M. H. Hagemann, U. Born, G. Weber and J. N. Wünsche

Gene structure and allelic variations of hop (*Humulus lupulus* L.) bitter acid biosynthesis genes

Hop breeding to date is time-consuming and costly, thus the development of molecular markers that assist the knowledge-driven selection of genotypes from breeding populations is needed. To meet this requirement, modern technologies such as targeted genotyping-by-sequencing are now readily available. Consequently, to initiate the development of markers useful in hop breeding through targeted sequencing approaches, key genes involved in the bitter acid biosynthesis were structurally analyzed. The genes valerophenone synthase *(VPS)* and the two prenyltransferases 1 *(PT1)* and 2 *(PT2)* all had two copies, presumably allelic variations in the genome of the cultivar 'Cascade'. Gene structure analysis revealed two exons in *VPS* and 10 exons in both *PT1* and 2 *PT2* with large introns leading to a total length of up to 9 kb and 31 kb, respectively. The long PT2 gene size is due to transposable elements of the Gypsy and Copia family. Real-time quantitative PCR further suggests that, among the three bitter acid genes, the *VPS* had the statistically most significant association with the genotype typical bitter acid concentration. Moreover, identifying transcription factor binding site (TFBS) modules in all genes reveal promising sequence regions with known regulatory functions of the secondary plant metabolisms. It will be further evaluated whether those TFBS-modules are promising targets for future targeted sequencing approaches.

Descriptors: humulone, real time quantitative PCR, gene structure, promoter analysis, transcription factor binding site, single nucleotide polymorphisms

1 Introduction

The global hop market is in constant movement and transformation and currently driven by the rising demand of craft brewers for flavour hops. However, the introduction of new and highly productive cultivars, global warming and newly emerging pests and diseases are also concerning issues, which have to be addressed in modern hop breeding programs. Since the plant breeding process is very time consuming [1], new technologies such as marker-assisted selection based on genotyping-by-sequencing (GBS) are developed in order to support and enhance the selection of genotypes with preferred traits from breeding populations [2–4]. Consequently, the current study is aimed at initiating the development of molecular markers for the most important hop metabolites, the bitter acids (BA), which produce the typical bitter taste of most beer styles. However, marker development with non-targeted GBS might not be the only promising option in breeding programs of horticultural crops such as for hop, because GBS approaches tend to be extensive and costly. Typically, non-targeted GBS is based on amplifying and sequencing random fragments of a genome, consequently capturing genes or gene regulatory elements occurs only coincidentally rather than intentionally. Furthermore, most GBS approaches produce large amounts of sequence data, which require costly IT-infrastructure. Although sequencing techniques such as restriction site–associated DNA sequencing (RAD-sequencing), which are aimed at reducing the complexity of the genome are more practical for GBS, they are still based on random sequencing approaches [5] (Fig. 1A, see page 208). Considering the 2.6 Gb size of the hop genome and considering the high proportion of repetitive sequences [6], it is likely that "junk DNA" rather than important trait-associated genes and gene regulatory elements are sequenced.

Therefore, targeted sequencing of functional genes with „Genotyping-in-Thousands by sequencing" [7] or amplicon sequencing [8] (Fig. 1B, see page 208) are cost-efficient complementation's to GBS strategies for marker development. Targeted sequencing can be based on the PCR amplification of functional regions as genes or gene regulatory promoter regions [9]. Gene promoters contain elements for gene regulation and consequently there is a high probability that this regulatory elements are directly trait-associated sequences and consequently they are of interest in targeted sequencing approaches [9]. Several studies showed the importance of regulatory elements for gene activity. For ex-

https://doi.org/10.23763/BrSc19-25hagemann

Authors

Michael Helmut Hagemann, Ute Born, Jens Norbert Wünsche, University of Hohenheim, Crop physiology of Specialty Crops, Stuttgart, Germany; Gerd Weber, University of Hohenheim, Institute of Plant Breeding, Seed Science and Population Genetics, Stuttgart, Germany; corresponding author: Michael@Uni-Hohenheim.de

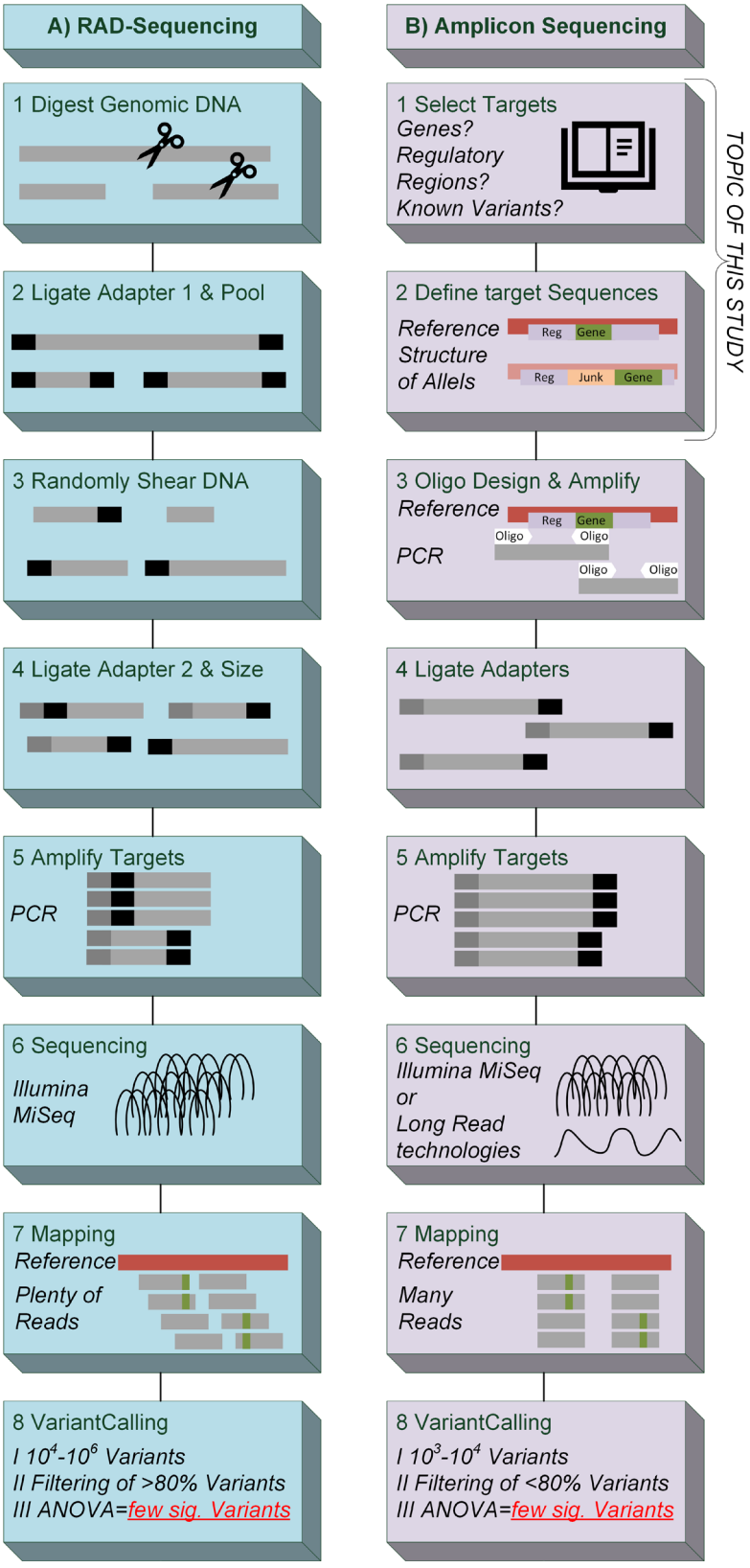

Fig. 1 **Comparison of two sequencing approaches (A) Restriction site–associated DNA sequencing (RAD-sequencing) based on Illumina MiSeq and (B) Amplicon Sequencing based on either Illumina MiSeq or Long Read technology sequencing. While RAD-sequencing is based on mostly randomly distributed restriction enzyme cut sites and on random shearing amplicon sequencing, amplicon sequencing is a knowledge driven sequencing approach, which produces less data but potentially more relevant sequence data. The focus of the current study is on how to select targets for a target sequencing approach aimed at the development of molecular markers for the trait bitter acid**

ample, the tale to tale motif, which was first functionally studied in potato, is necessary for certain transcription factors to bind and thereby enhancing the transcription of the related gene such as the isopentenyl transferase [10]. Specifically, the deleting a short part of the tale to tale motif of the promoter of the StBEL5 transcription factor resulted in reduced gene expression [10]. Such studies demonstrate that even small changes within the regularity region of a promoter can have an impact on gene activity and downstream processes. Generally, a putative promoter sequence can be defined as the region about 2.000 bp upstream of the first exon of a given gene [10, 11]. Consequently, an analysis of the gene structure is needed to reliably identify the position of exon 1 and thereby the position of the corresponding promoter. About 75 % of plant genes have less than 7 exons and an exon and intron length of less than 500 bp, respectively, thus most genes are typically only a few kilobases long [12]. However, few genes are more than 10 kb long as known for example for maize or some gymnosperms [13].

Structural analysis of complex hop genes in the genomes of the cultivars ‘Shinzu Wase’ and ‘Teamaker’ was not possible due to the high number of short scaffolds (e.g. 132.476 scaffolds ranging between 1 kb and 300 kb for ‘Shinzu Wase’) with long assembly gaps and thus the incomplete alignment of transcripts to genomic sequences. However, the ‘Cascade’ genome released in February 2019 consists of 11.705 contigs, ranging between 18 kb to 8.000 kb, allowing a more detailed structure analysis of particularly more complex genes compared to the previously published draft genomes [6, 14]. A characterisation of regions within a putative promoter with regulatory function is then based on transcription factor binding sites (TFBS) and higher order TFBS-module identification by using nucleotide weighting of TFBS matrix patterns [15, 16]. These patterns extend the conserved core of a TFBS by considering the neighbouring sequences and using this information for TFBS scoring and TFBS-module identification. Furthermore, the TFBS-modules are based on functional experiments as for example hormonal treatments of transgenes with constructs containing the promoter elements and reporter genes [17, 18]. Thus, the TFBS-modules of a gene promoter are evidence-based potential regulatory regions and therefore useful as marker sequence for an associated gene function.

This study focused on the structural analysis of hop genes coding for specific enzymes with a known function within the BA biosynthesis pathway: valerophenone synthase *(VPS)*, prenyltransferase 1 *(PT1)*, and prenyltransferase 2 *(PT2)* [19–21]. Real-Time quantitative PCR (RTqPCR) was then employed to identify the BA-synthesis gene that had the strongest association with BA concentrations in hop cones of the evaluated genotypes. This specific gene was then analyzed for TFBS-modules and module-related transcription factors. These transcription factors were also analyzed for their structure and TFBS-modules. Moreover, TFBS-modules present in BA and BA-related transcription factor genes with known regulatory functions in the secondary metabolism are suggested to have regulatory function in the BA genes. Here we describe a process for a knowledge-based selection of targets for targeted sequencing (Fig. 1B Box 1 and 2) by using a small set of bioinformatical tools. Follow-up experiments will shown that the identification of molecular markers can be done more efficiently by using targeted sequencing rather than random sequencing, at least for well-known traits such as bitter acid metabolism in hops.

2 Materials and methods

2.1 Gene structure analysis

The gene structure analysis was based on the functionally verified sequences of the valerophenone synthase (*VPS*, GenBank ID: AB047593), prenyltransferase 1 (*PT1*, GenBank-ID: KM222441) and prenyltransferase 2 (*PT2*, GenBank-ID: KM222442) [19, 21, 22], as well as on the Basic Leucine Zipper (bZIP) Transcription Factors *HlbZip1* (GenBank ID: FN395065) and *HlbZip2* (GenBank ID: AM998490). The complementary DNAs (cDNAs) were used as input for a local BLAST search (Geneious 11.1.5, New Zealand) within the genome of the hop cultivar 'Cascade' [14]. This led to two contigs that corresponded to the two alleles of *VPS*, *PT1*, *PT2*, *HlbZip1* and *HlbZip2*, respectively, in the diploid 'Cascade' genotype. The target containing contig sequences have then been examined for the most likely gene structure by using SPLIGN, a BLAST-based tool which uses refined alignments and splicing site prediction in order to improve gene structure probability [23]. The resulting gene structure has then further been compared to the *ab initio* results of *Augustus* [24] and by aligning

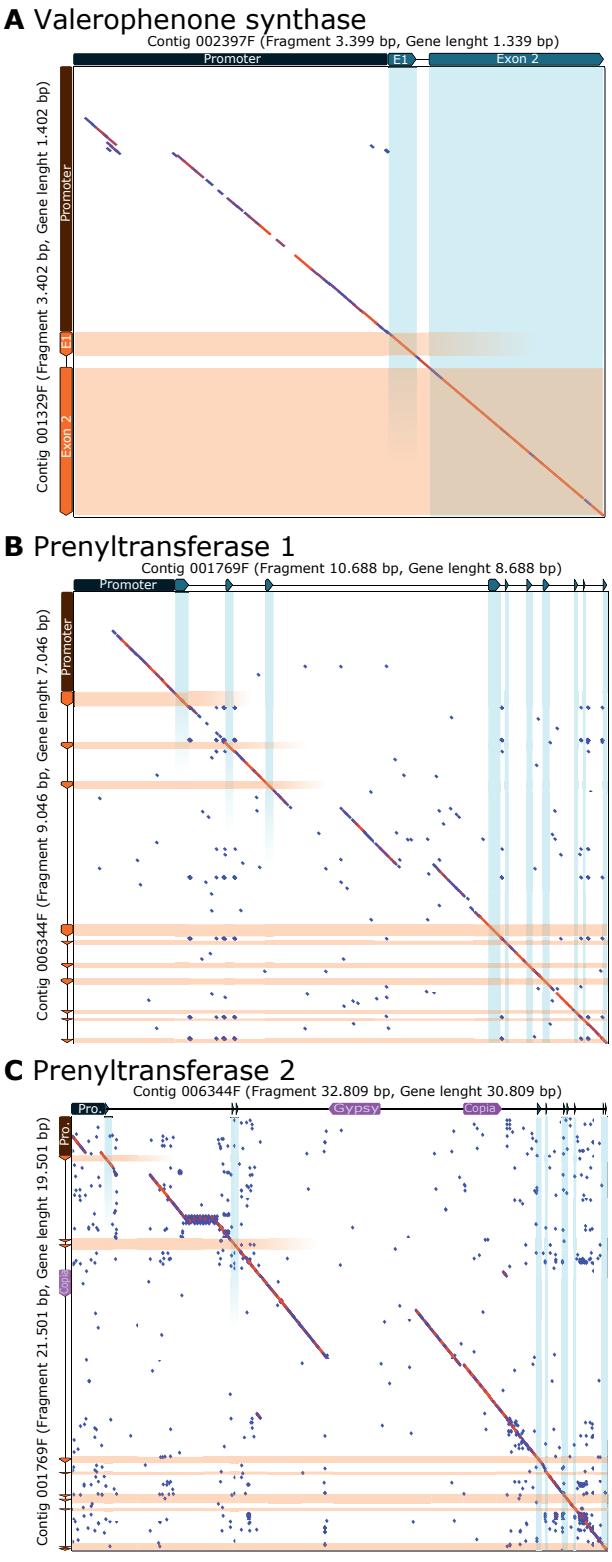

Fig. 2 **Sequence alignment dotplot graph, comparing the two alleles (contigs) of the promoter and gene structure of the bitter acid genes (A) valerophenone synthase, (B) prenyltransferase 1 and (C) prenyltransferase 2, respectively. A 2 kb long putative promoter region (grey) is followed by the exons (blue or red), while introns are represented as lines connecting the exons. Only retrotransposons of more than 1 kb are presented (pink). The alignment is shown as diagonal line, ranging from blue (low % identity) to red (high % identity), while blue dots result from random sequence homology. Gaps within the alignment line represent either insertions within one of the sequences or the lack of sequence homology. The locations of exons along the alignment are shown by overlapping transparent bars, blue bars representing exons from the longer gene (x-axis) and orange bars represent exons from the shorter gene (y-axis). Tile size is 2 kb**

transcriptome sequences from RNA sequencing projects [6, 25]. Additionally to the gene structure an analysis of potential TFBS within the putative promoter region was conducted (Genomatix, Germany; Promoter Module Library Version 6.2) [15, 16]. Finally, contigs containing the target gene have been analyzed for the presence of potential transposable elements and repeats by the CENSOR tool (https://www.girinst.org/censor/index.php; Datenbank RepBase Version 23.12) [26]. The resulting annotations were stacked with Geneious and the key results were presented in a sequence alignment dotplot graph (Fig. 2).

2.2 Plant Material

To identify major differences at the sequence level, two groups of hops, aroma hops with low concentrations of alpha acids (humolone) and bitter hops with high concentrations of alpha acids, were compared (Supplement 1). Sampling of those genotypes for RTqPCR-analysis was conducted in August 2018 at the hop research station Straß near Tettnang, Germany. Samples were taken at the cone developmental stage 7 [27]. Earlier findings showed that the expression levels of *VPS* and/or *PT1* are still high at this stage and thereafter either decrease [27] or stay high until harvest as described by Castro et al. [28]. Furthermore, *Clark* et al. [25] showed that the expression of both genes is much higher in mature cones compared to leaves. Three samples from each of three plants per genotype were taken. In a preliminary experiment three samples from each of three plants of genotype H024 were taken to estimate the intra-genotype variability that was found not to be significant ($p > 0.05$). Subsequently, it was extrapolated that this sample size is sufficient for all genotypes. Samples were immediately frozen on dry ice and stored at –80 °C until further processing.

2.3 Gene expression studies

Plant samples were ground in liquid nitrogen to a fine powder with mortar and pestle. Total RNA was extracted from 100 mg subsamples with the InviTrap Spin Plant RNA Mini Kit (Stratec, Berlin, Germany), following the manufacturer's recommendations, including DNase treatment and the use of the Lysis buffer RP. The RNA quality and quantity were determined by gel electrophoresis and photospectrometry. RNA samples were stored at – 80 °C until cDNA was synthesized using the TaqMan Reverse Transcription Kit (Applied Biosystems, Foster City, USA). For cDNA synthesis 500 ng of total RNA was used for each reaction. cDNA quality and subsequently transcription levels of the target genes *VPS*, *PT1* and *PT2* were tested by RTqPCR using a Rotor-Gene 6000 cycler (Qiagen, Hilden, Germany). The parameters for the reaction were as follows: initial denaturation (3 min, 95 °C), 40 cycles of denaturation (20 s, 95 °C), annealing (20 s, 58 °C), extension (20 s, 72 °C) and followed by a final melt curve from 60 °C to 99 °C in 0.5 K steps. A pooled sample, composed of 1 µl of 1 to 5 diluted cDNA from each cDNA sample, was used in each run as a reference for the relative gene expression and as a standard for the different runs.

A set of four previously characterized reference genes were re-evaluated for expression stability in the cones and leaves under the given conditions [29]: *7SL*, *ACTIN*, *GAPDH*, and *NADH*. The reference gene analysis was performed with the BestKeeper tool [30]. *7SL*-RNA signal recognition particle for protein transport (*7SL*) was selected as reference gene because it revealed the highest expression stability, indicated by the lowest standard deviation of the absolute regulation coefficient: 1.84 for *7SL*, 1.98 for *ACTIN*, 2.12 for *GAPDH*, and 5.23 for *NADH*. It is important to note that the primers of the reference gene *GAPDH*, which has been used in previous expression studies of hop genes [29], produce two amplicons of different size as conclusively confirmed by melt curve analysis and gel electrophoresis.

Table 1 **Hop genotypes of the Hüll collection of cultivars and breeding lines were grouped into two categories: Low and high alpha acid, respectively, based on genotype-specific concentrations at harvest. Data were provided by the Hop Research Center Hüll of the Bavarian State Research Center for Agriculture. Data represent average values of at least 5-year-(semi-)commercial trials**

Group	Genotype	Alpha acid [g/100gDW]	Beta acid [g/100gDW]	Beta acid to alpha acid ratio
Low alpha	H002	4.1	5.3	1.3
	H003	4.0	5.6	1.4
	H006	3.1	7.4	2.4
	H026	3.3	6.6	2.0
High alpha	H024	16.7	5.0	0.3
	H031	18.6	5.6	0.3
	H129	19.3	4.8	0.2
	H132	16.8	4.4	0.3

Three primer pairs for each of the three target genes were designed, except the pair for PT1/1 (Supplement 1). The efficiency of each primer pair was determined with DART tool [31]. Primer specificity was confirmed by melt curve analyses for each individual run and by sequencing of the resulting amplicons. The resulting amplicons have been blasted against the NCBI nucleotide collection and against the 'Cascade' genome [14], resulting in two hits per amplicon, which correspond to the two alleles of the diploid cultivar. Relative expression of the target genes was analyzed with the ΔΔCt-method, since the primer efficiency between amplicons was similar at 0.94 ± 0.1 (statistically insignificant). In order to choose the best primer pair per target gene, RTqPCR runs were performed with 10 cone and 2 leaf samples, respectively, focusing on the highest expression stability and sensitivity.

3 Results and discussion

All evaluated genes had two copies (Fig. 2), which are most likely alleles, since 'Cascade' is a diploid hop cultivar (hopbase.org). Sequence alignment was performed to evaluate structural commonalities and to proof that the two copies indeed belong to the same gene. In agreement with earlier data on the structure of the *VPS* gene, the analysis of the 'Cascade' genome showed that *VPS* consists of a short initial exon, followed by a short intron and a longer second exon (Fig. 2A) [19]. The first 1.400 bp of the putative *VPS* promoter region, upstream of exon 1, were less heterozygous (92 % identity) compared an overall identity of 66 % for the whole promoter region assumed to be 2 kb upstream of exon 1 (Fig. 2 A, Supplement 2). Unlike *VPS*, the principal gene structures of

Fig. 3 Promoter comparison of genes from the (A) bitter acid biosynthesis and (B) related transcription factors. Sequences are represented by horizontal bars with the conserved regions (>90% identity) coloured in dark grey and the heterogeneous regions in bright grey. Transcription factor binding site (TFBS)-modules, resulting from the Genomatix Model-Inspector analysis were divided into three groups: TFBS-modules associated with basic functions (black), those associated with general gene regulatory transcription factors (purple), and those associated with the secondary metabolism (red)

the *PT1* and the *PT2* consist each of 10 exons and thus are more complex (Fig. 2B). It is noteworthy, that this gene structure is not shown in the alignments presented on the hopbase.org genome browser. However, comparing the *PT1* alleles to those of the *VPS* promoters revealed similarities with a conserved region of 93 % identity, starting at 1.200 bp prior to exon 1, whereas the full 2 kb of the putative promoter region have an identity of only 73 % (Fig. 2 A and B; Fig. 3 A, Supplement 2). The putative promoter sequences of the *PT2* alleles also share a conserved region, however, one allele (contig 006344F) seems to be interrupted by an insert. Aligning the alleles revealed high sequence identity for the first 250 bp upstream of exon 1 and again starting at 1200 bp upstream of exon 1 (Fig 2 C; Fig 3 C). The CENSOR analysis of the *PT2* promoter insert in contig 006344F showed that it contains two short but independent DNA MuDR transposons of the highly active Mutator superfamily, originally described for maize [32] but also for example in grapevine [33]. Generally, the insertion of transposable elements can have tremendous effects on gene regulation with, for example, disruption of a promoter function [34]. Therefore, it is important to consider these sequence variations in targeted sequencing approaches to detect the occurrence of inserts and to evaluate whether they are correlated to an overall reduced gene function as for example in a lowered concentration of BAs.

By comparing the *VPS* and PT gene structures it became evident that there is an enormous span from the first to the last exon of these genes. It is 5-times or up to 20-times longer for *PT1* or *PT2*, respectively, when compared to the *VPS* (Fig. 2). Since the length of the coding regions of all three genes is similar, with around 1.200 bp for the *VPS* [28], 1.245 bp for the *PT1* and 1.227 bp for the *PT2* [21], respectively, the enormous length of the PT must be the result of the presence of large introns. Generally, less than 2 % of plant introns are longer than 2 kb [12] and only 16 genes have been identified in the model plant Arabidopsis which have a length greater than 5 kb [35]. However, from the Arabidopsis model it is known that (1) such long introns can actually be spliced out and (2) that these long introns are mainly composed of transposable elements [35]. The sequencing of the hop cultivar 'Shinsu Wase' already showed that about 32 % of the hop genome consists of long terminal repeat retrotransposons (LTRs) [6]. Evaluating the 'Cascade' genomic region of *PT1* and *PT2* for longer than 1 kb LTR-sequences showed that intron 3 in *PT2* contains LTRs; independent ones from the Copia family in each of the two contigs and one LTR from the Gypsy family only in contig 006344F (Fig. 2). The insertion of the Gypsy-LTR is mainly responsible for the size difference between the two *PT2* alleles. Compared to intron 3 of *PT1*, the first 3 kb of the *PT2*-intron 3 are highly similar (> 90 % identity), then an almost 8 kb long insert, containing the Gypsy-LTR, is only present in contig 006344F, and the remaining intron is again similar (> 85 % identity). This difference in intron length between the PT alleles raises the question, if a large intron is correlated with less effective post-transcriptional modifications, thus reduced gene expression and subsequently lower enzyme concentrations and metabolic activity. It is noteworthy that both *PT1* and *PT2* genes have 10 exons in the same genomic region, separated by only around 100 kb, and have for intron 3 a sequence identity of 63 %. The findings suggest that both genes are the result of an ancient gene duplication. A similar gene duplication has recently been postulated for two metabolic genes of cannabis, a close relative of hop [36].

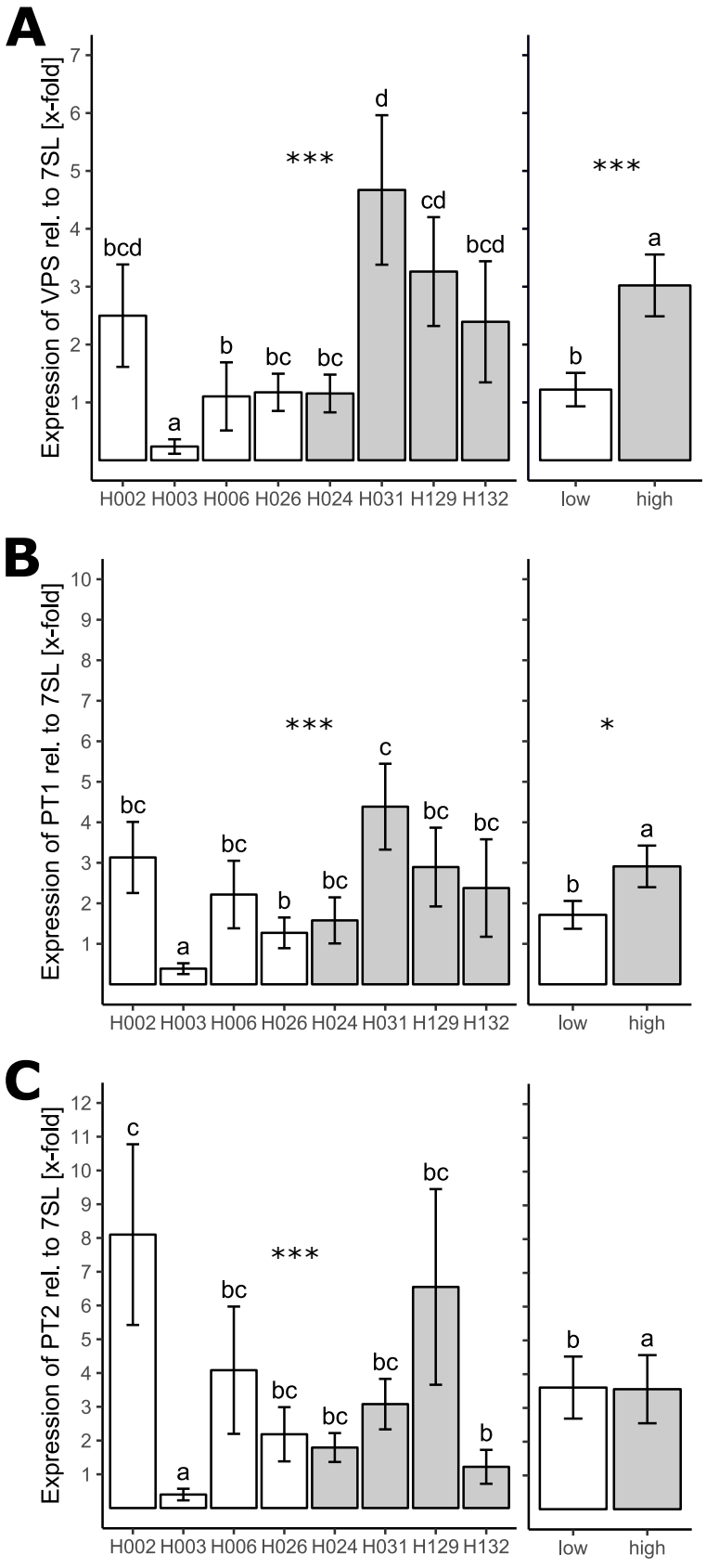

Fig. 4 **RT-qPCR analysis results of the bitter acid biosynthesis genes relative to the reference gene (7SL-RNA signal recognition particle for protein transport (7SL)). Left columns in the graphs show the results for individual cultivars, while the right colum shows the average of the cultivars based on the categories low alpha hops (open column) and high alpha hops (shaded column). The examined genes are (A) the valerophenone synthase (VPS), (B) the prenyltransferase 1 (PT1) and (C) the prenyltransferase 2 (PT2)**

The expression studies of the BA-biosynthesis genes indicate that the *VPS* transcript level was lowest in the low alpha genotype H003 and highest in H031 (Fig. 4), which is known for very high alpha acid concentrations. Moreover, *VPS* expression levels were low in the low alpha genotypes H006 and H026 but were high in the high alpha genotypes H129 and H132. These results generally support an association of the *VPS* gene activity with the alpha acid concentration typical for the genotype. However, two genotypes did not follow this pattern. The low alpha genotype H002 had comparatively high *VPS* expression levels, while the high alpha genotype H024 showed a below average *VPS* expression level compared to the other high alpha genotypes. This characteristic might be based on different temporal expression pattern of these genotypes, thus gene expression time-series as done for the landrace 'Tettnanger' [27] would need to be conducted for each of these cultivars in order to support this explanation. The expression level of *PT1* has the same pattern compared to the *VPS* (Fig. 4B), yet to a lesser but still significant extent ($p > 0.05$). However, *PT2* expression does not fit this pattern and the comparison of low and high alpha cultivars revealed no significant differences. It is noteworthy that the genotype H003 showed the lowest expression levels of all three analyzed genes. Expression analysis of the three BA genes indicated that the *VPS* has the strongest association between gene expression levels and BA concentrations. This is not surprising, considering that *VPS* catalyses the synthesis of phlorisovalerophenone, the first BA-specific metabolite [19, 37]. The next step of the BA biosynthesis is mediated by *PT1*, which has a double function in hop metabolism. It is not just a key enzyme for BA-biosynthesis, but it plays also a role in the prenylation of naringenine chalcone, a precursor for the hop specific pharmacologically important prenylflavonoids xanthohumol and 8-prenylnaringenine [21]. The two biosynthesis pathways catalysed by *PT1* might lead to a "dilution" of the correlation between the *PT1* expression level and the BA concentration. The *PT2* is thought to catalyze by single prenylation alpha acids and by double prenylation beta acids [21]. Consequently, the *PT2* regulates the alpha acids to beta acids ratio by the level of prenylation reactions. Figure 4 clearly shows that the *VPS* transcript levels had the strongest association with the BA concentration and, therefore, this gene was analyzed in more detail.

Further characterisation of the putative promoter of the BA genes resulted in the identification of 500 TFBS from over 100 TFBS-families in each of the promoters (data not shown). However, only 14 TFBS-modules have been identified within the BA gene promoters (Fig. 3, Supplement 3) of which most modules are associated with basic functions as the sugar responsive module SUCB_SURE_01 in *VPS* and *PT2* or the light or senescence responsive modules in *PT1*, GTBX_GTBX_01 and ABRE_GBOX_01, respectively. The study concentrated on the *VPS* promoters associated due to the association between *VPS* expression level and BA-concentration. It is interesting to note that both *VPS* promoters have the same two TFBS-modules, the sucrose response module (SUCB_SURE_01) and the module OPAQ_LEGB_01, with the latter module functionally proven in the context of *Arabidopsis* seed maturity [18]. However, the OPAQ_LEGB_01 module has binding sites for the transcription factors basic leucine-zipper (*bZIP*) and *AtFUSCA3* (GenBank-ID: AT3G26790.1) and both are co-regulated and important for petal organ identity [18, 38]. In agreement with this finding, *bZip* was also found to be expressed in mature hop cones [39], however, to our knowledge, AtFUSCA has not been discussed in hop. A BLAST search for *AtFUSCA3* within the hop transcriptome released by Clark et al. [25] revealed a single hit, the transcript comp487760_c0_seq1. This sequence has an *AtFUSCA3* typical 4 exon gene structure when it was aligned to the 'Cascade' genome (data not shown). Furthermore, the transcript is expressed in mature hop cone (supplemental data of Clark et al. [25]). These results support the authenticity of the OPAQ_LEGB_01 module, since its related transcription factors, *AtFUSCA3* and *bZip*, are both expressed in the hop flower (cone), which is in agreement with the findings for Arabidopsis [38] and hop [39]. The *bZip* transcription factors were already functionally characterized by Matousek et al. [39], who cloned and analyzed two hop *bZip* transcription factors, *HlbZip1* and *HlbZip2*. Transient expression of these factors in tobacco showed that they can activate the *VPS* promoter. Because functional evidence to date only exists for the *bZip* factors from hop but not for *FUSCA3*, only *HlbZip1* and *HlbZip2* were included in the gene structure and promoter analysis (Fig. 3B, Supplement 2). Both *bZip* genes have a *bZip*-typical 4 exon structure (data not shown) and a conserved promoter region, which is shorter for *HlbZip2* (770 bp) compared to that of *HlbZip1* (full 2 kb). The analysis of TFBS-modules suggested that the *HlbZip1* promoter regions have little allelic variation (90 % identity) and both alleles share 5 TFBS-modules (Fig. 2B, Supplement 2). In detail, the two *HlbZip1* allelic promoter regions share modules associated with basic plant function such as sugar response (SUCB_SURE_01), oxidative stress (SWNS_SWNS_02) reaction, dehydration (GTBX_MYCL_01), or regulation of the circadian rhythm (CCAF_CCAF_03), but the *HlbZip1* promoter in contig 000839F contains a TALE_TALE_01 module, which can lead to enhanced terpene biosynthesis [40]. Interestingly, such a module is also present in the *PT1* promoter (Supplement 3). Consequently, the OPAQ_LEGB_01 modules of the *VPS* promoters and the TALE_TALE_01 modules of *PT1* and *HlbZip1* promoters, respectively, are promising candidates for a targeted sequencing approach.

In summary, two gene copies, presumably allelic variations, were identified for all examined target genes within the 'Cascade' genome. While the gene structure of *VPS* consists of two exons, the gene structure of *PT* is rather complex with a span of up to 31 kb. This enormous length can partly be explained by the presence of transposable elements. They have also been found in the putative regulatory promoter sequences and should thus be included as targets in future targeted sequencing approaches. Further comparison of alleles of the promoter sequences showed that the region approximately 1 kb upstream of the first exon is best conserved with increasing levels of heterozygosity thereafter. Few TFBS-modules have been identified in both alleles of the examined genes and were identified as potential targets for a knowledge-guided sequencing approach.

4 Conclusion

A targeted sequencing approach is needed for identifying genetic variants of genotypes with very different gene structure. Moreover, using TFBS-modules, to guide targeted sequencing in comparison to non-targeted random GBS sequencing, will most likely identify more relevant genetic variation that assists the development of robust molecular markers. However, in continuation of this preparatory work, it will be demonstrated experimentally that targeted sequencing is a key tool for modern breeding strategies of specialty crops.

Acknowledgement

The project was financially supported by funds of the German Government's Special Purpose Fund held at Landwirtschaftliche Rentenbank. The authors thank Dr. Elisabeth Seigner and Anton Lutz for providing data on the genotypes and for critical discussions of the manuscript.

5 References

1. Lutz, A. and Seigner, E.: Innovationen rund um die Hüller Hopfenzüchtung, BRAUWELT, **155** (2015), no. 3, pp. 57-59.
2. Matthews, P.D.; Coles, M.C.; Pitra, N.J.; Matthews, P.D.; Coles, M.C. and Pitra, N.J.: Next generation sequencing for a plant of great tradition: Application of NGS to SNP detection and validation in hops (*Humulus lupulus* L.), BrewingScience – Monatsschrift für Brauwissenschaft, **66** (2013), no. 11/12, pp. 185-191.
3. Jakse, J.; Cerenak, A.; Radisek, S.; Satovic, Z.; Luthar, Z. and Javornik, B.: Identification of quantitative trait loci for resistance to Verticillium wilt and yield parameters in hop (*Humulus lupulus* L.), Theoretical and Applied Genetics, **126** (2013), no. 6, pp. 1431-1443.
4. Henning, J.A.; Gent, D.H.; Townsend, M.S.; Woods, J.L.; Hill, S.T. and Hendrix, D.: QTL analysis of resistance to powdery mildew in hop (*Humulus lupulus* L.), Euphytica, **213** (2017), no. 4, p. 98.
5. Bergey, C.M.; Pozzi, L.; Disotell, T.R. and Burrell, A.S.: A New Method for Genome-wide Marker Development and Genotyping Holds Great Promise for Molecular Primatology, International Journal of Primatology, **34** (2013), no. 2, pp. 303-314.
6. Natsume, S.; Takagi, H.; Shiraishi, A.; Murata, J.; Toyonaga, H.; Patzak, J.; Takagi, M.; Yaegashi, H.; Uemura, A.; Mitsuoka, C.; Yoshida, K.; Krofta, K.; Satake, H.; Terauchi, R. and Ono, E.: The draft genome of hop (*Humulus lupulus*), an Essence for Brewing, Plant & cell physiology, **56** (2015), no. 3, pp. 428-41.
7. Campbell, N.R.; Harmon, S.A. and Narum, S.R.: Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping

method based on custom amplicon sequencing, Molecular Ecology Resources, **15** (2015), no. 4, pp. 855-867.

8. Bybee, S.M.; Bracken-Grissom, H.; Haynes, B.D.; Hermansen, R.A.; Byers, R.L.; Clement, M.J.; Udall, J.A.; Wilcox, E.R. and Crandall, K.A.: Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics., Genome biology and evolution, **3** (2011), pp. 1312-23.
9. Senapathy, P.; Bhasi, A.; Mattox, J.; Dhandapany, P.S. and Sadayappan, S.: Targeted genome-wide enrichment of functional regions, PLoS ONE, **5** (2010), no. 6, p. e11138.
10. Lin, T.; Sharma, P.; Gonzalez, D.H.; Viola, I.L. and Hannapel, D.J.: The Impact of the Long-Distance Transport of a BEL1-Like Messenger RNA on Development 1W.OA., Plant Physiology Ò, **161** (2013), no. 2, pp. 760-772.
11. Mishra, A.K.; Duraisamy, G.S.; Matoušek, J.; Radisek, S.; Javornik, B. and Jakse, J.: Identification and characterization of microRNAs in *Humulus lupulus* using high-throughput sequencing and their response to Citrus bark cracking viroid (CBCVd) infection, BMC Genomics, **17** (2016), no. 1, p. 919.
12. Li, Y.; Xu, Y. and Ma, Z.: Comparative analysis of the exon-intron structure in eukaryotic genomes, Yangtze Medicine, **01** (2017), no. 01, pp. 50-64.
13. Stival Sena, J.; Giguère, I.; Boyle, B.; Rigault, P.; Birol, I.; Zuccolo, A.; Ritland, K.; Ritland, C.; Bohlmann, J.; Jones, S.; Bousquet, J. and Mackay, J.: Evolution of gene structure in the conifer Picea glauca: a comparative analysis of the impact of intron size, BMC Plant Biology, **14** (2014), no. 1, p. 95.
14. Hill, S.T.; Sudarsanam, R.; Henning, J. and Hendrix, D.: HopBase: a unified resource for Humulus genomics, Database, **2017** (2017), no. 1.
15. Klingenhoff, A.; Frech, K.; Quandt, K. and Werner, T.: Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity, Bioinformatics, **15** (1999), no. 3, pp. 180-186.
16. Cartharius, K.; Frech, K.; Grote, K.; Klocke, B.; Haltmeier, M.; Klingenhoff, A.; Frisch, M.; Bayerlein, M. and Werner, T.: MatInspector and beyond: promoter analysis based on transcription factor binding sites, Bioinformatics, **21** (2005), no. 13, pp. 2933-2942.
17. Winterhagen, P.; Hagemann, M.H. and Wünsche, J.N.: Different regulatory modules of two mango ERS1 promoters modulate specific gene expression in response to phytohormones in transgenic model plants, Plant Science, **289** (2019), p. 110269.
18. Roschzttardtz, H.; Fuentes, I.; Vásquez, M.; Corvalán, C.; León, G.; Gómez, I.; Araya, A.; Holuigue, L.; Vicente-Carbajosa, J. and Jordana, X.: A nuclear gene encoding the iron-sulfur subunit of mitochondrial complex II is regulated by B3 domain transcription factors during seed development in arabidopsis, Plant Physiology, **150** (2009), no. 1, pp. 84-95.
19. Okada, Y. and Ito, K.: Cloning and analysis of valerophenone synthase gene expressed specifically in lupulin gland of hop (*Humulus lupulus* L.), Bioscience, Biotechnology, and Biochemistry, **65** (2001), no. 1, pp. 150-155.
20. Tsurumaru, Y.; Sasaki, K.; Miyawaki, T.; Momma, T.; Umemoto, N. and Yazaki, K.: An aromatic prenyltransferase-like gene HIPT-1 preferentially expressed in lupulin glands of hop, Plant Biotechnology, **27** (2010), no. 2, pp. 199-204.
21. Li, H.; Ban, Z.; Qin, H.; Ma, L.; King, A.J. and Wang, G.: A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway, Plant Physiology, **167** (2015), no. 3, pp. 650-659.
22. Xu, H.; Zhang, F.; Liu, B.; Huhman, D. V.; Sumner, L.W.; Dixon, R.A. and Wang, G.: Characterization of the formation of branched short-chain fatty acid:coas for bitter acid biosynthesis in hop glandular trichomes, Molecular Plant, **6** (2013), no. 4, pp. 1301-1317.
23. Kapustin, Y.; Souvorov, A.; Tatusova, T. and Lipman, D.: Splign: algorithms for computing spliced alignments with identification of paralogs, Biology Direct, **3** (2008), no. 1, p. 20.
24. Stanke, M. and Morgenstern, B.: AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints., Nucleic acids research, **33** (2005), no. Web Server issue, pp. W465-7.
25. Clark, S.M.; Vaitheeswaran, V.; Ambrose, S.J.; Purves, R.W. and Page, J.E.: Transcriptome analysis of bitter acid biosynthesis and precursor pathways in hop (*Humulus lupulus*), BMC Plant Biology, **13** (2013), no. 1, p. 12.
26. Kohany, O.; Gentles, A.J.; Hankus, L. and Jurka, J.: Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor, BMC bioinformatics, **7** (2006), p. 474.
27. Häntzschel, K.R.; Cevik, E.; Collenberg, M.; Born, U.; Gatica-Arias, A. and Weber, G.: Biosynthesis of flavonoids and phloroglucinols: gene expression during hop (*Humulus lupulus* l.) flower development, Acta Horticulturae, (2013), no. 1010, pp. 53-58.
28. Castro, C.B.; Whittock, L.D.; Whittock, S.P.; Leggett, G. and Koutoulis, A.: DNA sequence and expression variation of hop (*Humulus lupulus*) valerophenone synthase (*VPS*), a key gene in bitter acid biosynthesis, Annals of Botany, **102** (2008), no. 2, pp. 265-273.
29. Maloukh, L.; Matousek, J.; Van Bockstaele, E. and Roldán-Ruiz, I.: Housekeeping gene selection for real time-PCR normalization in female hop (*Humulus lupulus* L.) tissues, Journal of Plant Biochemistry and Biotechnology, **18** (2009), no. 1, pp. 53-58.
30. Pfaffl, M.W.; Tichopad, A.; Prgomet, C. and Neuvians, T.P.: Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations., Biotechnology letters, (2004).
31. Peirson, S.N.: Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis, Nucleic Acids Research, **31** (2003), no. 14, pp. 73e-73.
32. Robertson, D.S.: Characterization of a mutator system in maize, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, **51** (1978), no. 1, pp. 21-28.
33. Benjak, A.; Forneck, A. and Casacuberta, J.M.: Genome-wide analysis of the "cut-and-paste" transposons of grapevine, PLoS ONE, **3** (2008), no. 9, p. e3107.
34. Lerman, D.N. and Feder, M.E.: Naturally occurring transposable elements disrupt hsp70 promoter function in Drosophila melanogaster, Molecular Biology and Evolution, **22** (2005), no. 3, pp. 776-783.
35. Chang, N.; Sun, Q.; Hu, J.; An, C. and Gao, A.H.: Large introns of 5 to 10 kilo base pairs can be spliced out in arabidopsis., Genes, **8** (2017), no. 8.
36. Laverty, K.U.; Stout, J.M.; Sullivan, M.J.; Shah, H.; Gill, N.; Holbrook, L.; Deikus, G.; Sebra, R.; Hughes, T.R.; Page, J.E. and Bakel, H. van: A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci, Genome Research, **29** (2019), no. 1, pp. 146-156.
37. Goese, M.; Kammhuber, K.; Bacher, A.; Zenk, M.H. and Eisenreich, W.: Biosynthesis of bitter acids in hops. A (13)C-NMR and (2)H-NMR study on the building blocks of humulone., European journal of biochemistry, **263** (1999), no. 2, pp. 447-54.
38. Gazzarrini, S.; Tsuchiya, Y.; Lumba, S.; Okamoto, M. and McCourt,

P.: The transcription factor *FUSCA3* controls developmental timing in arabidopsis through the hormones aibberellin and abscisic acid, Developmental Cell, **7** (2004), no. 3, pp. 373-385.

39. Matoušek, J.; Kocábek, T.; Patzak, J.; Stehlík, J.; Füssy, Z.; Krofta, K.; Heyerick, A.; Roldán-Ruiz, I.; Maloukh, L. and Keukeleire, D. De: Cloning and molecular analysis of *HlbZip1* and *HlbZip2* transcription factors putatively involved in the regulation of the lupulin metabolome in hop (*Humulus lupulus* L.), Journal of Agricultural and Food Chemistry, **58** (2010), no. 2, pp. 902-912.
40. Xu, X.; van Lammeren, A.A.M.; Vermeer, E.; Vreugdenhil, D. and Hannapel, D.: The role of gibberellin, abscisic acid, and sucrose in the regulation of potato tuber formation in vitro, Plant Physiology, **117** (1998), no. 2, pp. 575-584.

Received 23 September 2019, accepted 28 November 2019

Supplement

Supplement 1 **Primer pairs used in RTqPCR expression analysis of the genes of interest**

Pair	Primer	Sequence (5'->3')	Amplicon length [bp]	Amplicon Quality
VPS/1	Hl_VPS_1F	CGACTACTACTTTCGTGTCA	240	Stable, selected
	Hl_VPS_1R	GGTGCATTGTACTCGCACAG		
VPS/2	Hl_VPS_2F	CTCCGATTGGGATTAATGAC	159	stable
	Hl_VPS_2R	TTGCACATGACATGTTCCCG		
VPS/3	Hl_VPS_3F	CTCCGATTGGGATTAATGAC	238	stable
	Hl_CHSfam_1R	CACTCCAGTCCATCTCCG		
PT1/1	PT1_qPCR-F. [1]	AGTCCCACTCAAAATGGCCAA	106	unstable
	PT1_qPCR-R. [1]	ACACCAAACTTGCGATCTCC		
PT1/2	Hl_PT1_Exon3_F	TGGCCTCTAATTTTTAGGGCACT	264	insensitive
	Hl_PT1_Exon4_R	AGTCCCACTCAAAATGGCCAA		
PT1/3	Hl_PT1_Exon4_F	TCAGGGCGTATTTCTGTGGAA	135	stable, selected
	Hl_PT1_Exon4_R	AGTCCCACTCAAAATGGCCAA		
PT2/1	Hl_PTR2_Exon6_F	ATGTTCCTGATGTAGAAGGTGAC	236	stable, selected
	Hl_PT2_qPCR_R	ATTTGCTTTGTCCAACAACCA A		
PT2/2	Hl_PT2_qPCR_2F	ACGATTTGGAAAGTGACAGGAT	200	stable
	Hl_PT2_qPCR_2F	GAGGAACAGAATACATAGCCCC		
PT2/3	Hl_PT2_Exon1_F	TCCACCAAGAAGTTGCAGACAA	101	unstable
	Hl_PT2_Exon1_R	AGGGAATTTGGCCTTGGAAGAT		

Supplement 2 **Gen structure analysis of the bitter acid biosynthesis and basic leucine-zipper (bZip) transcription factor genes and comparison of their promoter alleles (contigs) based on the 'Cascade' genome**

Gene	cDNA	Contig	Exon num.	Gen length [bp]	Promoter Ident. [%]	Conserved Promoter	
						Position	Ident. [%]
VPS	AB01 5430	002397F	2	1.399	65,8	0 to –1400	92,0
		001329F	2	1.402			
PT1	KM22 2441	001769F	10	8.688	72,6	0 to –1200	93,3
		006344F	10	7.046			
PT2	KM22 2442	001769F	10	19.501	43,0	–251 to –1060 / –1200 to –2000; 0 to –250	93,8
		006344F	10	30.809			
HlbZip1	FN39 5065	000839F	4	12.860	90,0	0 to –2000	90,0
		002829F	4	12.813			
HlbZip2	AM99 8490	006711F	4	3.152	63,3	0 to –670	87,2
		006271F	4	3.171			

Supplement 3 **Transcription factor binding site-Modules, resulting from the Genomatix Model-Inspector analysis of promoter alleles (contigs) of the bitter acid biosynthesis and bZip transcription factor genes based on the 'Cascade' genome**

Promoter	Contig	TFBS-Modul	Position	Modul description
VPS	002397F	SUCB_SURE_01	1701-1725	Sugar responsive element
		OPAQ_LEGB_01	1790-1861	bZIP and RY elements
	001329F	SUCB_SURE_01	1707-1728	Sugar responsive element
		OPAQ_LEGB_01	1799-1866	bZIP and RY elements
PT1	001769F	L1BX_MYBL_01	614-577	MYB and HOX3 elements
		GTBX_GTBX_01	793-843	Light-responsive element
	006344F	TALE_TALE_01	314-287	Correlated with Isoprenylmetabolism
		GBOX_ABRE_01	449-487	bZIP elements
		ABRE_GBOX_01	467-407	Senescence-responsive gene expression.
		MYCL_MYBL_01	468-442	BR responsive BES and MYB elements
		AHBP_DOFF_01	1900-1833	Cell-specific expression
PT2	001769F	MYBL_DOFF_02	301-156	GAMYB and DOF for full gene activation
		SUCB_SURE_01	498-523	Sugar responsive element
	006344F	GTBX_MYCL_01	1787-1647	Dehydration-response
bZip1	000839F	TALE_TALE_01	52-76	Correlated with Isoprenylmetabolism
		AHBP_AHBP_AHBP _AHBP_AHBP_01	262-386	Concentration dependent developmental regulation
		SUCB_SURE_01	267-242	Sugar responsive element
		AHBP_DOFF_01	368-434	Cell-specific expression
		SWNS_SWNS_02	1478-1431	Oxidative stress response
	002829F (reverse)	AHBP_AHBP_AHBP _AHBP_AHBP_01	1815-1679	Concentration dependent developmental regulation
		SUCB_SURE_01	1810-1835	Sugar responsive element
		AHBP_DOFF_01	1710-1644	Cell-specific expression
		CCAF_CCAF_03	804-780	Circadian regulation
		SWNS_SWNS_02	605-652	Oxidative stress response
		MYCL_MYBL_01	197-222	BR responsive BES and MYB elements
		GTBX_MYCL_01	76-215	Dehydration response
bZip2	006711F	CCAF_CCAF_03	618-641	Circadian regulation
		GTBX_MYCL_01	1312-1171	Dehydration response
		WBXF_WBXF _DOFF_01	1341-1385	GA-dependent repressome element
	006271F (reverse)	MYCL_GCCF_01	1469-1450	Starch biosynthesis related element
		AHBP_DOFF_01	680-614	Cell-specific expression

References

1. Clark, S.M.; Vaitheeswaran, V.; Ambrose, S.J.; Purves, R.W. and Page, J.E.: Transcriptome analysis of bitter acid biosynthesis and precursor pathways in hop (Humulus lupulus), BMC Plant Biology, **13** (2013), no. 1, p. 12.