T. Albrecht, B. Büttner, S.B. Carey, R. Seidenberger, A. Lutz, A. Harkess and S. Gresset

Independent validation of molecular markers for sex determination on diverse sex chromosomes in hops (*Humulus lupulus* L.)

***Humulus lupulus* L. is a perennial and dioecious crop important for the beer and food industry. Sex determination in early stages of the breeding cycle is important to secure the selection of female candidates. Phenotypic selection for sex is time consuming and greenhouse capacities are limited generating the demand for an alternative method for early sex screening. Therefore, the need to develop genotypic markers for sex determination is high. Sex-linked markers developed in biparental populations were not applicable across diverse germplasm due to incomplete linkage. Recent sequence analyses have increased the knowledge about the structure and diversity of the sex chromosomes and newly developed markers are promising for the application in breeding programs. To identify sex-linked markers for the German hop breeding program, we analysed a diversity panel comprising 190 international hop varieties, wild hops and breeding material. SNP (single nucleotide polymorphism) markers were assessed with genotyping by sequencing and a genome-wide association study was performed based on the phenotypic sex expression. Significant SNP markers from the association study were analysed in two validation sets comprising selection candidates. Furthermore, recently published sex-linked markers were analysed in the diversity panel and in both validation sets. Based on both analyses, we were able to validate two SNP markers ("Scaffold_1533_203371763" and SM1) located in the sex determining region of the X and Y chromosomes. Both markers together were able to correctly assign sex to the selection candidates in both validation sets. Furthermore, genotypic clusters derived from the assays of marker "Scaffold_1533_203371763" separated female, monoecious and male plants enabling a more targeted allocation of breeding resources in terms of field and greenhouse space as well as phenotyping and genotyping capacities. This advancement will result in an increased selection intensity and thus progress in developing more resilient hop varieties for a sustainable brewing industry.**

Descriptors: *Humulus lupulus*, sex chromosomes, genome-wide association study, SNP markers, hop breeding

1 Introduction

Humulus lupulus L., common hop, is a perennial crop with its components being important for the beer, pharmaceutical and food industry. Hop is dioecious with a diploid (2n = 18 + XX, XY), large (2.8 Gb) and highly complex genome [1, 2]. Female plants develop cone-like structures called strobili. They consist of yellowish-green overlapping bracts attached to a central axis. Mature strobili harbor a significant amount of glandular trichomes, the so called lupulin glands, containing most of the hop specific compounds essential for brewing, pharmaceutical and industrial applications [3]. Male hop plants produce no strobili and thus are not commercially cultivated. As wind based pollination of female flowers results in the formation of seeds that potentially destabilise beer aroma [4], male hop plants are usually removed around main hop production regions. In the main hop production region of Germany, the Hallertau, removal of wild hop plants is prescribed by law to prevent the spread of pollen [5]. Therefore, determining sex as early as possible in the hop breeding process is important not only for an efficient utilization of breeding resources but also for commercial cultivation purposes.

In common hop, sex is determined by two sex chromosomes with XX generally resulting in female and XY resulting in male hop plants, respectively [6]. Occasionally, monoecious hop plants are reported in breeding gardens and commercial fields and are considered to be a consequence of abiotic stress and chromosomal number disorders (e.g. triploids or aneuploids) [7–9]. The X and Y chromosomes evolved from a pair of autosomes and one of their key characteristics is a sex determination region (SDR), where the Y does not recombine with the X, causing the sequence within it to diverge. In contrast, the so-called pseudo-autosomal region (PAR) of the sex chromosomes, exhibits similar characteristics as autosomal chromosomes, and recombines during male meiosis [10,

https://doi.org/10.23763/BrSc24-14albrecht

Authors

Theresa Albrecht, Bianca Büttner, Rebecca Seidenberger, Bavarian State Research Center for Agriculture, Institute for Crop Science and Plant Breeding, Biotechnology for Plant Breeding, Freising, Germany; Sarah B. Carey (ORCID ID: 0000-0002-6431-0660), Alex Harkess (ORCID ID: 0000-0002-2035-0871), HudsonAlpha, Institute for Biotechnology, Huntsville, USA; Anton Lutz, Sebastian Gresset, Bavarian State Research Center for Agriculture, Institute for Crop Science and Plant Breeding, Hop Breeding, Wolnzach, Germany; corresponding author: Sebastian.Gresset@lfl.bayern.de

11]. Recent analyses have shown that the SDR encompasses the majority of the Y chromosome sequence at almost 80 % [12–14].

Due to the lack of reliable sex-linked genetic markers, breeding programs globally rely on phenotypic evaluation, which is usually only possible from the second year of growth [8]. For several years, to increase breeding efficiency, the hop breeding programme at the Bavarian State Research Centre for Agriculture utilizes a greenhouse specifically designed for hop cultivation to promote hop growth in early spring enabling flower formation and thus sex determination within the first year after germination. The increase in breeding efforts due to the negative impacts of climate change on hop cultivation and the need for better adapted and more sustainable hop varieties led to an increase in breeding efforts [15]. Thus, greenhouse capacities being exceeded for phenotypic sex determination have generated demand for an alternative method for early sex screening.

Developing genotypic markers to improve the breeding process, has been a goal in hop breeding since first marker platforms have been developed. However, analysing markers is cost intensive and time consuming and most markers identified in biparental populations are not transferable to more distantly related breeding material. In recent years, genome-based genotyping platforms lead to decreased genotyping costs and better genome coverage even for niche crops like hop. Further advances were made with the availability of whole genome sequences from hop cultivars "Shinsu Wase", "Saazer", "Teamaker" and "Cascade" [16–18]. Recent developments using long sequence reads lead to a first physical map of the hop genome representing the ten chromosomes including the X chromosome [2]. Using this information, the first markers for sex determination identifying the X and Y chromosomes have been published and were recently summarized by Clare et al. [9]. Another advancement in the development of sex-linked markers is the recent generation of chromosome-scale genome references for males of three botanical varieties of hops, *lupulus*, *lupuloides*, and *neomexicanus* [13], which provide a better understanding of the XY pairs at a sequence level. The size of the Y chromosome varies across *Humulus* varieties, ranging from ~ 169 Mb in var. *lupulus* to ~ 270 Mb in varieties *lupuloides* and *neomexicanus* [14], consistent with observed cytological variation [7]. Importantly, the location of the boundary of the PAR with the SDR also varies across varieties [14], which complicates the development of sex-linked markers that work universally across hops if they land within this variable region.

To our knowledge, none of the markers have been developed or validated in current German breeding pools of hop. Although first validations in other breeding pools have shown the usefulness of the recently developed markers [9], validations in German breeding material and the reliable identification of monoecious phenotypes were missing. The validation of these markers is of interest for our own research in marker-assisted selection to improve the breeding process which is currently solely based on phenotypic selection. Therefore, the aims of this study were (1) to develop markers for sex determination in early stages of the breeding process using a large diversity panel comprising German and international varieties and breeding lines, and (2) validate the application of our own markers as well as published markers in independent validation sets.

2 Material and Methods

2.1 Genetic material

The diversity panel for identification of sex-linked genetic markers comprised 190 internationally derived hop genotypes of which 7 are European landraces and 85 registered varieties from Germany (31), UK (14), US (12), Czech Republic (8), Slovenia (5), Poland (4), Japan (3), South Africa (3), China (3), and Ukraine (2). Those varieties were complemented by 89 genotypes from the German breeding program, two German wild hops, five wild hops from Asia, and two genotypes from the UK breeding program.

Validation set 1 includes 25 genotypes of which 7 were overlapping with the diversity panel but with independent DNA sampling from different plants. Validation set 2 contains 82 genotypes of which 26 were overlapping either with the diversity panel or validation set 1 but with independent DNA sampling (Supplement 1).

2.2 Phenotyping

Phenotypic assessment of the sex for the diversity panel was conducted in the field on a plot basis with three plants per plot between 2017 and 2019 at two locations in Bavaria, Germany. Sex was recorded as female or male when most of the developed flowers in one plot were female or male, respectively.

Phenotypic assessment of the sex for the two validation sets was conducted in 2023 and 2024 on single plants of each genotype in two locations in Bavaria, Germany. Phenotyping was conducted as described previously with the difference that genotypes exhibiting male and female flowers on one plant were recorded as monoecious.

2.3 Genotyping

In the diversity panel, leaf samples of single plants were sent to LGC Genomics GmbH, Berlin, for DNA extraction and genotyping. For GBS (genotyping by sequencing), the genome was digested with restriction enzyme ApeKI for genome complexity reduction. A normalization for reduction of repetitive regions was performed and 1.5 million reads per sample produced on average. GBS analysis was performed using the Illumina NextSeq 500/illumina NovaSeq (2 x 150 bp) for de novo SNP detection. GBS and alignment based on the publicly available "Cascade" genome ("dovetailCascadeFullAssemblyMasked" provided by https://hopbase.org/ [2]) resulted in 368,688 single nucleotide polymorphism (SNP) markers across all samples. SNP markers were filtered by LGC for read counts above 16 resulting in 152,326 SNP markers of which 151,067 were biallelic. One cultivar was removed from the sample due to overall low genotyping quality. SNP markers were filtered according to minor allele frequency above 1 % and missing values below 10 % in TASSEL 5.2.93 resulting in 2494 high quality SNPs [19]. Missing data in the final data set was imputed with the LinkImpute algorithm developed for non-model organisms implemented in TASSEL 5.2.93 [20]. SNP genotypes were recoded according to the reference allele with 0 being homozygous for the reference allele, 1 being heterozygous and 2 being homozygous for the alternative allele.

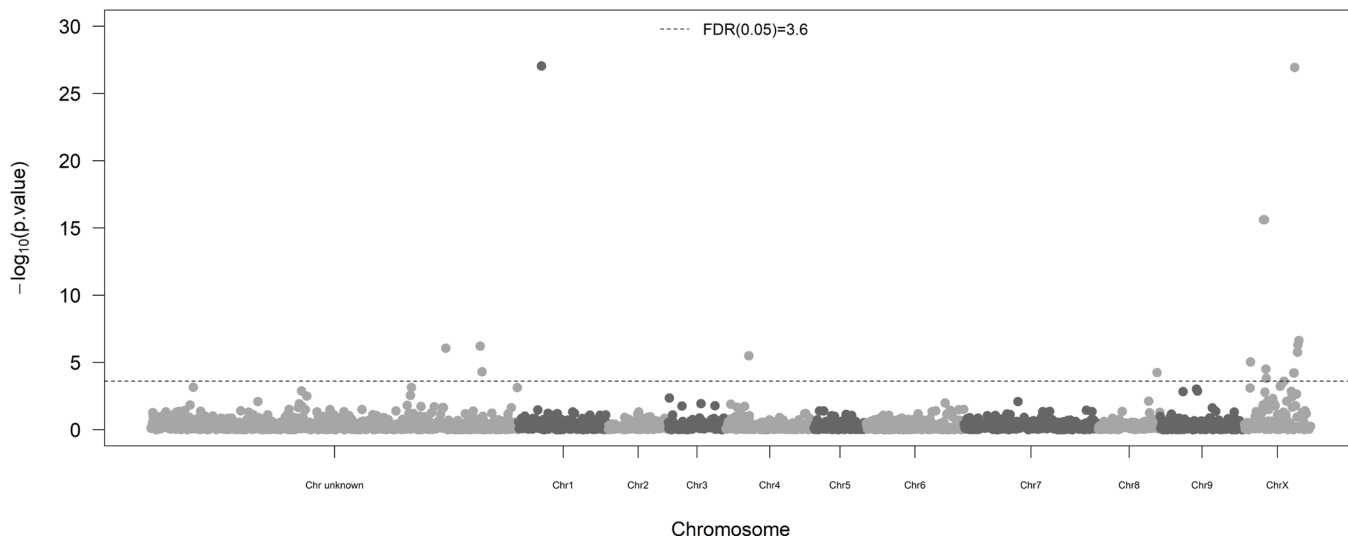

Fig. 1 **Manhattan plot of marker LOD scores (-$\log_{10}$(*p* value)) from the association mapping. SNP positions are indicated on the x-axis according to the Cascade full genome assembly. The LOD score is on the y-axis and the dashed line indicates the significance threshold according to the false discovery rate (FDR with α = 0.05) estimated on the *p* values of all 2492 markers**

2.4 Marker identification

Two methods for detecting sex-linked markers for the diversity panel were applied to the data set. First, genome-wide association mapping based on a mixed linear model with modelling the kinship between the hop cultivars as covariance was calculated in rTASSEL to identify significant associations between single SNP markers and phenotypic sex expression of the genotyped hops [21, 22]. Here, the identifier for the sex was coded as 1 for the male varieties and 0 for females. Second, we assumed that markers located within the SDR and which are hypothesized to be unique for the Y chromosome will not assemble on the X chromosome and therefore, will produce missing values in female lines but not in male lines ("Y-linked marker approach"). Here, the raw data (151,067 SNPs) was used without prefiltering for missing values.

2.5 Validation

In addition to our in-house research study, four studies with published sex-linked markers were chosen for analysis [8, 9, 23, 24]. In both validation sets, DNA was isolated with innuPREP Plant DNA Kit (iST-Innuscreen) according to the manufacturer's instruction.

In two of these studies, where the corresponding GBS SNPs were available [9, 24], SNP genotypes were first validated and selected within the diversity panel. For overall 10 SNP markers (9 identified SNP markers based on the GWAS and Y-linked marker approach and one primer set for a published SNP marker, SM1 [9].), KASP (Kompetitive allele specific polymerase chain reaction (PCR)) primers were designed by 3CR Bioscience based on publicly available sequences from Cascade. The primers were each tested in validation set 1. The KASP runs were performed with the PACE™ Genotyping Master Mix (3CR Bioscience) on a CFX Touch RealTime PCR Detection System (Bio-Rad) according to the manufacturer. Of these markers, three correctly predicted sex in validation set 1 and were then tested in validation set 2.

Parallel to the KASP markers, the published simple sequence repeat (SSR) marker HlAGA7 [8] and a multiplex PCR based on four DArT

Table 1 **Significant markers above false discovery rate (FDR) correction threshold obtained from association mapping based on 190 diverse hop genotypes using quality filtered GBS data and phenotypically observed sex**

Marker	Chromosome	Position (Mb)[a]	*P* value	LOD[b] score	R^{2c}
Scaffold_1531_168733693	1	168.734	9.03E-28	27.0	95 %
Scaffold_1533_390472306	X	390.472	1.14E-27	26.9	89 %
Scaffold_1533_203371763	X	203.372	2.43E-16	15.6	47 %
Scaffold_1533_414225557	X	414.226	2.31E-07	6.6	18 %
Scaffold_1533_414108608	X	414.109	5.03E-07	6.3	17 %
Scaffold_27_465238	unknown	0.465	6.03E-07	6.2	14 %
Scaffold_246_179307	unknown	0.179	8.78E-07	6.1	14 %
Scaffold_1533_410557854	X	410.558	1.72E-06	5.8	15 %
Scaffold_24_93252202	4	93.252	3.15E-06	5.5	12 %
Scaffold_1533_35415045	X	35.415	9.16E-06	5.0	13 %
Scaffold_1533_207262950	X	207.263	3.06E-05	4.5	12 %
Scaffold_144_537929	unknown	0.538	4.91E-05	4.3	9 %
Scaffold_49_243168600	8	243.169	5.57E-05	4.3	11 %
Scaffold_1533_386707347	X	386.707	6.06E-05	4.2	11 %
Scaffold_1533_214114887	X	214.115	0.00015	3.8	10 %
Scaffold_1533_343616467	X	343.616	0.00025	3.6	7 %

[a] Physical position of the SNP marker in mega bases

[b] Logarithm of odds

[c] Proportion of phenotypic variation explained by the marker

(Diversity Arrays Technology) markers which are assumed to be linked to sex chromosomes in hop were tested in validation set 1 [23]. Here, the PCRs were performed according to the publications [7, 8] except for using the GoTaq® G2 Hot Start Polymerase (Promega) in both PCRs. The fragments of the SSR marker were separated on a 5 % polyacrylamide gel (bisacrylamide (19:1) based on urea) by 50 watt for 1,5 hours and the fragments of the multiplex PCR on a 2 % agarose gel by 140 volts for around 1 hour.

2.6 Genomic coordinates of the sex-linked markers

To gain a better understanding of the genomic coordinates of the sex-linked markers, we used BLAST version 3.0.2 on the newly released X and Y chromosomes from across *Humulus lupulus* varieties [13, 25]. We plotted these results using the R package karyoploteR version 1.26.0 (R version 4.3.1) [26].

3 Results

3.1 Phenotypic sex determination

Phenotyping of the diversity panel across three years resulted overall in 40 predominantly male and 150 female phenotypes. Of the 89 genotypes randomly sampled from the Huell breeding program, 39 were male and 50 were female. All genotypes representing registered varieties were phenotyped as being female. Within validation set 1, the ratio between males and females was 12 : 11. Two additional genotypes showing predominantly male flowers but with a few female flowers on the end of the side branches were recorded as monoecious phenotypes. The proportion of monoecious phenotypes was even higher in validation set 2 with 14 genotypes developing mainly male flowers but with a few female flowers which even developed to cones. Additionally, 54 genotypes were phenotyped as being female and 14 as being male.

3.2 Association mapping and markers identified using GBS data

Using the mixed linear model, 16 significant SNP markers above the FDR correction thresholds were identified (Fig. 1, Table 1). Most of these SNP markers (ten) were located on the X chromosome, for three SNP markers with significant association to the phenotypically observed sex no chromosome assignment was available. Three SNP markers statistically associated with sex expression were detected on chromosomes one, four and eight, based on the alignment against the publicly available “Cascade” genome. The marker with the highest LOD (Logarithm of odds: -log10(p)) score of 27 was identified on chromosome 1 at 169 Megabases (Mb). Two significant markers on the X chromosome had a LOD score above 15 and were identified in two regions at 203 Mb and 390 Mb. All other significant markers had LOD scores below 7. For developing KASP primers, we focused on the three SNP markers with the highest LOD scores “Scaffold_1531_168733693”, “Scaffold_1533_390472306” and “Scaffold_1533_203371763”.

Comparison of missing values for the SNP markers in male and female varieties identified 13 markers with more than 90 % missing values in female lines (y-axis) but less than 30 % of missing values

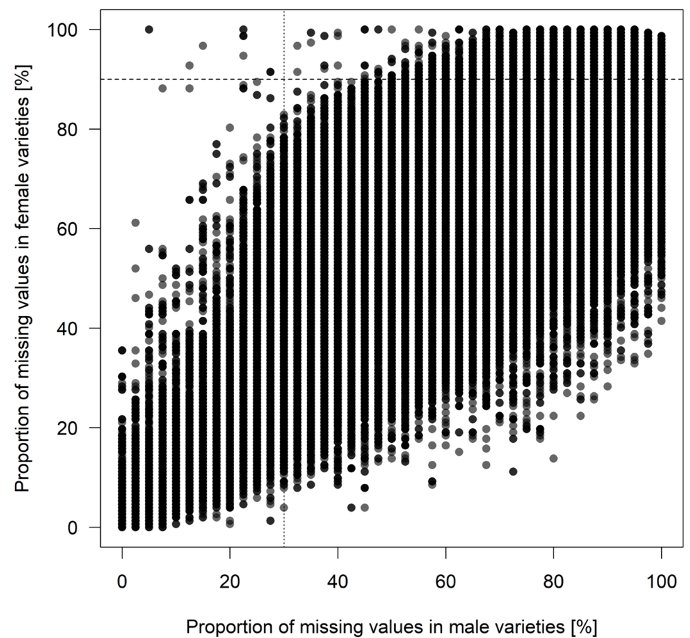

Fig. 2 Proportion of missing values of markers in 150 female (y-axis) and 40 male hop varieties (x-axis) based on 151,067 biallelic SNP markers. SNP markers are depicted as black diamonds with 60 % transparency to indicate overlaps. Dashed lines mark 90 % value for missing values in female varieties and 30 % value for missing values in male varieties, respectively

in male lines (x-axis, Fig. 2). In contrast, no markers were identified with a high proportion of missing values in males but low rates of missing values in females. According to the alignment against the most recent Cascade full genome assembly, these markers were located on chromosomes 1, 2, 3, 9, and X. Six of these markers were unique and had different allele classes clearly separating male and female plants in the diversity panel.

3.3 Validation of sex-linked markers

Overall, using genome wide association studies (GWAS) and the Y-linked marker approach, nine SNP markers were identified and the corresponding KASP primers were developed to validate their linkage to the observed sex. In figure 3 (see page 176), results from testing these nine SNP markers together with three markers previously published are illustrated. Five of these markers were able to predict the phenotypically observed pattern for sex across validation set 1. Three markers were able to detect different alleles in validation set 1, but the patterns were not in accordance with the phenotypically observed sex. Four markers were not able to detect different alleles for female and male plants. The SNP marker “Scaffold_1531_168733693” which had the highest LOD score in the GWAS, was not able to detect different allele classes to predict sex in validation set 1. SNP markers derived from the Y-linked approach utilizing missing data of females in the diversity panel predicted on average 17 % of the gender correctly which is lower compared to a random assignment of the sex.

Four marker systems predicted the phenotypic sex with a precision above 80 %. Two of these SNP markers were originally identified based on the GWAS analysis in the diversity panel and two were previously published SSR marker systems. The two SNP markers

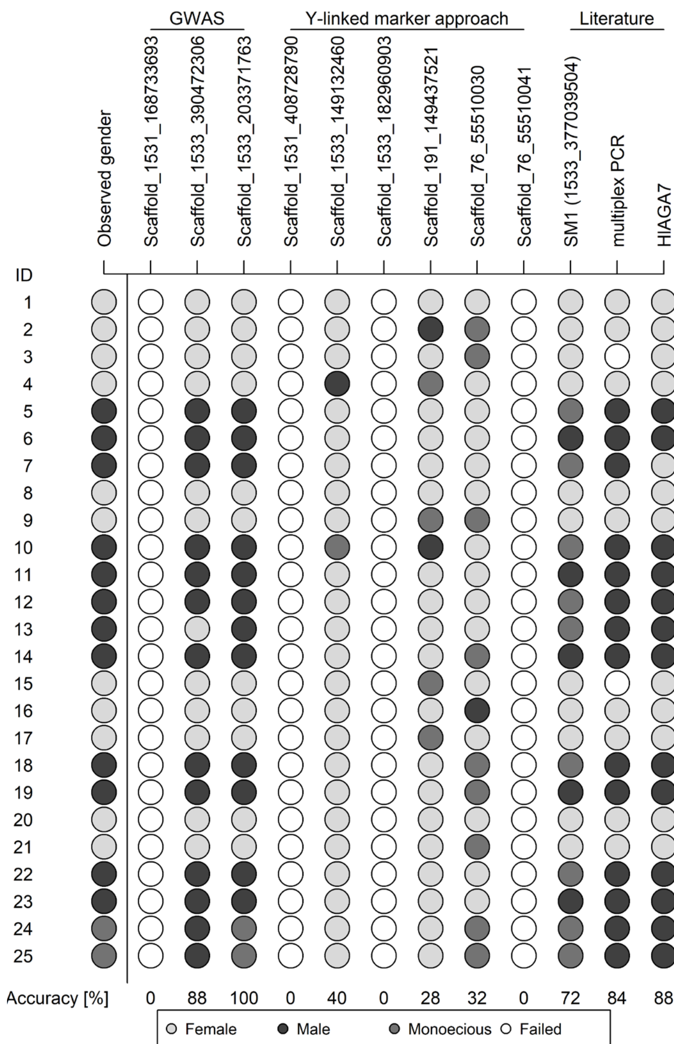

Fig. 3 **Consistency of phenotypically observed and genotypically predicted sex expression within validation set 1 for 12 tested molecular markers (three SNP markers identified with GWAS, six markers with Y-linked marker approach in the diversity panel and three markers from literature [8, 9, 23]). Percentages of overall accuracy for each marker is stated below**

"Scaffold_1533_390472306" and "Scaffold_1533_203371763" from the GWAS analysis identify male, female and monoecious breeding lines of validation set 1 with 88 % and 100 % accuracy, respectively (Fig. 3). These two markers were further validated in validation set 2, where both markers were able to predict 72 % and 96 % of the observed sex correctly (Supplement Fig. S1). In both validation sets, marker "Scaffold_1533_203371763" outperformed other markers with three distinct allele clusters, males as heterozygous and females as homozygous while all monoecious breeding lines appeared in the cluster between males and females (Fig. 4, Supplement Figs. S2 and S3).

The SNP markers mentioned in Havill et al. (FM_010, FM_003, FM_014 [24]) and Clare et al. (SM1, 1533_377039504 [9]) were checked first in the diversity panel. Only two markers FM_010 and FM_003 were available in our GBS data. Both were unable to distinguish male and female plants within the diversity panel and thus were not further analysed in the validation sets.

Two markers ("Scaffold_1533_377796553" and "Scaffold_1533_377039414") located close to FM_014 and SM1 were available in our diversity panel. Here, only the marker in proximity of SM1 was able to identify 138 out of 189 plants correctly with 51 genotypes of the diversity panel resulting in missing data for this SNP marker. Therefore, KASP primers for SM1 according to Clare et al. were applied in validation set 1, where all 11 female plants were correctly identified (Fig. 3), but the marker was not able to correctly separate male and monoecious plants (Supplement Fig. S3) [9]. In validation set 2, this marker correctly identified 76 out of 86 plants (Supplement Fig. S1).

In the multiplex PCR [23], the control band is present in all reactions, which indicates that the PCR has technically worked (see Supplement Fig. S4). There is a band pattern above and below the control band in line with the phenotypic sex expression for males and females, only the monoecious phenotypes cannot be correctly addressed. The SSR marker HlAGA7 was analysed in validation set 1 and was able to identify female plants, but it was not able to separate male and monoecious phenotypes (Supplement Fig. S5). The SSR and the multiplex marker were not pursued further because, on the one hand, the two markers cannot distinguish between the monecious and the male plants, on the other hand, they are much more complex in evaluating compared to the KASP markers. While the KASP marker works with fluorescence labelled primers and the PCR run can therefore be evaluated automatically, a gel must be made for the other two markers, which has to be visually evaluated.

Primer sequences of the developed markers were blasted against the sex chromosomes of the different hop botanical varieties *Humulus lupulus* var. *lupulus*, var. *lupuloides* and var. *neomexicanus* and the resulting physical positions are illustrated in figure 5. Here, the markers SM1, "Scaffold_1533_390472306" and HlAGA7 were located close to or even in the PAR region while SNP marker "Scaffold_1533_203371763" was clearly located in the SDR region of the X and Y chromosomes across all considered botanical varieties.

4 Discussion

Developing molecular markers for the early identification of male and female hop genotypes was a focus in hop breeding research for more than two decades with several marker systems being proposed [8, 9, 23, 27–30]. Broad application of the reported molecular markers in hop breeding were hampered due to low linkage of the proposed markers with the phenotypic sex expression across diverse germplasm. Technical properties of the proposed marker systems inhibited high throughput or automatization. As field, greenhouse and human capacities for phenotypic sex determination are limited, paralleled with an increased demand for breeding progress in terms of adaptation to rapidly changing production environments, developing sex-linked molecular markers suited for the German hop breeding program was one approach for increasing efficiency. Thereby, developed molecular markers must be codominant and tightly linked to the phenotypic sex expression at least across *Humulus lupulus* var. *lupulus*, *H. l.* var. *lupuloides* and *H. l.* var. *neomexicanus*, which are frequently used in germplasm development. Additionally, developed markers

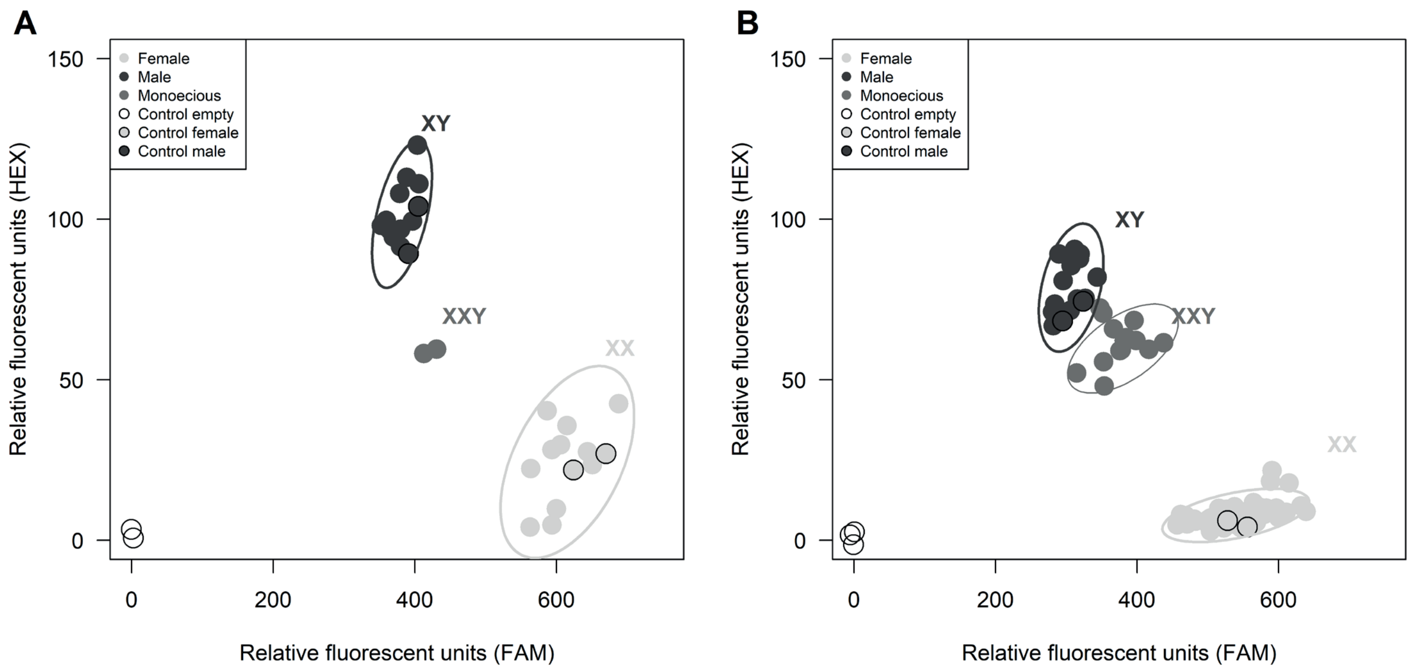

Fig. 4 **Scatterplot of KASP assay for SNP marker "Scaffold_1533_203371763" in (A) validation set 1 and (B) in validation set 2. Fluorescence from the FAM fluorophore is on the x-axis and HEX fluorophore is on the y-axis. Points are coloured according to the observed sex, while the ellipses indicate the predicted sex from allele clusters**

must be easily scalable to enable the handling of several thousand samples within a few days.

Identification of sex-linked molecular markers is always dependant on the utilized genotypes. Until recently, the validity of results was limited to the germplasm used for identification or empirical validation. With the high quality assembly of X and Y chromosomes for the botanical varieties *H. l.* var. *lupulus*, *lupuloides* and *neomexicanus* and the identification of the boundaries of the sex determining region (SDR) on the sex chromosomes [11, 13], we were able to develop a pair of codominant sex-linked SNP markers located within the SDR of hop sex chromosomes enabling the broad application of the developed markers across diverse germplasm.

For identification of SNP markers linked to the sex within the employed diversity panel, we followed two approaches. First, we identified genomic regions associated with sex expression using a GWAS mapping approach resulting in overall 16 markers above the FDR threshold. Three of these resulted in a LOD score above 15 and thus were significantly associated with the phenotypic sex expression based on the binary phenotypic sex records from the field trials with the diversity panel. Those three SNPs were transferred to the KASP marker system and evaluated for their predictive ability within validation set 1. The SNP marker "Scaffold_1533_203371763" predicted 100 % of the respective sexes within validation set 1 correctly with distinct groups of females, males and monoecious phenotypes. Within validation set 2, SNP marker "Scaffold_1533_203371763" predicted the sex for 79 out of 82 genotypes in accordance with the respective phenotypes. In one case, the KASP PCR failed to amplify, and no fluorescence signals could be detected while in two other cases the monoecious plants were located between the two clusters separating male and monoecious plants but could be clearly separated from the cluster of female plants (Fig. 4B). Analysis of the position of this SNP on the X and Y chromosomes placed SNP "Scaffold_1533_203371763" directly within the region of sex determination in all three botanical varieties of *Humulus lupulus* frequently used in germplasm development (Fig. 5). As the diversity panel and both validation sets comprised genotypes derived from different breeding programs including different botanical varieties of *Humulus lupulus* in

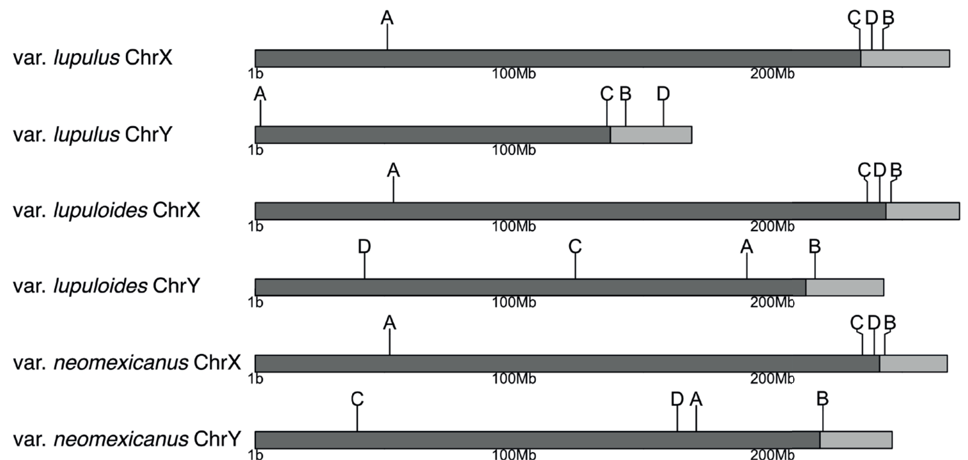

Fig. 5 **Physical coordinates of the four markers A: Scaffold_1533_203371763, B: Scaffold_1533_390472306, C:SM1, and D: HlAGA7 in X and Y chromosomes of the *Humulus lupulus* var. *lupulus*, var. *lupuloides* and var. *neomexicanus*. The dark grey area represents the sex determination (SDR) and the light grey area the pseudoautosomal region (PAR) on the sex chromosomes**

their ancestry, supported by the suppressed recombination rate within the SDR, we are highly confident, that SNP marker "Scaffold_1533_203371763" displays a highly predictive and reliable molecular marker for hop sex expression across a wide range of diverse germplasm. Additionally, this SNP also proved to be able to predict monoecious phenotypes within both validation sets. Monoecious phenotypes were only recorded for the validation sets on plants exhibiting predominantly male flowers with few female flowers on the end of the side branches. Those kinds of monoecious plants were previously described as being triploid with two X and one Y chromosome [7, 31, 32]. Monoecious hop plants with predominantly female and few male flowers were not observed within the two years of phenotyping of the two validation sets although they are occasionally observed in production fields and breeding gardens. Those kinds of monoecious plants are likely genetically diploid (XX) and initiated by an environmentally disturbed hormone balance during the shift from vegetative to reproductive growth [7,32]. Although the predictability of the SNP marker "Scaffold_1533_203371763" for sex expression in monoecious phenotypes with predominantly female flowers could not be tested within the current study, it seems unlikely that such phenotypes can be predicted by this SNP marker due to the anticipated large influence of environmental effects on the expression of male flowers on female genotypes.

Besides the highly functional SNP marker "Scaffold_1533_203371763", GWAS identified one additional SNP on the X chromosome based on the Cascade full genome assembly. KASP markers derived from this SNP marker ("Scaffold_1533_390472306") predicted 22 of 25 phenotypes within the validation set 1 and 59 of 82 in validation set 2 correctly. The imperfect linkage of this SNP marker with phenotypic sex expression is supported by the alignment to the sequence of the X and Y chromosomes of the three botanical varieties of *Humulus lupulus*. In all three botanical varieties SNP marker "Scaffold_1533_390472306" is outside of the non-recombining SDR but within the PAR (Fig. 4) [33].

SNP marker "Scaffold_1531_168733693" resulted in the overall highest significant association with phenotypic sex expression based on the GWAS analysis explaining almost 95 % of the phenotypic variance. In contrast to the results from the GWAS where all female phenotypes were linked to the T-allele at this SNP, the derived KASP markers failed to amplify different allele classes within validation set 1. Although this marker was found to be physically located on chromosome 1 based on the latest Cascade full genome assembly [2] as well as on the recently published full genome assembly of USDA 21110M [13], this marker was genetically linked to SNP marker "Scaffold_1533_390472306" located on the sex chromosomes in the diversity panel (linkage disequilibrium above 0.9, data not shown). Considering the significant associations of SNPs assigned to scaffold 1531 and phenotypic sex expression already reported by Havill et al. and the significant association of SNP marker "Scaffold_1531_168733693" to the phenotypic observed sex in the diversity panel, a sequence duplication of scaffold 1531 might be located in proximity to SNP marker "Scaffold_1533_390472306" on the sex chromosomes [24]. Alternatively, this SNP marker on chromosome 1 might be tightly linked to a sex determining region on the autosomal chromosome as previously mentioned for hops by McAdam et al. and discussed as contributing to sexual plasticity in Canabis sativa [34]. A third option might be that chromosome 1 harbours a genomic region under strong co-selection with female sex expression [30]. However, as no distinct allele clusters were observed in validation set 2 for SNP marker "Scaffold_1531_168733693" and the linked SNP marker "Scaffold_1533_390472306" is in the PAR of the sex chromosomes, our focus was on the remaining markers which were more promising in predicting sex based on the results of the validation panels.

For our second approach to detect sex-linked markers, we aimed to detect sex specific sequences as proposed by [35] by assuming that GBS reads based SNPs located on the Y chromosome will result either with "unknown" position or misplacing based on the alignment with the full genome assembly of Cascade. We detected 13 SNPs within our diversity panel which were characterized through low failure rate in phenotypical males and disproportionally high failure rate in females. Of those, six SNP markers were further validated but none proved to be suited for predicting sex within validation set 1 although two of these markers were in the scaffold associated with the X chromosome. Several reasons for the failure of this approach might be possible. On the one hand, a high number of SNP markers was observed with high rates of missing values and due to the low number of male varieties in the diversity panel, the SNP markers might have been selected only by chance but not due to linkage to the Y chromosome. On the other hand, we were able to identify SNP markers located on sex chromosomes but the chances to identify SNP markers within the SDR region might be low due to the applied GBS approach given that the SDR is assumed to carry a high number of repetitive elements with low methylation [13, 36–38]. Most SNP markers detected with the GWAS were also located in the PAR of the sex chromosomes and not in SDR and therefore are not tightly linked to phenotypic sex expression. Generally, both approaches were successful in identifying the sex chromosomes but with the second approach a lower number of markers was detected and the chances to detect markers in the SDR were higher with the GWAS method.

Another object of our study was to validate recently published markers for sex determination in our breeding material. Four studies were chosen using different marker technologies like SSR (HlAGA7), DArT (multiplex PCR) and GBS SNP markers [8, 9, 23, 24]. In a recent study, the marker HlAGA7 did not predict the sex 100 % correctly [39], and while we identified the marker to be within the SDRs of the *lupuloides* and *neomexicanus* references, in var. *lupulus* we instead found the marker is within the PAR (Fig. 5). In our validation set 1 HlAGA7 was able to correctly identify female plants and the same was observed for the multiplex PCR (with the exception of the monoecious phenotypes). The close proximity of the marker to the SDR in var. *lupulus* could explain the marker's efficacy in some materials, or the SDR boundary may be variable within var. *lupulus* but based on these results the chromosomal location of HlAGA7 explains why it may not be ideal for routine application. Therefore, our focus was on developing KASP assays based on our own and published GBS SNP markers. Unfortunately, the SNP markers published in Havill et al. were not linked to the phenotypically observed sex in our breeding germplasm [24]. Only, the KASP marker developed based on SM1 published by Clare et al was successfully validated in our material [9]. However, the SM1 marker was able to predict 85 % of the sex correctly which is less

compared to the results by Clare et al. [9]. Our results indicate that the SM1 marker is close to the border of the SDR and PAR region on the sex chromosomes in some botanical varieties which might be the reason for the observed recombination in our material.

5 Conclusion

Based on the two complementary approaches for identifying molecular markers for early prediction of hop gender, we were able to identify one SNP marker "Scaffold_1533_203371763" which is located in the SDR of the sex chromosomes and was consistently able to separate female and male plants as well as monoecious from male plants.

The results from our study show that markers must still be evaluated for specific germplasm before utilization in breeding programs. The SNP marker developed in our breeding material performed best in predicting phenotypic sex expression across the highly diverse German breeding germplasm. Nevertheless, we propose to apply a combination of SNP marker "Scaffold_1533_203371763" and SM1 to minimize the chance of failures in the assignment of sex to selection candidates and to back up the genotypic prediction for selection within males and females. Applying both proposed markers proved to be highly reliable for predicting phenotypic sex expression enabling a selection for sex within a few weeks after germination. Thus, sex specific selection intensities can already be applied in early selection steps, e.g. during greenhouse evaluation of downy mildew (Pseudoperonospora humuli) tolerance, with higher selection pressure on male offspring. Additionally, applying the proposed SNP marker set, male hop genotypes can be completely excluded from field trials envisaged for evaluation of quality and yield potential increasing trial homogeneity, thereby repeatability of phenotyping and reducing field space and labour demand. Overall, the marker set proposed for genotypic prediction of sex within the frequently utilized botanical varieties of *Humulus lupulus* L. enables a more targeted allocation of resources ultimately resulting in more selection gain and more sustainable future hop varieties for the brewing sector and beer enthusiasts.

Data availability

Sequences of primers used in this study can be requested through the corresponding author as well as Supplement 1.

Acknowledgement

The authors thank Jutta Kneidl, Sonja Ostermeier, Maximilian Schleibinger, Anton Hartung, Daniel Ismann, Petra Hager, Brigitte Haugg, Brigitte Forster, Andreas Roßmeier, Brigitte Brummer and Katja Merkl for their excellent technical assistant. We express our gratitude to apl. Prof. Dr. Volker Mohler and Dr. Christina Krönauer for critical discussions and guidance in interpreting results.

This research was supported in part by the Wissenschaftliche Station für Brauerei in München e.V. and the German Society of Hop Research (GfH). Additional support was provided by the United States Department of Agriculture National Institute of Food and Agriculture Postdoctoral Fellowship no. 2022-67012-38987 (SBC).

Conflict of interest

The authors declare there are no conflicts of interest.

6 References

1. Edwardson, J.R.: Hops – Their botany, history, production and utilization, Economic Botany, **6** (1952), no. 2, pp. 160-175.
2. Padgitt-Cobb, L. K.; Pitra, N. J.; Matthews, P. D.; Henning, J. A. and Hendrix, D. A.: An improved assembly of the "Cascade" hop (*Humulus lupulus*) genome uncovers signatures of molecular evolution and refines time of divergence estimates for the Cannabaceae family, Horticulture research, **10** (2023), no. 2, uhac281.
3. Biendl, M.; Engelhard, B.; Forster, A.; Gahr, A.; Lutz, A.; Mitter, W. et al.: Hopfen: Vom Anbau bis zum Bier, Fachverlag Hans Carl, 2016.
4. Kavanagh, T. E.; Greenaway, R. W.; Hildebrand, R. P. and Clarke, B. J.: Trace volatile constituents of beer. Hop seed oil., Journal of the Institute of Brewing, **81** (1975), no. 4, pp. 314-317.
5. Freistaat Bayern: Wildhopfenverordnung (BayWHopfV) in der in der Bayerischen Rechtssammlung (BayRS 7823-6-L) veröffentlichten bereinigten Fassung, die zuletzt durch § 5 der Verordnung vom 5. Dezember 2017 (GVBl. S. 589) geändert worden ist: BayWHopfV, 1956.
6. Neve R.A.: Sex determination in the cultivated hop *Humulus lupulus*., PhD thesis, University of London, UK, 1961.
7. Shephard, H. L.; Parker, J. S.; Darby, P. and Ainswoth, C. C.: Sexual development and sex chromosomes in hop, New Phytologist, **148** (2000), no. 3, pp. 397-411.
8. Jakse, J.; Stajner, N.; Kozjak, P.; Cerenak, A. and Javornik, B.: Trinucleotide microsatellite repeat is tightly linked to male sex in hop (*Humulus lupulus* L.), Molecular Breeding, **21** (2008), no. 2, pp. 139-148.
9. Clare, S. J.; King, R. M.; Tawril, A. L.; Havill, J. S.; Muehlbauer, G. J.; Carey, S. B. et al.: An affordable and convenient diagnostic marker to identify male and female hop plants, G3 Genes | Genomes | Genetics, **14** (2024), no. 1, jkad216.
10. Divashuk, M. G.; Alexandrov, O. S.; Kroupin, P. Y. and Karlov, G. I.: Molecular cytogenetic mapping of *Humulus lupulus* sex chromosomes, Cytogenetic and genome research, **134** (2011), no. 3, pp. 213-219.
11. Hill, S. T.; Coggins, J.; Liston, A.; Hendrix, D. and Henning, J. A.: Genomics of the hop pseudo-autosomal regions, Euphytica, **209** (2016), no. 1, pp. 171-179.
12. Razumova, O. V.; Divashuk, M. G.; Alexandrov, O. S. and Karlov, G. I.: GISH painting of the Y chromosomes suggests advanced phases of sex chromosome evolution in three dioecious Cannabaceae species (*Humulus lupulus*, *H. japonicus*, and *Cannabis sativa*), Protoplasma, **260** (2023), no. 1, pp. 249-256.
13. Carey S. B., Havill J. S., Bentz P. C., Akozbek L. M., Lovell J. T., Osmanski A., Padgitt-Cobb L., Lynch R., Allsing N., Easterling K., Orozco L., Hale H., Mueller H., Meharg Z., McKay J., Guerrero R., Vergara D., Kane N., Michael T. P., Muehlbauer G. J., Harkess A.: The evolution of the heteromorphic sex chromosome in plants., in prep.
14. Padgitt-Cobb, L. K.; Kothen-Hill, S.; Henning, J. and Hendrix, D. A.: The long-read genome assembly of hop (*Humulus lupulus*) uncovers the pseudoautosomal region and other genomic features, Acta Horticulturae, International Society for Horticultural Science (ISHS), Leuven, Belgium, 2021, pp. 1-16.

15. Seigner, E.; Lutz, A.; Oberhollenzer, K.; Seidenberger, R.; Seefelder, S. and Felsenstein, F.: Breeding of hop varieties for the future, Acta Horticulturae, International Society for Horticultural Science (ISHS), Leuven, Belgium, 2009, pp. 49-58.
16. Natsume S.; Takagi H.; Shiraishi A.; Murata J.; Toy onaga H.; Patzak J. et al.: The Draft Genome of Hop (*Humulus lupulus*), an Essence for Brewing, Plant and Cell Physiology, **56** (2015), no. 3, pp. 428-441.
17. Hill, S.; Sudarsanam, R.; Henning, J. and Hendrix, D.: HopBase: a unified resource for *Humulus* genomics, Database: The Journal of Biological Databases and Curation, **2017** (2017), no. 1.
18. Padgitt-Cobb LK.; Kingan SB.; Wells J.; Elser J.; Kronmiller B.; Moore D. et al.: A phased, diploid assembly of the Cascade hop (*Humulus lupulus*) genome reveals patterns of selection and haplotype variation, bioRxiv, 2019.
19. Bradbury, P. J.; Zhang, Z.; Kroon, D. E.; Casstevens, T. M.; Ramdoss, Y. and Buckler, E. S.: TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics, 23 (2007), no. 19, pp. 2633-2635.
20. Money, D.; Gardner, K.; Migicovsky, Z.; Schwaninger, H.; Zhong, G.-Y. and Myles, S.: LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms, G3 Genes|Genomes|Genetics, **5** (2015), no. 11, pp. 2383-2390.
21. Zhang, Z.; Ersoz, E.; Lai, C.-Q.; Todhunter, R. J.; Tiwari, H. K.; Gore, M. A. et al.: Mixed linear model approach adapted for genome-wide association studies, Nature Genetics, **42** (2010), no. 4, pp. 355-360.
22. Monier, B., Casstevens, T.M., Bradbury, P.J. and Buckler E.S.: rTASSEL: An R interface to TASSEL for analyzing genomic diversity., Journal of Open Source Software, **7** (2022), no. 76.
23. Čerenak, A.; Kolenc, Z.; Sehur, P.; Whittock, S. P.; Koutoulis, A.; Beatson, R. et al.: New Male Specific Markers for Hop and Application in Breeding Program, Scientific Reports, **9** (2019), no. 1, p. 14223.
24. Havill, J. S.; Richardson, B. J.; Rohwer, C. L.; Gent, D. H.; Henning, J. A. and Muehlbauer, G. J.: Identification of quantitative trait loci associated with R1-mediated resistance to powdery mildew and sex determination in hop (*Humulus lupulus* L.), Theoretical and Applied Genetics, **136** (2023), no. 7, p. 154.
25. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K. et al.: BLAST+: architecture and applications, BMC Bioinformatics, **10** (2009), no. 1, p. 421.
26. Gel, B. and Serra, E.: karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data, Bioinformatics, **33** (2017), no. 19, pp. 3088-3090.
27. Polley, A.; Ganal, M. W. and Seigner, E.: Identification of sex in hop (*Humulus lupulus*) using molecular markers, Genome, **40** (1997), no. 3, pp. 357-361.
28. Seefelder, S.; Ehrmaier, H.; Schweizer, G. and Seigner, E.: Male and female genetic linkage map of hops, *Humulus lupulus*, Plant Breeding, **119** (2000), no. 3, pp. 249-255.
29. Buck, E. J.; Wiedow, C.; Carlisle, C.; Chagne, D. and Beatson, R.: Validation of new sex specific DNA markers for hops (*Humulus Lupulus* L.), Acta Horticulturae, International Society for Horticultural Science (ISHS), Leuven, Belgium, 2009, pp. 323-328.
30. McAdam, E. L.; Freeman, J. S.; Whittock, S. P.; Buck, E. J.; Jakse, J.; Cerenak, A. et al.: Quantitative trait loci in hop (*Humulus lupulus*L.) reveal complex genetic architecture underlying variation in sex, yield and cone chemistry, BMC Genomics, **14** (2013), no. 1, p. 360.
31. Haunold, A.: Cytology, Sex Expression, and Growth of a Tetraploid X Diploid Cross in Hop (*Humulus lupulus* L.), Crop Science, **11** (1971), no. 6, pp. 868-871.
32. Skof, S.; Cerenak, A.; Jakse, J.; Bohanec, B. and Javornik, B.: Ploidy and sex expression in monoecious hop (*Humulus lupulus*), Botany, **99** (2012), no. 3, pp. 355-361.
33. Carey, S.; Yu, Q. and Harkess, A.: The Diversity of Plant Sex Chromosomes Highlighted through Advances in Genome Sequencing, Genes, **12** (2021), no. 3.
34. Monthony, A. S.; Ronne, M. de and Torkamaneh, D.: Exploring ethylene-related Genes in Cannabis sativa: Implications for sexual plasticity, bioRxiv, 2023.
35. Palmer, D. H.; Rogers, T. F.; Dean, R. and Wright, A. E.: How to identify sex chromosomes and their turnover, Molecular ecology, **28** (2019), no. 21, pp. 4709-4724.
36. Vyskot, B.; Araya, A.; Veuskens, J.; Negrutiu, I. and Mouras, A.: DNA methylation of sex chromosomes in a dioecious plant, Melandrium album, Molecular and General Genetics MGG, **239** (1993), no. 1, pp. 219-224.
37. Fraser, J. A. and Heitman, J.: Chromosomal sex-determining regions in animals, plants and fungi, Current Opinion in Genetics & Development, **15** (2005), no. 6, pp. 645-651.
38. Kejnovsky, E.; Hobza, R.; Cermak, T.; Kubat, Z. and Vyskot, B.: The role of repetitive DNA in structure and evolution of sex chromosomes in plants, Heredity, **102** (2009), no. 6, pp. 533-541.
39. Driskill, M.; Pardee, K.; Hummer, K. E.; Zurn, J. D.; Amundsen, K.; Wiles, A. et al.: Two fingerprinting sets for *Humulus lupulus* based on KASP and microsatellite markers, PLOS ONE, **17** (2022), no. 4, e0257746.

Received 24 October 2024, accepted 6 December 2024

Supplementary Information

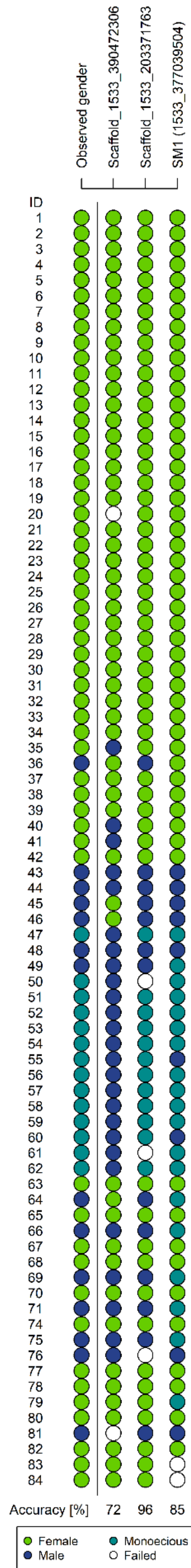

Fig. S1 **Gender of single plants in validation set 2 as observed in the field trials or estimated based on three SNP-markers. Female plants are coloured in green, monoecious plants are coloured in turquoise and male plants are coloured in blue. Overall accuracy for each marker is stated below**

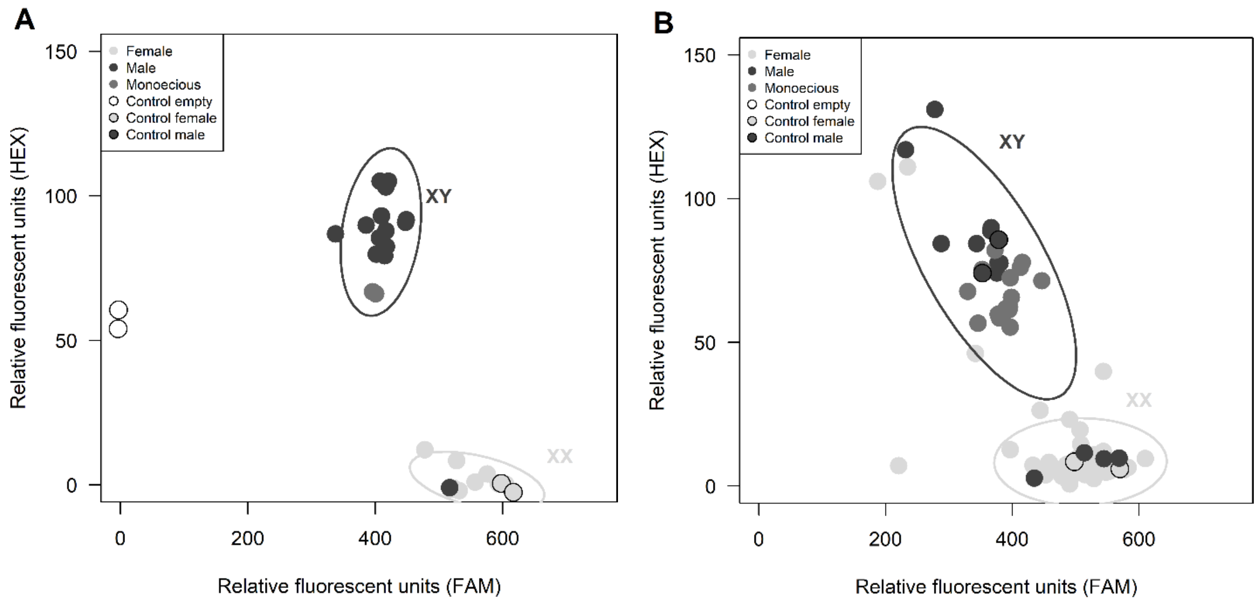

Fig. S2 Scatterplot of KASP assay for SNP-marker "Scaffold_1533_390472306" in (A) validation set 1 and (B) in validation set 2. Fluorescence from the FAM fluorophore is on the x-axis and HEX fluorophore is on the y-axis. Points are coloured according to the observed gender, while the ellipses indicate the predicted gender from allele clusters

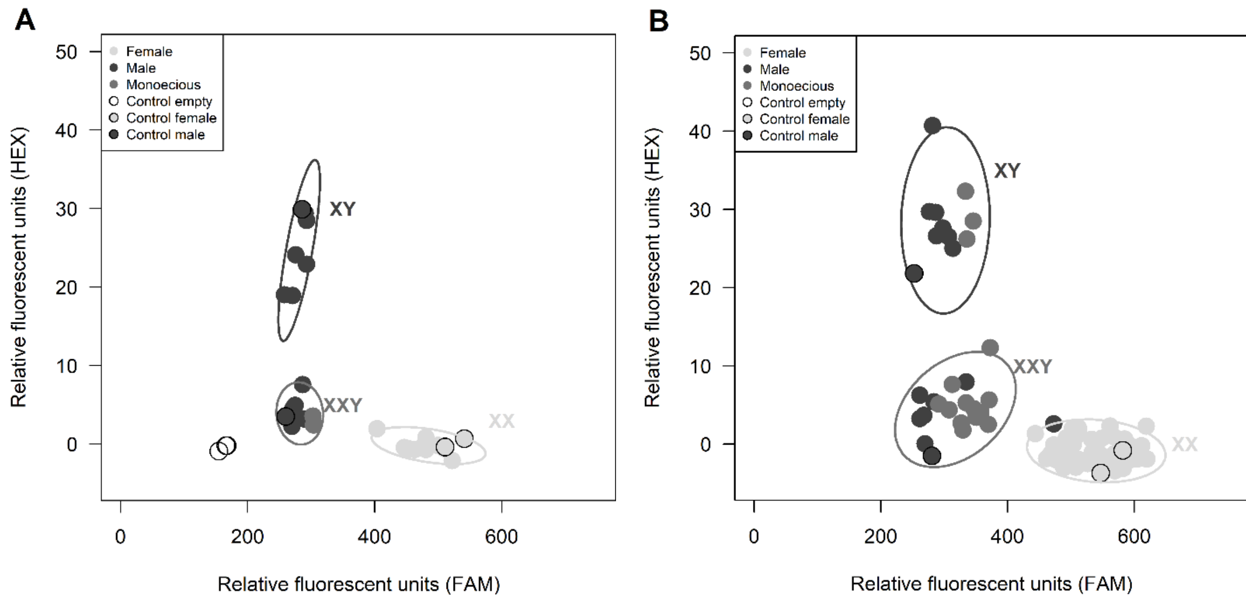

Fig. S3 Scatterplot of KASP assay for SNP-marker SM1 (1533_377039504) (A) validation set 1 and (B) in validation set 2. Fluorescence from the FAM fluorophore is on the x-axis and HEX fluorophore is on the y-axis. Points are coloured according to the observed gender, while the ellipses indicate the predicted gender from allele clusters

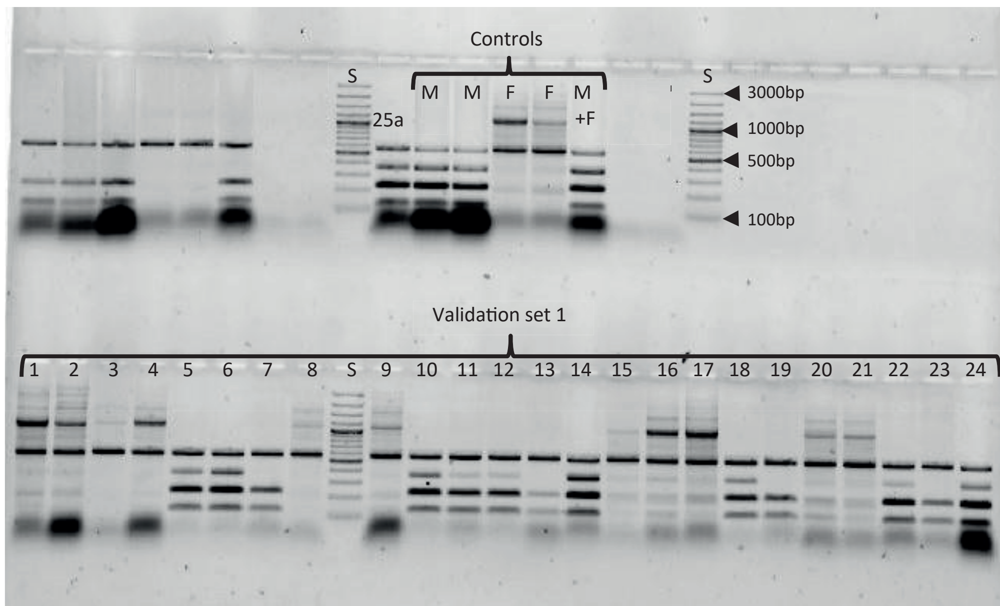

Fig. S4 **Fragments of the multiplex PCR on a polyacrylamide gel according to Čerenak et al 2019. Controls for male (M) and female (F) plants are in the first row and the ladder for size standards is indicated with S**

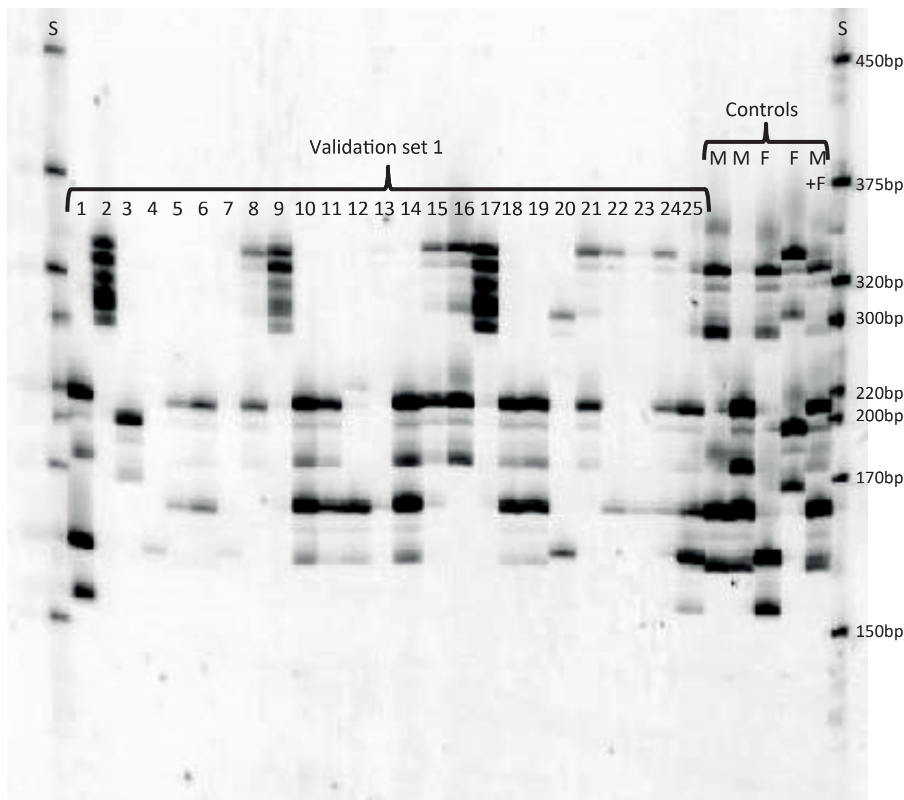

Fig. S5 **Fragments of the SSR marker HlAGA7 on a polyacrylamide gel according to Jakše et al. 2008. Controls for male (M) and female (F) plants are on the right and the ladder for size standards is indicated with S**