 170 SRAP and 208 SSR marker alleles.

For traits with significant TAM, one-way ANOVA was performed to determine the proportion of explained variance and the effect of the TAM (difference between plants showing the presence and absence of the respective TAM). Correlations among phenotypic traits were calculated as Pearson's correlation coefficients. All statistical analyses and calculations were performed within the R-environment (R Development Core Team, 2014).

# RESULTS

# Large variation in agronomic traits

All agronomical traits showed clear segregation among the crossings ( $F_1$  progenies), with observed values following a nearly normal distribution except for number of seeds (SN), which was positively skewed due to a large number of plants having less than 500 seeds per plant (Figure 2). Therefore, a logarithmic transformation was applied to SN [ln(SN)] when testing for TAM. To a lesser extent, rust susceptibility also deviated from normal distribution, caused by many plants showing a low rust susceptibility (score = 2). SN per plant ranged from 0 to 4608 and large variation was also observed for vigor-related traits such as Vigor, HSu, LA, and LL (Figure 2, Table 1). The difference between the earliest and latest flowering plant, as assessed by FT, was 46 days. For agronomic traits, the observed ranges of values were comparable between crosses and selfings except for SN, where selfings (mean/max value of 316/1114 in selfings of parent 1 and 447/1483 in selfings of parent 2) showed largely reduced seed numbers per plant compared to crosses (mean/max values of 1265/4608; Figures S1 and S2).

Genetics and Molecular Research 16 (1): gmr16019483



**Figure 2.** Frequency distribution of agronomic traits and concentration of phenolic compounds in  $122 \text{ F}_1$  progenies derived from a cross between parent 1 and parent 2. For abbreviations of traits, see Table 1.

**Table 1.** Minimum, mean, and maximum observed values with standard deviation (SD), and number of plants with observations (# Ind) of agronomical and compositional traits examined in  $F_1$  progenies derived from of a cross between parent 1 and parent 2.

Trait		Range of obser	rved values	# Ind	# sig Marker		
	Min	Mean	Max	SD		SRAP	SSR
FT (days of May)	19.0	41.9	65.0	9.5	116	0	0
HSp (cm)	8.0	14.9	24.5	3.8	118	0	0
HSu (cm)	46.0	86.6	129.0	18.4	118	5	6
Vigor	1.0	6.7	9.0	1.7	118	0	1
SN	0	1265	4608	1254	118	1	6
LA (cm <sup>2</sup> )	3.6	12.7	31.8	5.5	112	0	0
LL (cm)	3.8	10.8	16.8	2.3	112	0	0
SPAD	33.0	46.4	57.1	4.7	117	0	0
Rust	0.0	2.6	7.0	1.7	111	0	0
CT (mg/g)	9.7	27.0	47.2	7.0	122	0	0
PD-share	0.76	0.86	0.93	0.04	122	1	0
Quercetin (mg/g)	5.8	12.1	32.1	4.4	122	0	0
Kaempferol (mg/g)	0.25	1.80	5.59	0.86	122	0	0
Myricetin (mg/g)	0.30	1.15	4.30	0.54	122	0	0
Quinic acid (mg/g)	0.1	3.8	10.2	2.0	122	0	0
Gallic acid (mg/g)	0.09	0.16	0.33	0.05	122	0	0

The number of sequence related amplified polymorphism (SRAP) and simple sequence repeat (SSR) markers significantly associated with the traits are shown (# sig Marker). Agronomical traits are time of flowering (FT), plant height in spring and summer (HSp, HSu), plant vigor (Vigor), seed number per plant (SN), leaf area and length (LA, LL), chlorophyll content (SPAD), and susceptibility to rust diseases (Rust). Among compositional traits, CT gives the total concentration of condensed tannins and PD-share the share of prodelphinidins on the total CT.

Genetics and Molecular Research 16 (1): gmr16019483

# Quantity and quality of condensed tannin varies among plants

Concentrations of polyphenolic compounds observed for crossings also showed a clear segregation, and were close to the normal distribution (Figure 2). For CT, concentrations ranged from 9.7 to 47.2 mg/g plant material, with a mean value of 27.0 mg/g, while the PD-share ranged from 0.76 to 0.93. (Table 1). Concentrations of polyphenolic compounds were also assessed in the parental plants, whereby parent 1 (Brunner) showed a CT concentration of 22.3 mg/g with PD-share of 0.88, while the corresponding values for parent 2 (Perdix) were 17.2 mg/g and 0.83, respectively (data not shown). The observed differences in the concentrations of polyphenolic compounds between crossings of parent 1 and parent 2 were also reflected by their respective selfings (Figures S1 and S2). For all compounds except for quercetin, where the difference in the concentration between parents was low, selfings of the parent with the lower concentration (data not shown). Apart from that, the range of the compositional traits was comparable between selfings and crosses.

#### Strong correlations among vigor related traits

Among agronomical traits, significant, positive correlations were observed among Vigor, HSp, HSu, and SN, with the correlation being strongest between HSp and HSu, (r = 0.76) and weakest between Vigor and HSp (r = 0.59; Table 2). These vigor-related traits were also positively correlated to LA and LL, but to a lower extent. FT was most strongly correlation with ln(SN) (r = -0.34), with later flowering plants exhibiting a lower seed yield.

**Table 2.** Phenotypic correlations among agronomical traits in 122 crossings ( $F_1$  progenies) derived from a cross between parent 1 and parent 2. Agronomical traits are plant height in spring and summer (HSp, HSu), time of flowering (FT), seed number per plant (SN), chlorophyll content (SPAD), susceptibility to rust diseases (Rust), leaf area and leaf length (LA, LL).

Trait	Correlated trait												
	HSp	HSu	FT	ln(SN)	Rust	SPAD	LA	LL					
Vigor	0.59**	0.62**	-0.23*	0.75**	0.18	0.26**	0.26**	0.18					
HSp		0.76**	-0.15	0.47**	0.09	0.36**	0.28**	0.22*					
HSu			-0.02	0.61**	0.10	0.47**	0.25**	0.20*					
FT				-0.34**	0.08	-0.09	0.20*	0.29**					
ln(SN)					0.30**	0.22*	0.20*	0.14					
Rust						0.14	0.23*	0.31**					
SPAD							0.14	0.11					
LA								0.78**					

\*P < 0.05; \*\*P < 0.01.

# Weak correlations between agronomic and compositional traits

Among polyphenolic compounds, CT concentration was positively correlated with PD-share (r = 0.60), indicating that prodelphinidins were most important for total CT concentration levels (Table 3). CT also showed weak, positive correlations to the other compounds, with the exception of kaempferol, which, together with myricetin, was negatively correlated to quinic acid. Correlations between agronomic and compositional traits were only significant in a few instances (Table 3). Flowering time showed weak-to-

Genetics and Molecular Research 16 (1): gmr16019483

moderate correlations with chemical composition, whereby later flowering was associated with higher concentrations of CT, quinic acid, and gallic acid, and lower concentrations of kaempferol and myricetin. Furthermore, Vigor, HSp, and SPAD showed weak, negative correlations to gallic acid concentrations. Correlations between selfings of parent 1 and parent 2 generally had lower significance levels due to the lower number of plants used (Tables S2 and S3). However, the patterns of correlations observed were generally similar to those observed in crossings.

**Table 3.** Phenotypic correlations among compositional traits and between compositional and agronomic traits in 122 crossings ( $F_1$  progenies) derived from a cross between parent 1 and parent 2.

Trait	Correlated trait									
	СТ	PD-share	Quercetin	Kaempferol	Myricetin	Quinic acid	Gallic acid			
Vigor	0.12	0.12	0.01	0.09	0.02	0.03	-0.25**			
HSp	0.21*	0.21*	-0.10	0.07	-0.05	0.15	-0.33**			
HSu	0.17	0.00	-0.20*	-0.06	-0.16	0.18	-0.17			
FT	0.30**	0.04	-0.15	-0.34**	-0.33**	0.27**	0.26**			
ln(SN)	0.19*	0.13	0.04	0.13	0.18	-0.01	-0.05			
Rust	0.20*	0.23*	-0.06	0.05	-0.08	0.13	0.03			
SPAD	-0.04	-0.06	-0.17	0.04	-0.23*	-0.07	-0.24*			
LA	0.1	0.11	-0.10	-0.21*	-0.02	0.11	0.05			
LL	0.18	0.13	-0.21*	-0.22*	-0.07	0.17	0.09			
СТ		0.60**	0.33**	0.04	0.22*	0.26**	0.26**			
PD-share			0.29**	0.32**	0.28**	0.18	0.21*			
Quercetin				0.32**	0.52**	-0.12	0.03			
Kaempferol					0.32**	-0.28**	-0.15			
Myricetin						-0.35**	0.07			
Quinic acid							0.03			

\*P < 0.05; \*\*P < 0.01. Agronomical traits are time of flowering (FT), plant height in spring and summer (HSp, HSu), plant vigor (Vigor), seed number per plant (SN), leaf area and length (LA, LL), chlorophyll content (SPAD), and susceptibility to rust diseases (Rust). Among compositional traits, CT gives the total concentration of condensed tannins and PD-share the share of prodelphinidins on the total CT.

#### TAMs were detected for three agronomic and one compositional traits

Across all individuals, 173 and 189 alleles were detected for the 16 SRAP and 48 SSR markers, respectively. In crossings, TAMs were detected for SN (7 TAM), HSu (11 TAM), Vigor (1 TAM), and PD-share (1 TAM, Tables 1 and 4), whereas no TAMs were found within selfings of parent 2 (data not shown). Analyses of logarithmic- and non-transformed values of SN revealed significant values for the same markers (data not shown), thus untransformed SN values are reported for simplicity. Among the observed TAMs, SSR markers were more frequently found to be linked to phenotypic traits (13 TAM) when compared to SRAP markers (7 TAM). TAMs associated to SN explained between 1.1 and 28.1% of the variation, where the presence of the allele had a positive effect on all SSR marker alleles +991 to 1448 seeds per plant) and a negative effect on the SRAP marker allele (-261 seeds per plant; Table 4). For HSu, values of explained variance for the 11 TAMs ranged from 4.4 to 28.6%, with the presence of the allele always having a positive effect and resulting in plants that were 7.9 to 25.6 cm taller. For compositional traits, only one TAM could be found for PD-share, with the SRAP marker allele explaining 12.3% of the observed phenotypic variance, and resulting in a slightly higher share of prodelphinidins.

Genetics and Molecular Research 16 (1): gmr16019483

**Table 4.** P value, proportion of explained phenotypic variance (% expl. Var.), and effect size (Effect) of trait associated marker (TAM) alleles identified in  $F_1$  progenies derived from of a cross between parent 1 and parent 2. Traits are seed number per plant (SN), plant height in summer (HSu), plant vigor (Vigor), and the share of prodelphinidins on the total amount of condensed tannins (PD-share).

Trait	Marker	Allele size	P value	% expl. Var.	Effect
SN (seeds/plant)					
	OVK_183	274	1.0E-06***	28.1	1447.93
	OVK_141	251	3.0E-04***	24.9	1352.43
	OVK_073	204	4.4E-06***	24.4	1333.25
	OVK_133	205	1.0E-04***	9.9	1025.61
	OVK_155	255	1.0E-04***	9.9	1025.61
	OVK_111	223	6.2E-06***	9.3	991.47
	me4em3	356	1.0E-03***	1.1	-260.66
HSu (cm)					
	OVK_111	223	4.7E-06***	28.6	25.62
	me1em1	299	1.0E-03***	27.3	25.46
	me1em1	175	6.0E-04***	25.4	23.66
	OVK_133	205	0.0011***	24	23.43
	OVK_155	255	0.0011***	24	23.43
	me4em1	104	4.0E-04***	22.8	19.12
	OVK_063	191	1.4E-03***	21.3	20.99
	me4em3	127	1.1E-03***	20	19.81
	OVK_183	274	2.0E-04***	8.2	11.45
	me1em1	217	1.0E-04***	7.6	10.12
	OVK_073	189	8.0E-04***	4.4	7.87
Vigor	OVK_183	274	4.5E-05***	23.9	1.77
PD-share	me4em4	525	1.0E-04***	12.3	0.02

\*\*\*P < 0.001.

For the majority of TAMs detected in crossings, the marker allele was transferred from one parental plant only (data not shown). This is also reflected by the occurrence of the marker alleles in respective selfings, whereby almost all plants derived from one parent generally possessed the allele and plants derived from the other parent generally did not (Figures 3-5).



**Figure 3.** Seed number per plant (SN) dependent on the presence (P) and absence (NP) of the marker allele of TAMs significantly associated with the trait for  $122 \text{ F}_1$  progenies (Crossings) derived from a cross between parent 1 and parent 2, 30 S<sub>1</sub> progenies of parent 1 (Selfings P1), and 85 S<sub>1</sub> progenies of parent 2 (Selfings P2).

Genetics and Molecular Research 16 (1): gmr16019483



**Figure 4.** Plant height during summer (Hsu; in cm) dependent on the presence (P) and absence (NP) of the marker allele of TAM significantly associated with the trait for  $122 \text{ F}_1$  progenies (Crossings) derived from a cross between parent 1 and parent 2, 30 S<sub>1</sub> progenies of parent 1 (Selfings P1), and 85 S<sub>1</sub> progenies of parent 2 (Selfings P2).



**Figure 5.** Plant vigor (Vigor) and share of prodelphinidin on total condensed tannins (PD-share) dependent on the presence (P) and absence (NP) of the marker allele of TAM significantly associated with the trait for 122  $F_1$  progenies (Crossings) derived from a cross between parent 1 and parent 2, 30  $S_1$  progenies of parent 1 (Selfings P1), and 85  $S_1$  progenies of parent 2 (Selfings P2).

Genetics and Molecular Research 16 (1): gmr16019483

In crossings, the presence of TAM marker alleles mostly had a positive effect on plant performance, with only one SRAP marker-allele having a negative effect on seed yield. Due to the fact that plants of one selfing group generally possessed or lacked the respective markerallele, the effects of alleles from TAM detected in crossings could not be confirmed within selfings of the same parent.

Several TAMs showed associations with multiple, mainly vigor-related traits. For example, OVK\_183\_274 was associated with SN, Vigor, and HSu, while OVK\_133\_205, OVK\_111\_223 and OVK 155\_255 were associated with HSu and SN. Correlations among TAMs associated with vigor related traits (SN, Vigor, and HSu) showed two main clusters of marker alleles with strong correlations within and weak correlations outside the clusters (Figure 6). The TAMs me4em3\_356 (associated to SN) and me4em4\_127 (associated to PD-share) showed no clear affiliation to either of the two clusters.

Associated Trait						C	Correla	ated N	1arker						
HSu <sub>1,3</sub> SN <sub>13</sub> Vigor <sub>13</sub> PD-share	OVK_073_189	OVK_073_204	OVK_183_274	OVK_141_251	me1em1_217	me4em3_356	me4em4_525	me4em1_104	me4em3_127	OVK_133_205	OVK_111_223	OVK_155_255	OVK_063_191	me1em1_175	000 11
OVK_073_189		0.44	0.51	0.48	0.54	0.14	0.17	0.16	0.28	0.30	0.32	0.30	0.28	0.26	0
OVK_073_204	0.44		0.89	0.87	0.82	0.39	0.20	0.19	0.29	0.28	0.29	0.28	0.27	0.26	0
OVK_183_274	0.51	0.89		0.98	0.82	0.38	0.23	0.19	0.30	0.26	0.22	0.26	0.29	0.23	0
OVK_141_251	0.48	0.87	0.98		0.80	0.39	0.22	0.20	0.27	0.27	0.23	0.27	0.30	0.24	0
me1em1_217	0.54	0.82	0.82	0.80		0.35	0.16	0.10	0.23	0.30	0.31	0.30	0.29	0.23	0
me4em3_356	0.14	0.39	0.38	0.39	0.35		0.31	0.25	0.19	0.21	0.23	0.21	0.27	0.25	0
me4em4_525	0.17	0.20	0.23	0.22	0.16	0.31		0.37	0.49	0.29	0.26	0.29	0.29	0.23	0
me4em1_104	0.16	0.19	0.19	0.20	0.10	0.25	0.37		0.69	0.57	0.55	0.57	0.51	0.48	0
me4em3_127	0.28	0.29	0.30	0.27	0.23	0.19	0.49	0.69		0.65	0.62	0.65	0.63	0.55	0
OVK_133_205	0.30	0.28	0.26	0.27	0.30	0.21	0.29	0.57	0.65		0.97	1.00	0.92	0.83	0
OVK_111_223	0.32	0.29	0.22	0.23	0.31	0.23	0.26	0.55	0.62	0.97		0.97	0.89	0.86	0
OVK_155_255	0.30	0.28	0.26	0.27	0.30	0.21	0.29	0.57	0.65	1.00	0.97		0.92	0.83	0.
OVK_063_191	0.28	0.27	0.29	0.30	0.29	0.27	0.29	0.51	0.63	0.92	0.89	0.92		0.82	0.
me1em1_175	0.26	0.26	0.23	0.24	0.23	0.25	0.23	0.48	0.55	0.83	0.86	0.83	0.82		0.
me1em1 299	0.26	0.24	0.17	0.17	0.25	0.17	0.20	0.53	0.55	0.83	0.86	0.83	0.76	0.83	

**Figure 6.** Correlation coefficients among TAM alleles showing a significant association with at least one of the phenotypic traits plant height during summer (HSu), seed number (SN), plant vigor (Vigor), and share of prodelphinidins on total condensed tannins (PD-share). Black squares indicate significant associations between the TAM and one of the respective traits. Intensity of the green color reflects the value of the correlation coefficient.

#### DISCUSSION

The present study reports, for the first time, associations between genetic markers and phenotypic traits in sainfoin, representing a preliminary, but essential step towards MAS for this species. The information gained on TAMs, and on correlations among phenotypic traits, could help to increase efficiency in sainfoin breeding. For example, favorable marker alleles of TAM can be traced in potential parental plants from new varieties to increase their performance. Furthermore, the work effort required in the selection for complex traits such as seed yield might be reduced by selection for correlated traits such as plant vigor or height that can be more easily determined.

Genetics and Molecular Research 16 (1): gmr16019483

# Detected trait variation as a basis for selection

The presence of sufficient variation in heritable traits is indispensable for the selection of superior individuals. In crossings, for flowering time, the difference between the earliest and latest flowering plant was more than 6 weeks (Table 1). This variation is larger than in other forage legumes like red clover, where the difference between the earliest and latest materials of a breeding program ranges between 2 and 3 weeks (Boller B, unpublished results). Hence, this variation indicates the potential to create early and late flowering varieties, adapted to different climate regions. Differences in Vigor, Hsu (83 cm), LA (28.2 cm<sup>2</sup>), and LL (13 cm) were also substantial, indicating the potential to breed generally stronger plants with a higher chance of survival.

For concentrations of polyphenolic compounds, including total CT, large variation could be found among crossings (Table 1). This within-population variation for CT concentration (range = 37.5 mg/g) was higher compared to that observed in earlier studies that analyzed variation among different sainfoin varieties, where ranges of 24.5, 21.8, and 18.5 mg/g were observed (Azuhnwi et al., 2011, 2012; Malisch et al., 2015). The benefit of most of the polyphenolic compounds analyzed in the present study is their antioxidant capacity and anthelmintic effectiveness (Thill et al., 2012). Some individual compounds from these groups are considered more abundant and effective in their anthelmintic properties than others. For example, breeding for higher flavonol composition, i.e., quercetin and its glycoside rutin, might improve the anthelmintic capacity of the forage (Barrau et al., 2005). The range of quercetin concentrations found in our study was between 5.8 and 32.1 mg/g (Table 1), showing a very strong correlation (r = 0.99) to rutin concentration (data not shown), which is comparable to the range found by Malisch et al. (2015) in a panel of 27 diverse accessions (1.9 to 21.3 mg/g). This indicates the potential to increase the concentration of this active compound via selection in existing breeding pools or by introgression of more exotic material.

# TAM detection and validation

PCO analysis (Figure 1A) revealed the existence of population structure, which would not be expected in traditional bi-parental mapping populations, as was the case for our F1 progenies. This could be due to the fact that samples were coded for the presence and absence of SRAP and SSR marker alleles, resulting in two classes of marker phenotypes. Hence, the actual genotype of plants could be biased by the negligence of the allele dose (simplex, 1000; duplex, 1100; triplex, 1110; and quadruplex, 1111). Another possible explanation could be preferential chromosome pairing, a phenomenon observed in allohexaploid wheat and other species, which is caused by the presence of pairing between control loci (Sears, 1976). In the present study, using the kinship matrix as a variance-covariance matrix was a powerful tool to correct for such population structure effects when performing marker regression.

# **Detected TAMs and potential benefits**

The three agronomical traits for which significant TAMs were found (SN, HSu, Vigor) are all related to general plant vigor, are moderately-to-strongly correlated among each other, and share the same TAM (Table 2 and Figure 6), indicating a pleiotropic mode of action for the potential underlying gene loci. Increased seed yield is crucial for sainfoin, and its selection via

Genetics and Molecular Research 16 (1): gmr16019483

TAM would be highly effective. This is because, to date, seed yield can only be assessed after the year of establishment in the field, and is highly labor intensive. Furthermore, seed yield is strongly influenced by the environment (Nowosad et al., 2016), reducing trait heritability and, therefore, the response to selection. An efficient strategy to increase seed yield might be to combine MAS using identified TAMs with indirect selection via positively correlated traits that are simple to determine. In our study, plant height (r = 0.47, 0.61) and vigor (r =0.75) showed moderate-to-strong correlations with seed yield, i.e., more vigorous plants had a higher seed production potential, and could be used for the indirect selection of plants with increased seed yield. To a lower extent, indirect selection for higher seed yield is also possible by selecting genotypes with earlier flowering (r = -0.34).

For compositional traits, one significant TAM was identified for PD-share (Table 4). Our results show that it would not be possible to indirectly select for chemical composition via simple agronomic traits due to the weak correlations between these two groups of traits (Table 3). The single TAM found in our study that was associated with higher PD-share could, therefore, be useful to efficiently select plants with a higher PD concentration. This could be desirable, because PDs are thought to be the most beneficial compound of total CTs with the strongest antiparasitic activity (Mechineni et al., 2014). Due to the positive correlation between the two traits (r = 0.60), MAS for PD-share would also increase the total concentration of CT.

# Accuracy of TAM to infer the presence of QTLs in sainfoin populations

TAMs detected by simple marker regression are good indicators of the co-occurrence of a marker with QTL (Collard et al., 2005). However, the exact location of the marker within the genome, and the assignment of several TAMs to the same or a different QTL remains unknown without verification in a linkage map; to date, this is not available for sainfoin. However, patterns of correlations among markers or marker alleles might provide an indication of the assignment of TAMs to potential QTL. For example, in perennial ryegrass (*Lolium perenne*), highly correlated markers (r = 0.95) linked to seed yield were found to be associated with the same QTL (Studer et al., 2008). In our study, the two clusters of correlated TAMs suggested the existence of two different potential QTL regions for vigor-related traits, with five and eight associated TAMs, respectively (Figure 6). One TAM (me4em3\_356) associated with SN could not be assigned to either one of the two clusters and might, therefore, describe a third potential QTL region affecting seed yield (Figure 6). This was also the case for the TAM associated with PD-share (me4em4\_525), which likely belongs to a different, independent QTL.

The number and effect size of TAMs or potential QTLs, respectively, detected in our study was comparable to results reported for other forage species. For seed yield, three QTLs were detected in red clover and two in perennial ryegrass, explaining a maximum 15.3 and 32.8% phenotypic trait variation, respectively (Herrmann et al., 2006; Studer et al., 2008). This is comparable to 28.1% of the explained variance as observed in our study. For plant height, six QTLs were reported for perennial ryegrass, and five QTLs were reported for switchgrass (*Panicum virgatum*), explaining 27.9 and 17.4% of phenotypic variation, respectively, which is consistent with the 28.6% found in the present study (Studer et al., 2008; Serba et al., 2015). For plant vigor in alfalfa, four QTLs were detected, explaining up to 23.3% of the phenotypic variance, which is close to the 23.9% found in the present study (McCord et al., 2014).

The precision with which a QTL can be localized by a genetic marker locus is proportional to the number of sampled meiosis (each one providing the opportunity for recombination

Genetics and Molecular Research 16 (1): gmr16019483

between the marker and the trait locus) and, therefore, depends on the type of population and its size (Mackay, 2001). TAMs detected in our study are based on 122  $F_1$  progenies derived from crossing two non-inbred parental plants (parent 1, parent 2). Therefore, the number of meiotic divisions was limited to one per parental plant and progeny. A larger population size would have been desirable; however, due to the large number of self-fertilizations (Kempf et al., 2015), and the generally low seed set, this could not be accomplished. Hence, due to the limited number of crossing-over events, larger parts of chromosome might be linked (and thus, always inherited together). Therefore, distances of observed TAM to the true QTL might be large, and it is likely that the association might be lost in the next generation of plants, when crossing-over occurs between the QTL and observed TAM (Humphreys, 2005). For this reason and the allelic marker diversity between populations, the application of detected TAMs in other sainfoin populations might be challenging. However, they may still be useful in populations derived from crosses of one of the parents used in this study with other materials, e.g., for introgression of a given trait.

# Future trends for sainfoin breeding and conclusions

Generally, some of the disadvantages associated with sainfoin could be compensated for by cultivation practices, e.g., by reducing competition from weeds and choosing appropriate companion plants to increase persistence. However, optimized cultivation practices cannot fully compensate the missing adaptation capacity of available sainfoin cultivars to diverse environmental conditions. Genetic improvement may, therefore, not be bypassed in the long term.

In this study, traits relevant for plant survival and reproduction were found to be positively correlated, showing potential for their simultaneous improvement via breeding. Furthermore, a subset of polymorphic SSR and SRAP markers showed its usability for screening populations consisting of self- and cross-fertilized progenies. Thereby, the identification of TAMs based on linear regression methods was feasible for vigor traits, seed yield, and one chemical compound. Those TAMs could be incorporated into sainfoin breeding programs.

In future studies on sainfoin breeding, more information on trait inheritance must be provided. It will be necessary to apply our marker systems to several, larger sainfoin populations with different genetic backgrounds. A further important step will be the establishment of linkage maps to gain information about the location of QTLs on the chromosome. The possibility of inbreeding, which was recently confirmed (Kempf et al., 2015), could, thereby, enable the development of experimental  $F_2$  mapping populations or recombinant inbred lines. With a larger number of sampled meiotic divisions, linkage mapping in such populations, in combination with an increased number of markers and individuals, might allow a more precise determination of QTL positions than would be currently possible with our  $F_1$  population.

#### **Conflicts of interest**

The authors declare no conflict of interest.

# ACKNOWLEDGMENTS

Research supported by the European Commission through the Marie Curie Initial Training Network Legume Plus (#PITN-GA-2011-289377; http://legumeplus.eu/). The

Genetics and Molecular Research 16 (1): gmr16019483

authors thank Juha-Pekka Salminen and the natural chemistry research group (University of Turku, Finland) for providing the laboratory and assisting in data interpretation for the analysis of phenolic compounds. We thank Delley Seeds and Plants Ltd. for providing the field site and assisting in the field trial. Furthermore, we thank Sonja Reinhard, David Peditto, and Robert Spiess for assistance in marker analysis.

#### REFERENCES

- Aulchenko YS, Ripke S, Isaacs A and van Duijn CM (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23: 1294-1296. <u>http://dx.doi.org/10.1093/bioinformatics/btm108</u>
- Azuhnwi BN, Boller B, Martens M, Dohme-Meier F, et al. (2011). Morphology, tannin concentration and forage value of 15 swiss accessions of sainfoin (*Onobrychis viciifolia* Scop.) as influenced by harvest time and cultivation site. *Grass Forage Sci.* 66: 474-487. <u>http://dx.doi.org/10.1111/j.1365-2494.2011.00811.x</u>
- Azuhnwi BN, Thomann B, Arrigo Y, Boller B, et al. (2012). Ruminal dry matter and crude protein degradation kinetics of five sainfoin (*Onobrychis viciifolia* Scop) accessions differing in condensed tannin content and obtained from different harvests. *Anim. Feed Sci. Technol.* 177: 135-143. <u>http://dx.doi.org/10.1016/j.anifeedsci.2012.08.004</u>
- Barrau E, Fabre N, Fouraste I and Hoste H (2005). Effect of bioactive compounds from Sainfoin (Onobrychis viciifolia Scop.) on the *in vitro* larval migration of *Haemonchus contortus*: role of tannins and flavonol glycosides. Parasitology 131: 531-538. <u>http://dx.doi.org/10.1017/S0031182005008024</u>
- Barrett B, Baird I and Woodfield D (2009). White clover seed yield: a case study in marker-assisted selection. In: Molecular Breeding of Forage and Turf, (Yamada T and Spangenberg G, eds.), Springer Science Business Media, New York.
- Castelli F, Contillo R and Miceli F (1996). Non-destructive determination of leaf chlorophyll content in four crop species. *J. Agron. Crop Sci.* 177: 275-283. http://dx.doi.org/10.1111/j.1439-037X.1996.tb00246.x
- Chen W-M and Abecasis GR (2007). Family-based association tests for genomewide association scans. Am. J. Hum. Genet. 81: 913-926. http://dx.doi.org/10.1086/521580
- Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142: 169-196. <a href="http://dx.doi.org/10.1007/s10681-005-1681-5">http://dx.doi.org/10.1007/s10681-005-1681-5</a>
- Doyle CJ, Thomson DJ and Sheehy JE (1984). The future of sainfoin in British agriculture: an economic assessment. *Grass Forage Sci.* 39: 43-51. http://dx.doi.org/10.1111/j.1365-2494.1984.tb01663.x
- Eding H and Meuwissen THE (2001). Marker-based estimates of between and within population kinships for the conservation of genetic diversity. J. Anim. Breed. Genet. 118: 141-159. <u>http://dx.doi.org/10.1046/j.1439-0388.2001.00290.x</u>
- Engström MT, Pälijärvi M, Fryganas C, Grabber JH, et al. (2014). Rapid qualitative and quantitative analyses of proanthocyanidin oligomers and polymers by UPLC-MS/MS. J. Agric. Food Chem. 62: 3390-3399. <u>http://dx.doi.org/10.1021/jf500745y</u>
- Faville M, Barrett B, Griffiths A, Schreiber M, et al. (2003). Implementing molecular marker technology in forage improvement. Proc. New Zeal. Grassl. Assoc. 65: 229-238.
- Goplen BP, Richards KW and Moyer JR (1991). Sainfoin for western Canada. Agriculture Canada Publication 1470/E. Herrmann D, Boller B, Studer B, Widmer F, et al. (2006). QTL analysis of seed yield components in red clover (*Trifolium*

pratense L.). Theor. Appl. Genet. 112: 536-545. http://dx.doi.org/10.1007/s00122-005-0158-1

- Hill SN (1980). An Evaluation of factors influenced by nitrogen fixation in sainfoin (*Onobrychis viciifolia* Scop.). Montana State University, Bozeman.
- Hoste H, Martinez-Ortiz-De-Montellano C, Manolaraki F, Brunet S, et al. (2012). Direct and indirect effects of bioactive tannin-rich tropical and temperate legumes against nematode infections. *Vet. Parasitol.* 186: 18-27. <u>http://dx.doi.org/10.1016/j.vetpar.2011.11.042</u>
- Humphreys MO (2000). Breeding methods for forage and amenity grasses. In: Molecular breeding of forage crops. Proceedings of the 2nd International Symposium, Molecular Breeding of Forage Crops, Hamilton and Lorne (Spangenberg G, eds.).
- Humphreys MO (2005). Genetic improvement of forage crops past, present and future. J. Agric. Sci. 143: 441-448. <u>http://</u> dx.doi.org/10.1017/S0021859605005599
- Kempf K, Grieder C, Walter A, Widmer F, et al. (2015). Evidence and consequences of self-fertilisation in the predominantly outbreeding forage legume Onobrychis viciifolia. BMC Genet. 16: 117. <u>http://dx.doi.org/10.1186/s12863-015-0275-z</u>
- Kempf K, Mora-Ortiz M, Smith LMJ, Kölliker R, et al. (2016). Characterization of novel SSR markers in diverse sainfoin (*Onobrychis viciifolia*) germplasm. *BMC Genet.* 17: 124. <u>http://dx.doi.org/10.1186/s12863-016-0431-0</u>

Genetics and Molecular Research 16 (1): gmr16019483

- Lee M, Godshalk EB, Lamkey KR and Woodman WW (1989). Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. *Cell Biol. Mol. Genet.* 29: 1067-1071.
- Li G and Quiros CF (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. *Theor. Appl. Genet.* 103: 455-461. <u>http://dx.doi.org/10.1007/s001220100570</u>
- Mackay TFC (2001). The genetic architecture of quantitative traits. Annu. Rev. Genet. 35: 303-339. <u>http://dx.doi.org/10.1146/annurev.genet.35.102401.090633</u>
- Malisch CS, Lüscher A, Baert N, Engström MT, et al. (2015). Large variability of proanthocyanidin content and composition in sainfoin (*Onobrychis viciifolia*). J. Agric. Food Chem. 63: 10234-10242. <u>http://dx.doi.org/10.1021/</u> acs.jafc.5b04946
- McCord P, Gordon V, Saha G, Hellinga J, et al. (2014). Detection of QTL for forage yield, lodging resistance and spring vigor traits in alfalfa (*Medicago sativa* L.). *Euphytica* 200: 269-279. <u>http://dx.doi.org/10.1007/s10681-014-1160-y</u>
- McMahon LR, McAllister TA, Berg BP, Majak W, et al. (2000). A review of the effects of forage condensed tannins on ruminal fermentation and bloat in grazing cattle. Can. J. Plant Sci. 80: 469-485. <u>http://dx.doi.org/10.4141/P99-050</u>
- Mechineni A, Kommuru DS, Gujja S, Mosjidis JA, et al. (2014). Effect of fall-grazed sericea lespedeza (Lespedeza cuneata) on gastrointestinal nematode infections of growing goats. Vet. Parasitol. 204: 221-228. <u>http://dx.doi.org/10.1016/j.vetpar.2014.06.002</u>
- Mora-Ortiz M, Swain MT, Vickers MJ, Hegarty MJ, et al. (2016). De-novo transcriptome assembly for gene identification, analysis, annotation, and molecular marker discovery in Onobrychis viciifolia. BMC Genomics 17: 756. <u>http://dx.doi.org/10.1186/s12864-016-3083-6</u>
- Nowosad K, Liersch A, Poplawska W and Bocianowski J (2016). Genotype by environment interaction for seed yield in rapeseed (*Brassica napus* L.) using additive main effects and multiplicative interaction model. *Euphytica* 208: 187-194. <u>http://dx.doi.org/10.1007/s10681-015-1620-z</u>
- Posselt UK (2010). Breeding Methods in Cross-Pollinated Species. In: Fodder Crops and Amenity Grasses, (Boller B, Posselt UK and Veronesi F, eds.). Springer Verlag, New York.
- R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Software. Available [http://www.R-project.org]. Accessed October 2, 2016.
- Robins JG, Bauchan GR and Brummer EC (2007). Genetic mapping forage yield, plant height, and regrowth at multiple harvests in tetraploid alfalfa (*Medicago sativa* L.). Crop Sci. 47: 11-18. http://dx.doi.org/10.2135/cropsci2006.07.0447
- Sears ER (1976). Genetic control of chromosome pairing in wheat. Annu. Rev. Genet. 10: 31-51. <u>http://dx.doi.org/10.1146/</u> annurev.ge.10.120176.000335
- Serba DD, Daverdin G, Bouton JH, Devos KM, et al. (2015). Quantitative trait loci (QTL) underlying biomass yield and plant height in switchgrass. *Bioenerg. Res.* 8: 307-324. <u>http://dx.doi.org/10.1007/s12155-014-9523-8</u>
- Studer B, Jensen LB, Hentrup S, Brazauskas G, et al. (2008). Genetic characterisation of seed yield and fertility traits in perennial ryegrass (*Lolium perenne L.*). *Theor. Appl. Genet.* 117: 781-791. http://dx.doi.org/10.1007/s00122-008-0819-y
- Thill J, Regos I, Farag MA, Ahmad AF, et al. (2012). Polyphenol metabolism provides a screening tool for beneficial effects of *Onobrychis viciifolia* (sainfoin). *Phytochemistry* 82: 67-80. <u>http://dx.doi.org/10.1016/j.phytochem.2012.05.030</u>
- Wright S (1978). Variability within and among natural populations. The University of Chicago Press, Chicago.

# **Supplementary Material**

Figure S1. Frequency distribution of performance in agronomic traits and concentrations of phenolic compounds in 30 S<sub>1</sub> progenies derived from parent 1.

Figure S2. Frequency distribution of performance in agronomic traits and concentrations of phenolic compounds in 85 S1 progenies derived from parent 2.

Table S1. Forward and reverse primers for SSR (Kempf et al., 2016) and SRAP (Li and Quiros, 2001) markers used to test for marker trait associations.

Table S2. Pearson's correlation coefficients among phenotypic traits assessed in offspring derived from the selffertilization of parent 1 (P1). Agronomic traits assessed were vigor (Vigor), plant height in spring (HSp) and summer (HSu), flowering time (FT), seed number per plant (SN), sensitivity to rust (Rust), SPAD, leaf area (LA), and leaf length (LL), with the number in subscript representing the year of measurement (2012, 2013). Quality traits were the concentration of condensed tannins (CT), prodelphinidin share (PD share), as well as concentration of

Genetics and Molecular Research 16 (1): gmr16019483

quercetin, kaempferol, myricetin, quinic acid, and gallic acid. Numbers highlighted in green and red indicate that the corresponding correlation coefficients in offspring derived from cross-fertilization between parent 1 (P1) and parent 2 (P2) were at least 0.30 higher or lower, respectively.

Table S3. Pearson's correlation coefficients among phenotypic traits assessed in offspring derived from selffertilization of parent 2 (P2). Agronomic traits assessed are vigor (Vigor), plant height in spring (HSp) and summer (HSu), flowering time (FT), seed number per plant (SN), sensitivity to rust (Rust), SPAD, leaf area (LA), and leaf length (LL), with the number in subscript representing the year of measurement (2012, 2013). Quality traits are concentration of condensed tannings (CT) and prodelphinidin share (PD-share), as well as concentration of quercetin, kaempferol, myricetin, quinic acid, and gallic acid. Numbers highlighted in green and red indicate that the corresponding correlation coefficients in offspring derived from cross-fertilization between parent 1 (P1) and parent 2 (P2) were at least 0.30 higher or lower, respectively.

Genetics and Molecular Research 16 (1): gmr16019483