

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Hop-variety Identification Using First- and Second-generation Sequencing

---

Hiromasa Yamauchi

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/61673>

---

## Abstract

Twenty-one hop varieties from Europe and the United States were successfully identified by DNA analysis, based on single nucleotide polymorphisms (SNPs; including insertion/deletion sequences) as identification markers. Several dozen megabases of transcriptome sequencing data were obtained by next-generation sequencing of samples from three hop varieties and compared to search for the regions containing SNPs. Consequently, four SNP-rich regions were selected as candidates for exploring identification markers in the hop varieties. Sequence data from these regions in all the tested varieties were obtained by the normal Sanger method and compared for the SNPs present. Combination of these SNPs could work well for identification of the 21 hop varieties. Moreover, the mixture of two varieties could be correctly evaluated by using this method. Hop pellet samples of two different varieties were mixed in various ratios and DNA sequencing was carried out. As a result, 5% contamination of a different variety could be detected by examining the electropherogram of the SNP positions. More quantitative methods for mixture evaluation could be expected using DNA techniques, such as quantitative real-time PCR. Because this SNP-based identification method utilizes the DNA sequence itself, it could be a reproducible tool for accurate identification of the hop varieties.

**Keywords:** hop, identification, NGS, SNP, transcriptome, variety

---

## 1. Introduction

Accurate identification and use of hop (*Humulus lupulus*) varieties is very important for the production of quality beer. Hop varieties are usually identified by sensory analysis (taste, appearance, and smell), detection of the differences in the cone structure, and content of different biochemical substances, including alpha acids and essential oils. However, these methods present certain limitations, as accurate identification of the pelletized hop is not

possible by observing its external appearance. Moreover, the content of biochemical substances in hop can vary depending on the cultivation conditions.

There are several reports on the use of DNA-based analytical techniques such as random amplified polymorphic DNAs (RAPDs) [1–5], restriction fragment length polymorphisms (RFLPs) [3,6], amplified fragment length polymorphisms (AFLPs) [3,4,7,8], and microsatellites (single sequence repeats, SSRs) [7, 9–14], besides other approaches like markers in the spacer or noncoding regions [15], sequence-tagged site (STS) markers [3, 13, 16], and diversity arrays technology (DArT) [17] for the assessment of the genetic diversity in hops. Since most of these methods are based on fingerprinting approach, the results obtained are sometimes unclear and prone to misjudgment, thereby limiting the detection of contaminating varieties.

Analysis of single nucleotide polymorphisms (SNPs; differences of single nucleotide in homologous DNA among different varieties) in genomic DNA might be a better and more reproducible tool for the identification of varieties. SNPs are widely distributed in the genome and could be used as markers for the assessment of genotypes. For example, variation in DNA sequence and expression of valerophenone synthase (VPS) gene, a key gene of the bitter acid biosynthesis pathway, has been investigated in hop, using SNPs [18]. However, a large amount of DNA sequence is needed to obtain sufficient SNPs in order to identify the different varieties. In this context, high throughput next-generation sequencing (NGS) generally provides several hundred thousand–times more sequence data in a single analysis compared to the conventional Sanger method, but the whole genome sequencing by either method is still very expensive and time-consuming. In fact, genome sizes of two representative hop varieties, *lupulus* and *neomexicanus*, are 2.74 and 2.97 Gb, respectively [19], which are comparable to that of the human genome [20].

To overcome these problems, transcriptome analysis has been employed for the identification of hop varieties. Transcriptome is the entire mRNA content, transcribed from the genome, and its size ranges from one hundredth- to two hundredth-parts of the genome. Nevertheless, even by a conservative estimate of an average of one SNP per 1000 bp, based on the frequency of SNPs observed in the human genome [20], 13.5K to 30K SNPs could be expected in a relatively short period. Such a high frequency of SNPs would be enough for the identification of hop varieties. The discovery of a large number of SNPs and their specific combinations in each variety could lead to the identification of many hop varieties and detection of contaminants in mixed varieties using these specific SNP-combination-based markers. In the present study, we developed an SNP-based identification method for the hop varieties [21].

## 2. Research protocols, methods, results, and discussion

### 2.1. Identification of SNP markers by second generation sequencing and transcriptome analysis

#### 2.1.1. SNPs

In order to obtain intravariety DNA polymorphic regions required for developing a hop variety identification technique, we attempted searching for SNPs in a large amount of

sequence data obtained using NGS. We focused on the transcriptome because transcriptome analysis requires about 100-times less data processing than whole-genome analysis, thereby reducing the lead time and cost. Besides, according to our calculation, there are already available data for as many as 15000 SNPs in the transcriptome analysis; the data size, however, is smaller than that for the whole-genome analysis. Thus, we performed an intravariety SNP analysis, using this technique, based on the assumption that SNPs required for the identification of many varieties could be obtained, thereby.

### 2.1.2. *Sample collection and storage*

The hop varieties to be identified were Saaz, Sládek, and Premiant, which originated in Czech Republic; Tradition, Spalter, Spalter Select, Perle, Tettnang, Brewer's Gold, Northern Brewer, Magnum, Herkules, German Nugget, and Taurus, which originated in Germany; and Cascade, Zeus, Summit, Galena, Super Galena, Nugget, and Columbus/Tomahawk, which originated in the United States (here, as is widely assumed, we considered that Columbus and Tomahawk were genetically identical). Pellets or dried samples of these varieties were obtained from appropriate suppliers. Three varieties (referred to as A, B, and C for convenience) were selected, and fresh leaves were collected from these varieties. The leaves were sampled and stored according to the procedure described below:

*Tissue:* Leaves as young (small, yellow-green, and soft) as possible were collected. Those with white foreign matter on the surface were excluded.

*Methodology:* To prevent RNase contamination, leaves were collected with gloved hands and were soaked in a reagent (RNA Save; Biological Industries Israel Beit Haemek Ltd., Israel) for preventing RNA degradation.

*Storage:* Although RNA was stable for at least 1 week even at room temperature, the leaves were stored under refrigeration for as much time as possible until being used for RNA isolation and transcriptome and SNP sequence analysis.

### 2.1.3. *Transcriptome analysis*

The collected samples were used for transcriptome analysis. The procedure was subcontracted to Eurofins Genomics (Ebersberg, Germany). Briefly, the total RNA was extracted from the samples using RNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) according to manufacturer's procedure. The quality of total RNA was evaluated in terms of the degree of degradation of rRNAs with Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Normalized cDNA library for use in Roche GS FLX Titanium sequencing was prepared as follows: poly (A) + RNA was isolated from the total RNA, and the first-strand cDNA synthesis was primed with an N6 randomized primer. Normalization was carried out by one cycle of denaturation and reassociation of the cDNA. Reassociated double-stranded (ds) cDNA was separated from the remaining single-stranded (ss) cDNA by passing the mixture through a hydroxylapatite column. The ss-cDNA was amplified by 9 PCR cycles. The cDNA library in the size range 500–700 bp was eluted from a preparative agarose gel. Emulsion PCR and sequencing were conducted according to standard protocols of Roche and the normalized

cDNA library was sequenced in 1/2-plate run of GS FLX Titanium. Library preparations and their sequencing were carried out by Eurofins Genomics.

#### *2.1.4. Preparation and assembly of contigs for SNP searches*

Using the transcriptome sequence data obtained as described above, contigs were prepared under a subcontract to Eurofins Genomics. Briefly, the procedure comprised of sequence clustering and assembly for each of the varieties, based on the nucleotide sequences of the DNA fragments. De novo assembling from the unique single-read data was performed by MIRA Assembler Version 2.9.45 x 1 (for sequence assembly; Rheinfelden, Germany). To search for SNPs, contig and singlet data obtained in one of the hop varieties served as a reference for the single-read data obtained from the other two varieties. Specifically, with the nucleotide sequences of the contigs and singlets of variety C being used as reference sequences, single reads of varieties A and B were each applied and mapped to the reference sequences according to whether they shared a common portion. Further, contigs constituted by the mapped single reads were identified from the assembling information deployed on the analysis software. The reference sequences as well as the contigs and/or singlets of the other varieties were aligned to search for SNPs using bioinformatics analysis. Average reads per contig were 7 and 6 in variety A and B, respectively. The detected SNPs were reproducibly present in each variety and were therefore not the artifacts of error. Further, the nucleotide sequences of the contigs and singlets of variety B were used as reference sequences, and single reads of varieties A and C were compared and mapped to the reference sequences to search for SNPs in the same way, as mentioned above. A similar exercise was performed with the nucleotide sequences of the contigs and singlets of variety A being used as reference sequences, and the single reads of varieties B and C were applied and mapped to the reference sequences to search for SNPs.

The NGS performed in the 3 varieties generated a total of 589K to 638K reads with the total number of bases without keys, tags, and bad-quality bases being 191 to 227 Mb and the average read length without keys, tags, and bad-quality bases being 299 to 367. These values were comparable to the equipment spec (Table 1). Numbers of contigs (part of cDNAs) assembled in each variety were 42K to 45K. Among these contigs, there were about 4500–6700 contigs with a length of 1000 bp or more (Table 2).

#### *2.1.5. Evaluation of SNP detection by NGS RNA method*

SNPs were searched in the contigs by comparing among the 3 varieties. As a result, 10.4K to 19.3K SNPs were obtained, as shown in Table 2. The numbers of SNPs were almost compatible with the expected numbers, 13.5K to 30K.

#### *2.1.6. Results for SNP analytical regions*

To call variants, mapping analysis was performed among the three hop varieties, which were combined with the contigs (as reference sequences) and single reads of each other. They were mapped and called by GS Mapper Software (Roche Applied Science, Penzberg, Germany).

Large run results	Variety A	Variety B	Variety C	Equipment spec
Total number of reads	618207	588711	638443	450–650 K
Total number of bases without keys, tags, and bad-quality bases	227199992	210506914	191071744	180–280 Mb
Average read length without keys, tags, and bad-quality bases (bp)	367	357	299	350–450

**Table 1.** NGS run results for 3 hop varieties.

		Variety A	Variety B	Variety C
Assembly results	Number assembled	546411	516494	549226
	Number too short	14436	16175	25343
Sum of large contigs <sup>a</sup>	Total number of reads	290726	274772	232407
	Number of large contigs	6725	6157	4563
	Total number of bases	10096155	9165669	6562275
Sum of all contigs	Total number of reads	546411	516494	549226
	Number of all contigs	43638	42395	45432
	Total number of bases	29918028	28215799	26612919
Total number of SNPs		10472	14408	19330

<sup>a</sup>Contigs  $\geq 1000$  bp.

**Table 2.** NGS results of assembly, contigs, and SNPs.

Numbers and types of SNPs per contig or singlet were obtained. For example, if 16 single reads were selected as candidate DNA fragments containing SNPs, in some cases, all of them had the same nucleotide and were different from the reference, whereas in other cases, some of them had same nucleotide and the others had a nucleotide same as that of the reference. We called the former case “HOMO,” in which identification was thought to be done more easily than in the latter case, designated as “HETERO.”

We selected the contigs that had more “HOMO” SNPs per single contig. For example, a contig (A1 c1675) containing 6 different SNPs was selected when the mapping was performed in the A1 region contig as a reference with single reads of variety B. This contig (A1 c1675) had 4 SNPs when single reads of variety C were used. In the same way, the other SNP-rich regions were selected. Thus, 4 SNP-rich regions, A1, B1, C1, and A1-2, were obtained.

The results are shown in the Table 3.



Reference/ variety supplied	Total number of contigs	Total number of SNPs detected	Number of SNPs homozygous in 2 varieties other than reference (per contig)	Contig size (bp)	Analysis region
Variety A	43638	10472	6	1.4k 719	A1 A1-2
Variety B	42395	14408	9	1.0k	B1
Variety C	45432	19330	24	2.8k	C1

**Table 3.** Results for SNP analytical regions.

Primers were designed for each region using DNASIS Pro software (Hitachi Software Engineering Co., Ltd.; Tokyo, Japan), and PCR amplifications of the four regions were performed for further confirmation of the analysis regions.

**2.2. Application of NGS-SNP genotypes to identify hop varieties by Sanger sequencing**

*2.2.1. DNA extraction from hop varieties*

DNA was extracted from the pellets or dried cones of the three hop varieties (A, B, and C), used in the transcriptome analysis described above. DNA extraction by the cetyltrimethylammonium bromide (CTAB; Sigma-Aldrich Co. LLC., St. Louis, USA) method was carried out as follows. The CTAB solution comprised of 2% (w/v) CTAB, 100 mM Tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 1.4 M NaCl, and 1% (w/v) polyvinylpyrrolidone (PVP). About 10–50 g of hop pellets or 1 g of dried cones were ground in a mortar with or without liquid nitrogen. Next, 650 µL of the CTAB solution and 2 µL 1 mg/mL RNaseA solution were added to the ground material in a 1.5 mL microtube and stirred well. This microtube containing the sample and the CTAB solution was submerged in a constant temperature water bath held at 65°C such that the content of the tube was completely underwater; the mixture was incubated for 1 h to disrupt hop cells. An equal volume (650 µL) of chloroform/isoamyl alcohol (24:1; Sigma-Aldrich Co. LLC.) was subsequently added, and the mixture was manually mixed by inversion for 3 min. The mixture was centrifuged in a Beckman’s Allegra 21R centrifuge (with F2402H rotor) at 15,000 rpm (ca. 15,000×g) for 1–5 min and thereby fractionated into an organic solvent layer (lower) and an aqueous layer (upper). The aqueous layer (ca. 400 µL) was removed to a new tube. After the addition of an equal volume of isopropyl alcohol (Wako Pure Chemical Industries, Ltd.), the mixture was mixed by inversion and then centrifuged in the same way as described above. The supernatant was discarded and the remaining sediment was rinsed with about 500 µL of 70% ethanol (Wako Pure Chemical Industries, Ltd.) and dried in an MV-100 MicroVac Mini Vacuum-Centrifugal Evaporator (TOMY Seiko Co., Ltd.) for about 5 min. The residue was dissolved in 20–50 µL of TE buffer (pH 8.0, Nippon Gene Co., Ltd.) to obtain the DNA sample.

The extracted DNAs were subjected to amplification of the respective regions A1, B1, C1, and A1-2 using the primer sets shown in Table 4. The primers were designed using the DNASIS Pro software (Hitachi Software Engineering Co., Ltd.).

Primer	Nucleotide sequence	Length of amplified product (bp)
A1_3L	TAAGGTGTTGGGAGGGTTGA	651
A1_3R	CCACCAATAACAGGCTCCAC	
B1_1L	CAGACTTGTGGCTGTCAAAAA	729
B1_1R	CTTCTCCTTCGAACCTGTCG	
C1_1L	CGGCGTTTTTCAATTTTCAT	646
C1_1R	GTGATGACTCGGGCTTCAGT	
A1_2_1L	GAAATCTGCTTKGAGAAACCTGG	ca 1500
A1_2_1R	GCAGGTATCTTTGTAGGTACATC	
A1-2-M_F	ATTTTGTCTATGCCTGGCA	507
A1-2-M_R	ATTAGACCAGCACCAGTATG	

**Table 4.** PCR primers used in the study.

PCR was performed in a Veriti 96 Well Thermal Cycler (Life Technologies Japan Ltd.) with PerfectShot Ex Taq (Loading dye mix, Takara Bio Inc.), as described in its manual. The temperature during the 30 PCR cycles was 98°C for 10 s, 60°C or 50°C for 30 s, and 72°C for 60 s.

#### 2.2.2. Confirmation of PCR amplification

Five microliters each of the PCR product obtained, as described above, was subjected to agarose gel electrophoresis, performed using a mini gel electrophoresis system, Mupid (gel: 3% NuSieve 3:1 Agarose; FMC BioProducts or Cambrex Bio Science Rockland). The electrophoresis was carried out at 100 V for 30 min, and the amplification products were visualized under ultraviolet light illumination (Printgraph, Atto Corp.) after staining with ethidium bromide (2 µg/mL) for about 40 min. The presence of the amplified DNA fragment of the intended size was ascertained for each DNA sample. The sizes of DNA fragments amplified using the respective primer sets are described in Table 4.

The purification of the resulting PCR products was performed using the QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA, USA) according to the protocol recommended by the manufacturer, as described below. Five volumes of PBI buffer (225 µL) was added to the PCR product (45 µL) and mixed. The resulting solution was placed in the QIA quickspin column and centrifuged at 13,000 rpm (ca. 11,000×g) for 1 min. The supernatant was discarded, 750 µL of PE buffer was added, and the mixture was centrifuged at 13,000 rpm for 1 min. The supernatant was again discarded and the remaining solution was centrifuged at 14,000 rpm (ca. 13,000×g) for 1 min. The supernatant was discarded, the column was transferred to a new 1.5 mL Eppendorf tube, and 30 µL of EB buffer was added. After being left to stand for 1 min, the solution was centrifuged at 13,000 rpm for 2 min to elute the purified PCR product.



2.2.3. Confirmation of SNP markers by Sanger DNA sequencing

Cycle sequencing was performed using the BigDye terminator v1.1 cycle sequencing kit (Life Technologies Japan Ltd.), according to the protocol prescribed by the vendor. The purification of the product as the template DNA for sequencing was performed by the Centri-Sep Spin Column (Life Technologies Japan Ltd.) according to the manufacturer’s protocol. DNA sequencing was performed on an ABI PRISM 310 genetic analyzer (Life Technologies Japan Ltd.). Sequence data obtained from the 5’ and 3’ ends were checked, and the correct base sequence was determined. The nucleotide sequences determined from the tested varieties were aligned in each analysis region by using ClustalW (DDBJ; DNA Data Bank of Japan), which is a popular multiple sequence alignment program for DNA.

Thus, the amplified regions were confirmed. Comparison was also made with the data obtained using the next-generation sequencer, to confirm that the identification of the three varieties A, B, and C was possible. The size of the amplified products from the regions, A1, B1, C1, and A1-2, were 651, 729, 640/646, and ca. 1500 bp, respectively.

The PCR product from the B1 region was bigger in size than expected, and on close scrutiny, it was found that this DNA region included a 111-bp insertion sequence, which might be that of an intron.

For reference, the two different sizes of the amplicons obtained from the C1 region represented the size with or without the 6-nucleotide insertion in this region.

Since the fragment of region A1-2 had a length of about 1,500 bp, the sequencing data obtained for this region consisted of sequences from two regions: an analytical region, A1-2-L of 538 bp from the 5’ end (primer L), and an analytical region A1-2-R of 516 bp from the 3’ end (primer R).

It was determined that the analysis of regions A1, B1, C1, and A1-2 (and the SNPs contained, therein) was also applicable to the identification of other varieties in addition to the three mentioned above. Thus, analysis of a number of varieties was carried out.

Name	No. of SNPs <sup>a</sup>	Size (bp)
A1	24	541
B1	14	645
C1	43	559 <sup>a</sup>
A1-2-L	21	538
A1-2-R	67	516
A1-2-R-2	59	542

<sup>a</sup>Including indel (insertion/deletion) sequence.

**Table 5.** Properties of SNP-rich analytical regions.

2.2.4. Identification markers for 21 hop varieties

Twenty-one varieties of hops were selected for identification based on the nucleotide sequence in region A1.

Nucleotide sequences were aligned to determine the consensus sequence in region A1. The nucleotide sequences determined in the varieties were evaluated for 24 SNPs in region A1 corresponding to the nucleotide positions 74, 77, 87, 103, 116, 118, 121, 125, 134, 135, 148, 192, 195, 197, 199, 203, 204, 226, 230, 235, 306, 316, 330, and 532. These 21 hop varieties showed 12 types (types 1 to 12) of combinations of 24 SNPs in region A1. The SNP positions are depicted in red in Figure 1.

A1 REGION

TGGTGGTGCA	AAAAAGTCAG	GATTTTGATG	GTGGGGAGTC	AGAGTTTAGG	GCCAGTGAGG	60
AGGTAGAAGT	AGAWGAYGGT	AAAATTKTGG	ATGGTGGGAA	TGRTAAAGAT	AACTCWGYCA	120
DTTWWAAGA	GAAKRATGAG	AAGTTGGYCG	AAGAAGATGG	GGTGAGTCTT	GGCGGAGATG	180
AGTCAGTGGT	GRAAKCWGYG	CARSTTAATG	TCCCAGCTTC	AAGAGMAGCY	GATGYTGGAG	240
TTTCAGAAGA	ACTTGATGAA	GCTGAAATTA	GAGGTGTAGA	GCCTCCAGGA	GGCGGGAACC	300
TTGGCYCTGG	TTCTGMAGAA	TTTGTGGGCW	CAAAGTTGAT	GCCTACAGAT	TCTGAATCTG	360
ATGGAAATGT	TGTGGGGTCT	GTTACTGGTG	GTCTGATGA	AGTTGATACC	AAGCACGTAT	420
CAGCAGGAGA	AGATGGTGGG	CTAAAAGCTA	ATTCTGAAGT	TCATCAGAGC	GGCCCCGTGG	480
TCGAGAAAGG	TGCTGATAAT	GAAAAAGTTC	TGTCAGGTGA	TGGAGTTGGT	CSAAAGTTGA	540
T						541

**Figure 1.** Nucleotide sequence of region A1. Each nucleotide is shown as per the IUPAC definition. Gaps are inserted after every 10 letters to increase clarity. Number at the right in every line denotes the number of nucleotides up to that position. Red letters represent the SNP positions.

Nucleotide sequences determined in the 21 varieties were aligned to determine the consensus sequence in region B1. These sequences were evaluated for 14 SNPs in region B1, corresponding to the nucleotide positions 178, 204, 227, 234, 245, 246, 247, 248, 370, 426, 439, 547, 562, and 624. The 21 hop varieties to be identified showed 9 types (types a to i) of combinations of 14 SNPs in region B1. The SNP positions are depicted in red in Figure 2.

Similarly, 43 SNPs were found in region C1 (6 of these 43 SNPs, at positions 129–134, constituted an indel portion) corresponding to the nucleotide positions 3, 13, 17, 76, 77, 87, 88, 93, 129, 130, 131, 132, 133, 134, 136, 138, 163, 165, 245, 254, 313, 321, 331, 356, 373, 375, 376, 380, 396, 398, 399, 421, 435, 438, 460, 474, 475, 476, 477, 480, 481, 500, and 547. The 21 hop varieties to be identified showed 5 types (types i to v) of combinations of 43 SNPs or an indel portion in region C1. The SNP positions are depicted in red in Figure 3.

The analyses of 24 SNPs in region A1, 14 SNPs in region B1, and 43 SNPs or an indel portion in region C1 made it possible to classify 16 of the 21 varieties, excluding Perle, Northern Brewer, Premiant, Zeus, and Summit, into distinct types, respectively. For identifying Perle, Northern

**B1 REGION**

CATTCCTCAG	GATCTAAAGC	GAATTATTAA	GAAATACATT	CAAATGAGAT	TGCAAGATGG	60
AAAAGATGTG	GACATCGACC	ATATTTTGCA	TATTCTTCCT	TTGTCACAGG	GAATGCTAAT	120
TAAGAAACAC	ATGTGTTTCC	CTATGTAAAA	AAAAGTAAGT	TATCTTTTAT	AGTATATRTA	180
TGTCATTATT	AAGTTTGATC	TTTAATTGGA	GCTATAACAT	AGTATTRTTA	TTTRATTGAT	240
TGATTGATTA	TATGATTAAT	TATAGGTTCC	ATTATTTCAA	AACACGGATG	AACATGTTTA	300
CGAAACAATA	TGCAAATATC	TGAAACCAGT	CACATATTCA	GAGAAAAGTT	ATATCATTCC	360
AAAAGGAGAR	CCACTTGATA	TGATTCTCTT	CATCACACAA	GGTGTGTGT	GGGCATTGG	420
AAATGMTACT	ACTCCCATRA	GTCGACTTCA	AAAAGGTGAT	TTCTATGGAA	ATGAACTCAT	480
AGAATGGCAA	TTAAAGTCAA	CATCCATTGA	TGAGTTTCCT	ATTCGGTTG	CTAATCTTAA	540
ATCCCAAYACC	AATGTTGAAG	CYCTTGCTCT	TATGGCCAAT	GACTTGGAAC	ATGTAGTCTC	600
CAATTGCTGG	TTCAAGTTCT	CCCBACCTG	CTCATCTATG	CCTGA		645

\* Variety Galena shows 2 genotypes, one of which contains deletion sequences from 245th to 248th nucleotide.

Figure 2. Nucleotide sequence of region B1. Please consult the legend of Figure 1 for details.

**C1 REGION**

TTYGACGTGA	TCSAGTKGAG	TGGTGGTATA	TACGTTCCAT	ATTATAAAGA	TTACATTGTT	60
TTAATGGATA	TACCARRGAA	GGAGAAWKTA	AAMCTCTCCA	TCGCACTGCA	AGCCAACCCC	120
GATGATACSA	TGACTYTSTA	CACTGATGCC	ATCTTGAATG	GCRTSGAAAT	CTTCAAATTA	180
AGCGACTCCA	ACGGAAATCT	CGCCGGTCTG	AATCTTGACC	TTGTTCCGCA	TGCGAAGCCA	240
CCAARCCAAC	CATYATGGAA	GTTGAAAAAT	CATCAATCAG	AGATTATTGG	CATTTCCATT	300
GGTGTAGTTT	GCWGC GTTGT	YTTGCTCTGC	RTAATTGGAT	TCCTATTTCT	CCGGCRAGGT	360
AGAAAAGTCA	AARGMWCAAK	GCTCATTTCC	TTGTCHASWA	CCAGGTCAAG	CAAGACCGGA	420
RAATCATCAT	TGCCRTTKGA	TCGGTGTCTG	TACTTTTCAY	TGGCTGAGAT	CATWRMYGCM	480
RCAACAACACT	TCGAAGATAK	TTTCATTATT	GGGTTGGAG	GATTCGGAAA	CGTGTATAAA	540
GGGCACRTCG	ACAATGGGA					559

\* Underlined part contains deletion sequence.

Figure 3. Nucleotide sequence of region C1. Please consult the legend of Figure 1 for details.

Brewer, Premiant, Zeus, and Summit, the nucleotide sequences in region A1-2 of these varieties were further analyzed. These nucleotide sequences were aligned to determine the consensus sequence in region A1-2-L. The nucleotide sequences were evaluated for 21 SNPs in region A1-2-L (10 of these 21 SNPs constitute indel portions) corresponding to the nucleotide positions 34, 101, 118, 124, 164, 168, 171, 186, 187, 188, 189, 190, 191, 192, 193, 194, 393, 398, 399, 459, and 502 (positions 186–194 and position 399 each constituted an indel portion). The SNP positions are highlighted in red in Figure 4.

These nucleotide sequences were aligned to determine the consensus sequence in region A1-2-R. The nucleotide sequences determined in the varieties were evaluated for 67 SNPs in region A1-2-R (4 of these 67 SNPs, at positions 57–59 and 65, constituted an indel portion) corresponding to the nucleotide positions 1, 2, 3, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 25, 26, 27, 28, 29, 30, 31, 33, 35, 36, 37, 38, 41, 42, 43, 44, 46, 47, 48, 50, 51, 56, 57, 58, 59, 63, 65, 68, 72, 78, 79, 84, 86, 88, 90, 92, 118, 153, 154, 191, 205, 206, 226, 228, 233, 254, 289, 315, 350, 392, and 405. The SNP positions are highlighted in red in Figure 5.

Zeus, Summit, and Premiant were subjected to further analysis for region A1-2-R2. In region A1-2-R2 (34 of these 59 SNPs constituted an indel portion), the SNPs corresponded to the nucleotide positions 3, 8, 19, 20, 27, 28, 40, 41, 46, 47, 57, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 90, 112, 116, 118, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 140, 156, 178, 191, 192, 244, 266, 271, 377, and 430 (positions 64–85, and 120–131 constituted the indel portions). The SNP positions are depicted in red in Figure 6.

Using nucleotides at the SNP positions in the fragments of A1-2-L (sequenced from 5' primer) and A1-2-R/R2 (sequenced from 3' primer), Perle and Northern Brewer as well as Premiant, Zeus, and Summit were successfully identified, when 21 and 67 SNPs were found between Perle and Northern Brewer, and 59 SNPs were found among Premiant, Zeus, and Summit.

**A1-2-L REGION (PERLE AND NORTHERN BREWER)**

CAGCTGAGCT	TGAAAGTTAT	GGCTATCGCC	GCC	RCCATT	CGGGTCCTCT	GACAACTCCG	60		
TCTCGGCGGC	GCTCAGTCAC	CTCCTTCAGC	TACTCTCCA	W	TCTCATT	TCTTACGRGA	120		
CAAY	TGAGTT	TTCTCAGAGC	AAGTCGCC	CG	TGTATTGGCT	TGT	YAGCYAC	WGCTTCTTCT	180
TCTTC	<u>TTCTT</u>	<u>CTTC</u>	CATTTC	CATGAGTCTC	GAAGCCGCCG	AGAAGGCCTC	GCCGGTGTCTG	240	
TTCTTGGATC	GCAGGGAGAC	CGGTTTTCTT	CATTTTGTCA	AGTACCACGG	CGTCGGCAAT	300			
GACTTCATCT	TGGTAAACTC	GAAAATGGAG	CTAGTTTTTC	TCTATAGAGG	AAAAGAAAGA	360			
ATTTTTTTCT	TCTAAATTTA	TTGAGTAGTC	Y	TTTTTT	KIA	TTGATGTAGG	TTGATAATAG	420	
GGATTCTTCA	GAGCCTAGGA	TTACTCCCGA	GCAAGCGGYG	AAGCTCTGTG	ATCGGAACCTT	480			
TGGAATTGGA	GCTGATGGGG	T	SATTTTTGC	TATGCCTGGC	ATCAATGGCA	CTGACTAC	538		

\* Underlined part contains deletion sequence.

Figure 4. Nucleotide sequence of region A1-2-L. Please consult the legend of Figure 1 for details.

2.2.5. Characterization of SNP-rich regions and identification of hop varieties

In every SNP-rich region, consensus sequence and SNP positions were detected by the alignment analysis. Nucleotide polymorphisms in each variety were evaluated at the SNPs positions in all the regions. Within the 21 studied hop varieties, 24 SNPs were identified in the A1 region and 14 SNPs (including indel) were observed in the B1 region. *H. lupulus*, with normally a diploid chromosome, had heterozygous or homozygous SNPs in each DNA region. For example, in the A1 region, European varieties had homozygous SNPs at 77th (C/T) and

A1-2-R REGION (PERLE AND NORTHERN BREWER)

```
RSWASGAWRS YYYRSSYGAY WTATYKWYRW RCYTRWRMAT KSSWAKRWAW RCACASGTAA 60
CAYATCAYAC AYACACAYRC ACAYAYAYAY AYATATACAT ATAATGTAAA TGTGTATWCT 120
CTATTGTGAT GATTTGCTTC TGAAACATGA TAWWAGAAAC AATAAAAAGA ACAAAGAAGT 180
ATTGGTTATT RGGGAAGACCT CTTTYRGATT TCTTTGACAA TGCCTRTGTG TTYTATTGAT 240
GTAGCTTTAC TGTRCATACT GGTGCTGGTC TAATTATTCC AGAAATACWA GATGATGGTC 300
AGGTATAAAC TTCTMTACTT GCAAATTACC TAAATCTATT GGGTTCATAK GCTTCAAAAG 360
CAACATTCAT AAATGATTGC ATTTTGCTAC ARTTATGGCA GTTTKGCTTC TAATCTGGAT 420
TTGTTGAAGT TGCCAAACTA CTGGATAGCG ATTTTGTAGT ATTCAATTTA GCAGGGGTGC 480
TGTTTTGATT TCAGGTCAAA GTTGATATGG GTGAAA 516
```

\* Underlined part contains deletion sequence.

Figure 5. Nucleotide sequence of region A1-2-R. Please consult the legend of Figure 1 for details.

A1-2-R2 REGION (PREMIANT, ZEUS, AND SUMMIT)

```
AGYACGAKGC TTTGCCCGMK TTATTGYYGA GCTTGAGAAW WTGCAMRGAA AGCACASGTA 60
ACAYATMTAT AGACMTATAC ATGCATATAY ACACACACAC ACACACACAC AYACAYAYAC 120
ACACACACAC ATATACATAK AATGTAAATG TTGATWCTCT ATTGTGATGA TTTGCTTYTG 180
AAACATGATA WWAGAAACAA TAAAAAGAAC AAAGAAGTAT TGTTATTAG GAAGACCTCT 240
TTTRGATTC TTTGACAATG CCTATRTGTT YTATTGATGT AGCTTTACTG TGCATACTGG 300
TGCTGGTCTA ATTATTCCAG AAATACAAGA TGATGGTCAG GTATAAACTT CTATACTTGC 360
AAATTACCTA AATCTAWTGG GTTCATAGGC TTCAAAGCA ACATTCATAA ATGATTGCAT 420
TTTGCTACAR TTATGGCAGT TTGGCTTCTA ATCTGGATTT GTTGAAGTTG CCAAACACT 480
GGATAGCGAT TTTTGAGTAT TCAATTTAGC AGGGGTGCTG TTTTGATTTC AGGTCAAAGT 540
GA 542
```

\* Underlined part contains deletion sequence.

Figure 6. Nucleotide sequence of region A1-2-R2. Please consult the legend of Figure 1 for details.

103th (A/G) positions. Such homozygous SNPs can be analyzed more easily than the heterozygous SNPs when mixed with another variety, having different nucleotides at the SNP positions. In a case, where the variety Saaz was mixed with the variety Sládek, position 77 SNPs were T and C, respectively, in the A1 region (data not shown), and these nucleotides could be easily distinguished on their electropherograms. On the other hand, position 74 SNPs were W (A and T) and A, respectively, in the same region. In this case, it was difficult to recognize whether the variety Saaz was mixed with the variety Sládek or not. Each variety had a specific combination of SNPs as markers, and 12 and 9 DNA types in the A1 and B1 regions, respectively, were identified among the 21 varieties. Indel was also found in the C1 region. Nucleotides at the SNP positions including indel were observed. Forty-four SNPs were found



in this region, and 5 DNA types were identified among the 21 varieties. It was revealed that Galena had 2 DNA types in the B1 region and Cascade and Super Galena had 2 DNA types in the C1 region.

Additionally, the A1-2 region was searched because there was no difference in the combination of the SNPs in A1, B1, and C1 regions between Perle and Northern Brewer, among Premiant, Zeus, and Summit. Nucleotides at the SNP positions in the fragments of A1-2-L (sequenced from 5' primer) and A1-2-R/R2 (sequenced from 3' primer) were shown, when 21 and 67 SNPs were found between Perle and Northern Brewer, and 59 SNPs were found among Premiant, Zeus, and Summit. Perle and Northern Brewer as well as Premiant, Zeus, and Summit were successfully identified.

For confirmation, we tried to distinguish between Columbus and Tomahawk, considered genetically identical [22–24], by analyzing the 4 regions and, additionally, the middle area of the A1-2 region, with newly designed primers, A1-2-M\_F and A1-2-M\_R (Table 4). No differences in the nucleotide sequences were detected between them. Therefore, we confirmed that these 2 varieties are indeed identical and could be considered a single variety.

Consequently, 21 hop varieties were successfully identified by a combination of these genotypes of SNPs, with differences in the 4 SNP-rich regions. The summary results are shown in Table 6.

Origin	Variety	Diplotype				
		A1	B1	C1	A1-2	A1+B1+C1+ (A1-2)
Czech Republic	Saaz	1	a	i		A
	Sládek	2	b	i		B
	Premiant	2	c	i	$\beta 1$	C
Germany	Tradition	3	d	ii		D
	Spalter	4	e	i		E
	Spalter Select	2	f	ii		F
	Perle	2	d	i	$\alpha 1$	G
	Tettnang	5	g	i		H
	Brewer's Gold	7	c	i		I
	Northern Brewer	2	d	i	$\alpha 2$	J
	Magnum	2	d	iii		K
	Herkules	6	d	ii		L
	German Nugget	9	c	i		M
	Taurus	8	d	i		N
	Cascade	2	d	iv		O1



Origin	Variety	Diplotype				
		A1	B1	C1	A1-2	A1+B1+C1+ (A1-2)
				i		O2
	Zeus	2	c	i	β2	P
	Summit	2	c	i	β3	Q
	Galena	2	h	i		R1
			i			R2
	Super Galena	11	c	i		S1
				v		S2
	Nugget	10	c	i		T
	Columbus/Tomahawk	12	c	i	γ	U

**Table 6.** Successful identification of 21 hop varieties

2.2.6. *Comparison of DNA samples prepared from hop pellets and cones*

Comparison was made between the results obtained in the two cases where DNA samples were extracted from either the hop pellets or the dried hop cones, as described in Section 2.2. It was confirmed that DNA extraction and sequencing were possible with both types of DNA samples. Also, analyses using DNA samples of both types yielded the same results. This result further demonstrates that inspection at a processing step (e.g., inspection for contamination at a pelletization step) is technically possible.

2.2.7. *Comparison between three Saaz clones*

About 1 g each of dried cones of three Saaz clones (Osvald’s clones 31, 72, and 114) was ground in a mortar in the presence of liquid nitrogen, and the DNA was extracted from about 50 mg each of the ground materials by the CTAB method described in Section 2.2. Each of the extracted DNA sample was subjected to amplification of the DNA fragments from regions A1, B1, and C1 and followed by sequencing of the amplified fragments, according to the procedures as described above. For each region, the nucleotide sequences of the three clones were aligned for comparison with each other.

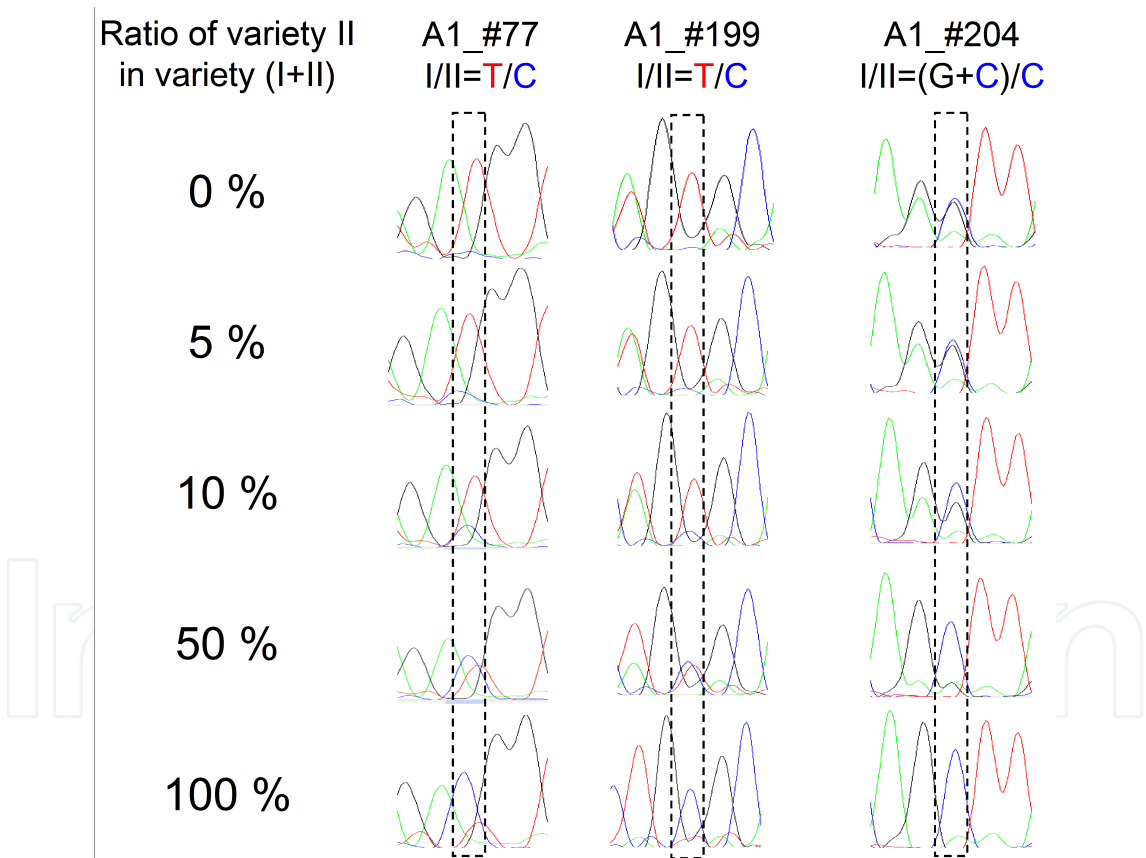
The results obtained confirmed that there was no difference among the three Saaz clones in terms of the nucleotide sequences in the analyzed regions. In other words, it was found that any of Saaz clones could be identified through the above-described analysis using the inventive variety identification regions. These results demonstrated that the analysis using regions A1, B1, and C1 as the variety identification regions required no determination by clone.

2.2.8. *Identification of variety mixtures using the newly identified SNP markers*

Because it is important to detect contamination of other varieties, we developed a method for detecting such contamination by using the SNPs markers. The mixed samples of Saaz and

Premiant hops were analyzed (for region A1). To prepare different samples, pellets of these varieties were ground as described above followed by mixing of the ground materials in the relative proportions by weight as mentioned in Table 7.

A representative electropherogram, containing the SNP position 77 in the A1 region (A1\_#77), is shown in Figure 7. Peaks in the electropherogram represent fluorescence intensity of 4 different nucleotides at each DNA positions, which are depicted in different colors; A, C, G, and T, are shown in green, blue, black, and red, respectively. Variety I is homozygous (TT) at position 77, and variety II is homozygous (CC) at the same position. In the case of mixing of varieties in 50% proportion, overlap of 2 peaks of T and C was observed at this position, reflecting contamination. Figure 7 also shows peaks in the same region when variety II was mixed at 5 and 10%. Peaks in blue at this position, derived from variety II, were detected even at 5% contamination. It was the same for A1\_#199 and A1\_#204, in which variety I was heterozygous (GC) and variety II was homozygous (CC), so it was not easy to recognize the contamination by peak height of electropherogram.



**Figure 7.** Representative electropherogram containing the SNP positions 77, 199, and 204 in the A1 region (A1\_#77, A1\_#199, and A1\_#204): “T” is represented in red color, and “C” is in blue color.

We estimated more SNPs, with a focus on homozygous SNPs, and calculated an average of the mixing rate in variety II. As a result, we obtained a closer estimate of the mixing rate. We postulate that contamination at levels lower than 5% could be estimated; however, the

acceptable contamination level for us might be 5% with reference to the genetically modified organism contamination in Japanese food. Thus, the values below 5% have a slight importance.

Proportion of Premiant as a contaminant	1 0%	2 5%	3 10%	4 50%	5 100%
Saaz	30 mg	28.5 mg	27 mg	15 mg	0 mg
Premiant	0 mg	1.5 mg	3 mg	15 mg	30 mg
Detection of Premiant	–	○	○	○	○

**Table 7.** Preparation of the 2-variety mixtures and results obtained for each sample.

Each sample was prepared by mixing pellets of two varieties in specified proportions by weight.

3. Conclusions and future directions

The results obtained in the present study suggest that the SNP-combination method has high reliability as it utilizes the sequence data. Computational handling of SNP data can be easily carried out so that analysis among many varieties could be easily performed with digital information. Other SNPs identified in the future could be used as additional identification markers. Therefore, using this method, enhancement of accuracy and repeatability in the variety identification could be accomplished in a relatively simple manner.

As SNP-rich DNA regions are only 0.6–1.5 kb in size, they are much less likely to be damaged during the processing of hop products. However, upon degradation and fragmentation of the SNP-rich regions, newly designed primers for the shorter fragments generated might be useful for the amplification of the fragments and may contain several useful SNPs which can be used as identification markers.

The SNP-based information could be also used for quantitative determination of the ratio of variety mixture. In the future, more accurate results will be obtained by using quantitative real-time PCR and/or NGS, which may provide a huge amount of sequence data and increase reliability.

Acknowledgements

I appreciate the technical support and advice offered by Dr. Toshiya Yamamoto (National Institute of Fruit Tree Science, Tsukuba, Japan), Dr. Jun Onodera (Operon Biotechnologies, K.K., Eurofins Genomics, Japan), Dr. Eiichiro Ono (Research Institute, Suntory Global Innovation Center Limited, Japan), and my collaborators.

## Author details

Hiromasa Yamauchi\*

Address all correspondence to: Hiromasa\_Yamauchi@suntory.co.jp

Suntory Business Expert Ltd, Kawasaki, Japan

## References

- [1] Araki S, Tsuchiya Y, Masachika T, Tamaki T, Shinotsuka K. Identification of hop cultivars by DNA marker analysis. *J. Am. Soc. Brew. Chem.* 1998;56:93–98. DOI: 10.1094/asbcj-56-0093.
- [2] Murakami A. The practical application of PCR for the verification of hop variety. *MBAA Technical Quarterly*. 1998;35:185–188.
- [3] Patzak J. Assessment of somaclonal variability in hop (*Humulus lupulus* L.) *in vitro* meristem cultures and clones by molecular methods. *Euphytica*. 2003;131:343–350. DOI: 10.1023/A:1024096401424.
- [4] Patzak J, Nesvadba V, Henychova A, et al. Assessment of the genetic diversity of wild hops (*Humulus lupulus* L.) in Europe using chemical and molecular analyses. *Bi-ochem. Syst. Ecol.* 2010;38:136–145. DOI: 10.1016/j.bse.2009.12.023.
- [5] Sustar-Vozlic J, Javornik B. Genetic relationships in cultivars of hop, *Humulus lupulus* L., determined by RAPD analysis. *Plant Breed.* 1999;118:175–181. DOI: 10.1046/j.1439-0523.1999.118002175.x.
- [6] Pillay M, Kenny S. T. Structure and inheritance of ribosomal DNA variants in cultivated and wild hop, *Humulus lupulus* L. *Theor. Appl. Genet.* 1996;93:333–340. DOI: 10.1007/BF00223173.
- [7] Jakse J, Kindlhofer K, Javornik B. Assessment of genetic variation and differentiation of hop genotypes by microsatellite and AFLP markers. *Genome*. 2001;44:773–782. DOI: 10.1139/g01-071.
- [8] Seefelder S, Ehrmaier H, Schweizer G, et al. Genetic diversity and phylogenetic relationships among accessions of hop, *Humulus lupulus*, as determined by amplified fragment length polymorphism fingerprinting compared with pedigree data. *Plant Breed.* 2000;119:257–263. DOI: 10.1046/j.1439-0523.2000.00500.x.
- [9] Andreja Č, Jernej J, Branka J. Identification and differentiation of Hop varieties using simple sequence repeat markers. *J. Am. Soc. Brew. Chem.* 2004;62:1–7. DOI: 10.1094/ASBCJ-62-0001.

- [10] Jakse J, Satovic Z, Javornik B. Microsatellite variability among wild and cultivated hops (*Humulus lupulus* L.). *Genome*. 2004;47:889–899. DOI: 10.1139/g04-054.
- [11] Kanai D, Kirita M, Sakamoto K. Method for judging kind of hop using microsatellite DNA. JP Patent Publication number 2006-34142, 2006-2-9.
- [12] Patzak J. Comparison of RAPD, STS, ISSR and AFLP molecular methods used for assessment of genetic diversity in hop (*Humulus lupulus* L.). *Euphytica*. 2001;121:9–18. DOI: 10.1023/A:1012099123877.
- [13] Patzak J, Nesvadba V, Krofta K, Henychova A, Marzoev A. I, Richards K. Evaluation of genetic variability of wild hops (*Humulus lupulus* L.) in Canada and the Caucasus region by chemical and molecular methods. *Genome*. 2010;53:545–557. DOI: 10.1139/g10-024.
- [14] Syvolap IuM, Zakharova O. O, Kozhukhova N. E, Ihnatova S. O, Prystavs'kyi M. S, Zelenina H. A. Modern biotechnologies in estimation of genetic diversity of Ukrainian varieties of hop (*Humulus lupulus* L.). *Tsitol Genet*. 2010;44:3–12. Cf. DOI: 10.3103/S0095452710050014.
- [15] Murakami A, Darby P, Javornik B, Pais MSS, Seigner E, Lutz A, Svoboda P. Molecular phylogeny of wild Hops, *Humulus lupulus* L. *Heredity*. 2006;97:66–74. DOI: 10.1038/sj.hdy.6800839.
- [16] Vrba L, Matoušek J, Patzak J. New STS molecular markers for assessment of genetic diversity and DNA fingerprinting in hop (*Humulus lupulus* L.). *Genome*. 2007;50:15–25. DOI: 10.1139/g06-128.
- [17] Howard EL, Whittock SP, Jakše J, Carling J, Matthews PD, Probasco G, Henning JA, Darby P, Cerenak A, Javornik B, Kilian A, Koutoulis A. High-throughput genotyping of hop (*Humulus lupulus* L.) utilising diversity arrays technology (DArT). *Theor. Appl. Genet*. 2011;122:1265–1280. DOI: 10.1007/s00122-011-1529-4.
- [18] Castro C. B, Whittock L. D, Whittock S. P, Leggett G, Koutoulis A. DNA sequence and expression variation of hop (*Humulus lupulus*) valerophenone synthase (VPS), a key gene in bitter acid biosynthesis. *An Bot*. 2008;102:265–273. DOI: 10.1093/aob/mcn089.
- [19] Grabowska-Joachimik A, Śliwińska E, Piguła M, Skomra U, Joachimik AJ. Genome size in *Humulus lupulus* L. and *H. japonicus* Siebold and Zucc. (*Cannabaceae*). *Acta Societatis Botanicorum Poloniae*. 2006;75:207–214. DOI: 10.5586/asbp.2006.024.
- [20] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. The sequence of the human genome. *Science* 2001;291:1304–1351. DOI: 10.1126/science.1058040.
- [21] Yamauchi H, Mukouzaka Y, Taniguchi T, Nakashima K, Furukubo S, Harada M. Newly developed SNP-based identification method of hop varieties. *J. Am. Soc. Brew. Chem*. 2014;72:239–245. DOI: 10.1094/ASBCJ-2014-1006-01.

- [22] Hatch L. Brewbase Encyclopedia. [http://members.tripod.com/~Hatch\\_L/bbase-hops.html](http://members.tripod.com/~Hatch_L/bbase-hops.html), (cf. 2013-08-20).
- [23] Hopsteiner. Varieties - United States. <http://www.hopsteiner.com/varieties/united-states.php>, (cf. 2013-08-20).
- [24] Lewis G. K, Zimmermann C. E, Hazenberg H. Hop variety named 'Columbus'. U. S. Patent 10,956. 1999-06-15.



