P. D. Matthews, M. C. Coles, and N. J. Pitra

Next Generation Sequencing for a Plant of Great Tradition: Application of NGS to SNP Detection and Validation in Hops (*Humulus lupulus* L.)

Application of next generation DNA sequencing technology to hops yielded an unprecedented, large number of novel single nucleotide polymorphisms (17, 128 SNPs). The markers were detected and then validated for use in genotyping and control of quality for hops. By using genotyping-by-sequencing (GBS) and a universal network-enabled analysis kit (UNEAK) designed for species with no "reference genome", we generated a set of molecular markers with a genome-wide distribution. Validation of the markers was accomplished by observation of metrics of sequencing quality, by marker behavior in genetic segregation and by application to genetic distance and hierarchical cluster analyses across a set of commonly known cultivars. The SNPs were characterized by average read depth of 3.7 and a call rate across 178 diverse individuals of 0.82. Many SNP alleles segregated with near test cross ratios of 1 : 3 or 3 : 1 and intercross ratios of 0.50 among 103 full-siblings. Erroneous SNPs, with unusually high or low allele segregation ratios were detected at a rate of 4.1 % and could be removed from further analyses. Filtering of SNPs for potentially higher quality was accomplished by selection of call rate thresholds above 0.5, 0.75 and 0.90 or, alternatively, by selection of markers with minimal segregation distortion. Genetic distance matrices and dendrograms for marker subgroups were similar as shown by Mantel's Z-tests and cophenetic correlation coefficients. Bootstrapping generated an exceptionally well-supported tree for genetic relationships among the hop cultivars.

Descriptors: genotyping-by-sequencing, single nucleotide polymorphisms, genetic distance analysis, *Humulus lupulus* L., hop quality

1 Introduction

Hop is a difficult species for genetic dissection and for application of molecular crop improvement methods, such as marker-assisted selection (MAS). In contrast to species with inbred lines; i.e., maize or rice; *Humulus lupulus* L. is dioecious and, thus, an outcrosser. Outcrossing portends a heterozygous species; which makes molecular marker isolation, validation and application more complex. Additionally, the hop genome is large, about 2.9 pg (2.8 GiB) [2]; twice the size of its con-generic species *H. japonicus* Siebold and Zucc. and approximately four times larger than its nearest con-familiar, *Cannabis sativa* L. [43]; suggesting, perhaps, two ancient genome-wide duplications or extensive repetitive element content [3, 16]; both of which complicate molecular marker detection and applications. Furthermore, hops tolerate multiple ploidy levels including tetraploid and triploid types and may exhibit aneuploidy. Chromosomal imbalances have been evidenced in hops by frequently observed sexual transmutation (sterile male flowers on female plants), given that the gender determination system is based on a XY system, where the balance of X to autosomes is sex determinative [26]. Since the hop genome is large and complex, hop is likely to exhibit significant copy number and presence and absence variations [37]. The size and variability of the hop genome make *de-novo* next generation sequence assembly of a draft reference sequence difficult. Without a reference genome large-scale detection, validation and referencing of genome-wide SNPs has been intractable. Although significant transcriptome assemblies are available [27, 44], there are no reports of transcriptome-wide SNP mining as yet. Application of GBS UNEAK TASSEL pipeline to hops has solved the above mentioned problems.

Applications of molecular markers in the hop growing, handling and brewing industries includes many important utilities, including (1) trueness-to-type genotyping for identification of hop propagation material in reference to commonly known cultivars (2) purity assessment of putatively mixed fields or lots of processed raw hop products (3) genetic similarity (nearest neighbor) analysis for guidance in substitution of hop cultivars (4) and development of novel, improved and agronomically sustainable hop varieties by MAS. DNA-based molecular markers have the greatest potential; because, unlike chemical markers, such as essential oils, molecular markers are insulated from sampling and environmental effects and only require very small samples of tissue; such as a single bract, a shoot tip, a plumule or a leaf fragment.

A number of technologies have been applied to increase the number of molecular markers for hops. *Pillay* and *Kenny* were the first to

Authors

Dr. Paul D. Matthews, Mark C. Coles, and Nicholi J. Pitra, Hopsteiner, S.S. Steiner, Inc., New York/USA; corresponding author: pmatthews@hopsteiner.com

publish a study of chloroplast DNA variations in hops in 1994 [18]. Since then, several research groups and an international consortium, the HumDArTium^{PM}, have made substantial progress. *Patzak* [29, 32], *Henning* [12, 14] and *McAdams-Howard* [15, 25] have astutely reviewed the progress in marker development. Molecular markers have been applied to a number of practical uses in breeding, for example, (1) genetic distance analyses among breeding stocks [11, 13, 14] (2) prioritizing germplasm collections [1, 8, 15, 32, 45] and (3) quantitative trait loci (QTL) detection [7, 8, 12, 17, 20, 25]. Recently, DNA polymorphs among candidate genes for valued hop chemical contents and disease resistance genes have been the focus of targeted marker development [6, 22, 30] and a patent [28]. The latest efforts [12, 25] in QTL mapping of traits, have employed a conglomeration of about 300-800 molecular markers of various types. On the recent forefront of marker development are bioinformatics approaches to mine genic SSRs from published transcriptomes (expressed genes) [31, 33]. More than 1000 novel SSRs are available for further development and application in hop [19]. Low-density, genic SSRs will be important complements to genome-wide, high density SNP analysis [36, 40], which are potentiated by our current study.

Massively parallel nucleic acid sequencing technologies, often called "next generation sequencing" (NGS) platforms have enabled rapid development of gene expression measurements and genome-wide SNP detection in hops. Next generation sequencing relies on massively parallel sequencing reactions produced in short reads of 20 to 250 nucleotides length in a microfluidics flowcell monitored by high-density charged couple devices. A single instrument "run" can produce 100–600 billion sequenced nucleotides in about two weeks. Excellent, current reviews of the revolutionary technologies in historical context of sequencing innovations [24], plant DNA finger printing and barcoding [38] and specific applications to SNP development for plants with reference [36], with partial reference [10, 42] and no reference genomes [21] are available. On the Illumina HighSeq 2000™ platform used in this study, 6 "lanes" of about 100 million sequencing reactions each were monitored to detect 100 nucleotides in sequence. For this study, we employed the non-reference approach of *Lu* et al., which uses a network-enabled filtering system to detect SNPs as binary allele character states [21]. Re-creation of the marker set does not require sequence information, and may be accomplished by replication of the specific genomic reduction, sequencing and computing protocol presented.

2 Materials and Methods

A plant collection of 178 distinct genotypes maintained by Hopsteiner, S.S. Steiner, Inc. was employed. All accessions are available from the United States Department of Agriculture, Agricultural Resource Service, Germplasm Resource Information Network (USDA-ARS-GRIN) (www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?19415) or are cultivars available from Hopsteiner (www.hopsteiner.com). Only commonly known cultivars (CKCs) were used, since the purpose of the study is validation of GBS TASSEL UNEAK generated SNPs over common commercial germplasm: no "wild hops" or other species of *Humulus* are included in this study. Additionally, a full-sib family of 103 individuals from a mating of Zenith X DG0117m was included to validate SNP segregation. The triploids Alpha Aroma, Crystal, Green Bullet, Liberty, Mt. Hood, Pacific Gem, Sticklebract, Willamette and Ultra the tetraploid USDA21055 and the remaining diploids are named in Figure 4 (see page 189). Twelve replicates of cultivar Saaz were used to judge repeatability of data production in an independent experiment.

Total DNA (genomic, chloroplast, mitochondrial, and contaminants) was extracted from 50 mg of young, expanding hop leaves with a Qiagen 96-well DNAEasy™. DNA extracts were assayed on 0.7 % argarose gels for amount (more than 1 ug) and median molecular mass (> 30 kb). A subset of DNAs was tested for restriction by *ApeKI* and separation of fragments on a Qiagen Qiaxel™ capillary electrophoresis system.

For each cultivar (N = 178) in the study, reduced representation libraries were prepared according to a published protocol [10]. *ApeKI* was used as the restriction enzyme and sets of 95 samples per lane were single-end sequenced to 100 nucleotides on Illumina HiSeq 2000. UNEAK was used to call SNPs, essential according to Lu [21]. The pipeline was run in LINUX as a set of PERL scripts initiating Java plugins in TASSEL version 3 or 4 [4], with the default parameters settings for each pipeline plugin, except the UMergeTaxaTagCountPlugin –c (minimum count of tag must be present to output) was set to 144, based on the desire for greater than 80% call rate of at least one tag per 178 genotypes. Manipulation of output files was accomplished with minor Python scripts. Common statistical methods and data plotting were implemented in R or Excel.

3 Results and Discussion

3.1 Detection of SNPs

The number, quality and distribution of molecular markers produced by the GBS UNEAK TASSEL pipeline depends on DNA quality, genome size, sequencing density and the natural distribution of restriction sites within the genome. Read depth and call rate are interdependent and useful general characteristics for SNP quality. Distribution of number of sampled sites depends on the restriction enzyme used for genome reduction [10, 21]. For hop, we chose *ApeKI* based on experience with read depth in maize, which has a similar genome size to hop. Being methyl-sensitive *ApeKI* should also preferentially sample genic regions of the hop genome. Because of unknown variability among plant species in repetitive element content and genome structure, the number and coverage of GBS tags is hard to predict [41], and various groups have empirically determined strategies to increase call rate and coverage [10, 35,36]. For example, in maize a theoretical average read depth of 14.6 reads per *ApeKI* SNP was calculated [41]. In practice for hop genomic DNA with *ApeKI*, preparation of libraries by standard GBS protocol [10] and 96-plex per lane of Illumina HiSeq 2000™ yielded an acceptable, average read depth 3.7 reads per SNP and a total of 17, 869 SNPs. Trials with 384-plex per sequencing lane with the same system applied to 12 independent Saaz genotypes yielded a read depth of about 0.89 and 7,529 SNPs. Such lower coverage may provide a more economical approach for some studies.

3.2 Validation of SNPs

SNPs from 96-plex experiments were further characterized by distribution of read depth, call rate and minor allele frequency (MAF). Since we were not attempting to maximize marker number in these experiments, we set a very high constraint on tag count number in UMergeTaxaTagCountPlugin (c = 144), compared to the default value of 5. A tag count of 144 distributed evenly among the 178 plants in our SNP detection routine favors a call rate of over 0.80 for read depths greater than 1. With this exception to otherwise default parameter settings, a posterior average read depth of 3.7 and a median call rate of 0.85 were obtained for an easily manageable number of SNPs (N = 17, 869). Much larger numbers of detected SNPs, by a factor of 10–100 times are possible (data not shown), but not necessary for many studies [9]. Imputation of missing data among low-coverage, low call rate SNPs [36] is not currently feasible in hops. The distributions of read depth (the number of SNPs within a category of read depth for a population) and the distributions of call rate (the number of SNPs with at least one tag within a category of proportion of individual cultivars) are shown in figure 1. The read depth of the minor allele across all SNPs was distributed around the mean of 3.70 with a minimum of 0.84, a median of 2.24 and maximum of 121.23. The mean call rate across all genotypes was 0.82 with a minimum of 0.16, a median of 0.85 and maximum of > 0.99.

Read depth and call rate are related [21]. Following the example of Lu et al., we calculated "coverage" as the mean number of reads per SNP locus per cultivar. A scatter plot of call rate vs. coverage shows call rate increases dramatically with coverage up to a point. Thus given our high-stringency approach to detection, a positively skewed distribution of call rate shown in figure 1 was expected. Figure 2 shows a dramatic increase in coverage at 1–15 reads per locus, which includes our average read depth. Coverage above 20 reads/SNP locus/cultivar has little value for increasing call rate. Therefore, most of the SNPs detected in hops had good coverage.

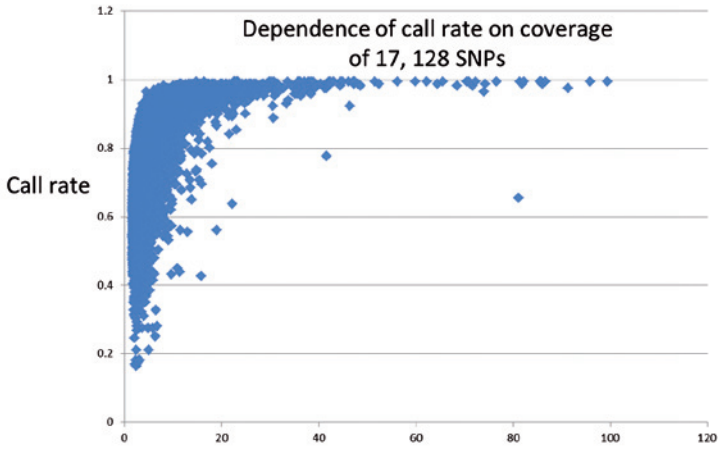

Fig. 2 Scatter plot of call rate and coverage demonstrates the dependence of call rate on read depth. Call rate increases dramatically with coverage and most of the markers provide sufficient call rate

Further validation of marker performance was accomplished by assessing trends in allele frequencies in subpopulations of the cultivars. The SNPs from the admixture of populations were subgrouped into individual families and panels. We also included a consideration of ploidy-level. The association panel was mostly diploids, but one tetraploid and 7 triploid hops were initially included. Exclusion or inclusion of the polyploidy hops did not affect the summary statistics or cluster analyses of SNPs presented here (data not shown). Nevertheless, the polyploids were removed from analyses of allele frequency distribution.

The distribution of the SNPs across categories of minor allele frequencies (MAF) is shown for the diploid association panel and the full-sib family in figure 3 (see next page). In the association panel (Fig. 3A), the distribution of MAF shows increased numbers of SNPs at some frequencies, suggesting population structure. Many of the hops in the panel are related by decent. The histogram for the full-sib family is strongly tri-modal (Fig. 3B) with frequency category optima suggesting Mendelian segregation frequencies (MAFs 0.25, 0.5 and 0.75) of many SNPs. The distribution of MFA among the family matches expectations for allele frequency distribution in the full-sib family from a crossing of heterozygous parents. A biparental mating of hops is expected to give a tri-modal distribution of MFA with optima at segregation frequencies: for test-cross loci (AA x Aa) and (aa x Aa) at expected ratios of 1 : 3 and 3 : 1 (MAFs = 0.25 and 0.75) and inter-cross loci (AA x aa and Aa x Aa) at a ratio of 1 : 1 (MAF = 0.50). A preponderance of SNPs segregating at a frequency of 0.50 is indicative of a heterozygous species. The symmetrically distributed proportion of SNPs

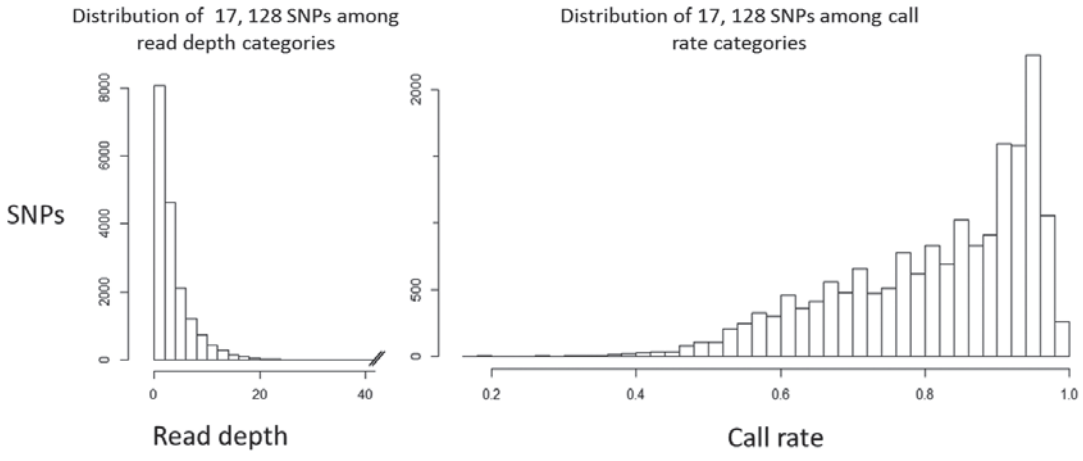

Fig. 1 Histograms of read depth and call rate for 17,869 SNPs detected in a population admixture. Read depth for most SNPs is less than 2.2 reads, nevertheless call rates are generally high. Stringent settings of parameters in the UNEAK pipeline lead to detection of a reduced number of markers with good read depth and call rates

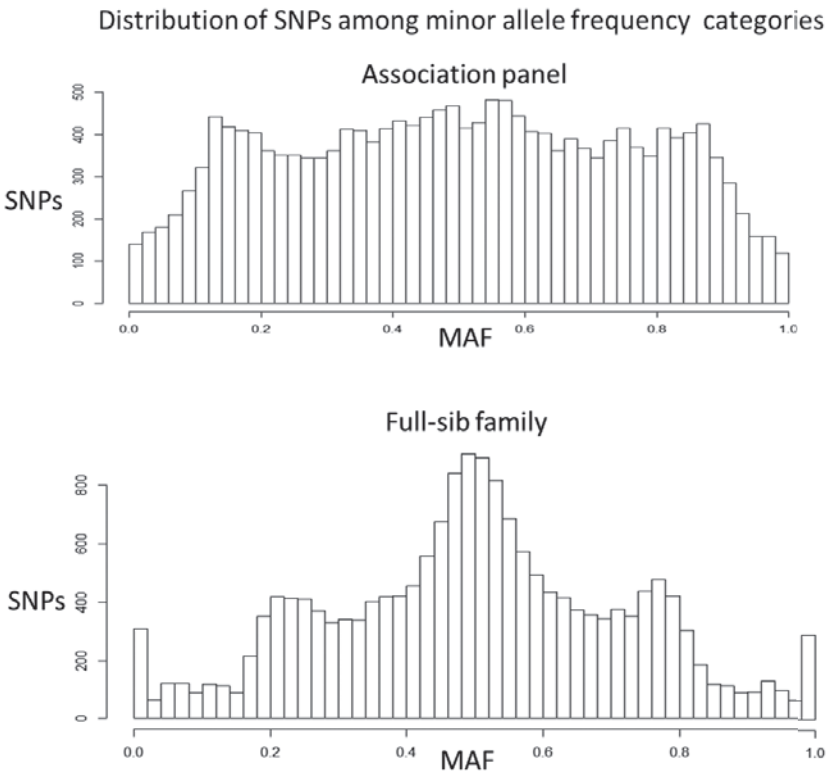

Fig. 3 Histograms of minor allele frequencies of 17, 869 SNPs among two populations of hops: (A) a diploid association panel (N = 60) and (B) a full-sib family (N = 103). Frequency optima in the full-sib family indicate Mendelian segregation for many SNPs

with extreme low (< 0 .05) or high frequency (> 0.95) (Fig. 3B) are likely to represent randomly distributed errors in SNP calling due to sequencing error and paralog-confusions [9]. Calling error of about 4.1 % (741/17,869) of SNPs is consistent with sequencing error rate of Illumina HiSeq 2000™ platform. Lu et al. found sequencing error in switchgrass to be just less than 3 % where default parameters in UNEAK were set (ETR = .03) to accommodate such [21]. Rare alleles in the population may be lost in the noise of the sequencing errors or unresolved paralogs, so SNPs with real MAFs below 0.05 are not likely to be detected. We removed the 741 putatively erroneous SNPs with 0.05 < MAF > 0.95 from further analyses.

To test the performance of the GBS UNEAK TASSEL markers in genetic distanced analyses for our CKCs, we employed Nei's neighbor joining (NJ) method with Jaccard's distances in TASSEL 4 for the association panel of diploids. Dendrograms were substantially congruent with published molecular studies of the CKCs (cited in Introduction) or with known pedigrees available at USDA-ARS-GRIN. Comparison of distance matrices and dendrograms were analyzed statistically (see below).

We tested the hyptothesis that more stringent filtering for subgroups of SNPs would not significantly affect the performance of the markers for genetic distance analyses. To provide a reduced set of more stringently filtered sets of SNPs, we filtered by three minimum call rates (0.5, 0.75, 0.90) within the diploid association panel. Filtering resulted in reduced sets of 16, 106; 10, 243 and 3,068 SNPs with a call rate => 0.5, => 0.75 and => 0.90, respectively. As an independent approach to filtering for high value SNPs suggested by Lu et al. [21], we also filtered and then grouped the SNPs based on MAF within the full-sib family of near 0.20–0.30, 0.45–0.55 and 0.70–0.80. Such SNPs should be segregating according to biological constraints, and are less likely to be erroneous. The dendrograms were substantially congruent for all filtered subgroups of markers. Following the standard methods of distance matrix and dendrogram comparison used previously for hops [29], we rejected the null hypothesis that the dissimilarity matrices were not related with Mantel's Z-tests [23] (331 < Z > 356, p < 0.0001, 5,000 permutations) and found all pairwise dendrogram comparisons to have large cophenetic correlation coefficients [34], i.e., > 0.95. Therefore, we concluded that further filtering of SNPs beyond UNEAK is unnecessary for genetic distance determination in hops. Figure 4 (see next page) shows a NJ tree constructed from 17, 128 SNP markers for the association panel using MEGA5 software [39]. Support for nodes is indicated as the percent recovery of the node among 10,000 bootstrap samplings. The large number of novel SNPs supplies exceptionally strong support for all nodes in the dendrogram.

Our success with SNP subgroups indicate that fewer markers may be enough for some studies and also suggest cost-efficiency might come from lower-coverage sequencing. Greater economy for sequencing and less coverage is accomplished by increasing the number of uniquely indexed genotypes per sequencing lane. A pilot run of 12 independent Saaz cultivars sequenced in a 384-plex lane of Illumnia HiSeq 2000™ generated an average of 11, 544 +/– 3578 SNPs per individual with an average read depth of 0.80 +/– 0.25, with our stringent UNEAK setting of c = 8 (minimum call rate = 0.67). While the average read depth of 0.80 is significantly lower than that for markers validated in this study (average read depth = 3.7), it is comparable to several successful, low-coverage studies that do not depend on imputation of missing data [5, 21].

4 Conclusions

We report the first application of next generation sequencing technology to simple and cost-effective molecular markers for hop research and for hop production quality control. More than an order-of-magnitude increase in the number of molecular markers for hops research potentially solves most previously reported problems. The power, ease and utility of SNP markers for true-ness-to-type determination, population genetic studies, and marker-assisted breeding in hops suddenly emerged from a combination of sequencing and computing technological advances. Additionally, SNPs markers, in contrast to other marker types, are superior in reproducibility, genome-wide distribution, co-dominance and "transfer-ability" to other laboratories and to other types of assay technologies. The sequenced-based markers can be re-created from the method without transfer of complicated sequence knowledge. The pipeline for marker production for "non-reference" genomes, GBS UNEAK TASSEL, has recently been validated for use in a several important, specialty crops; such as

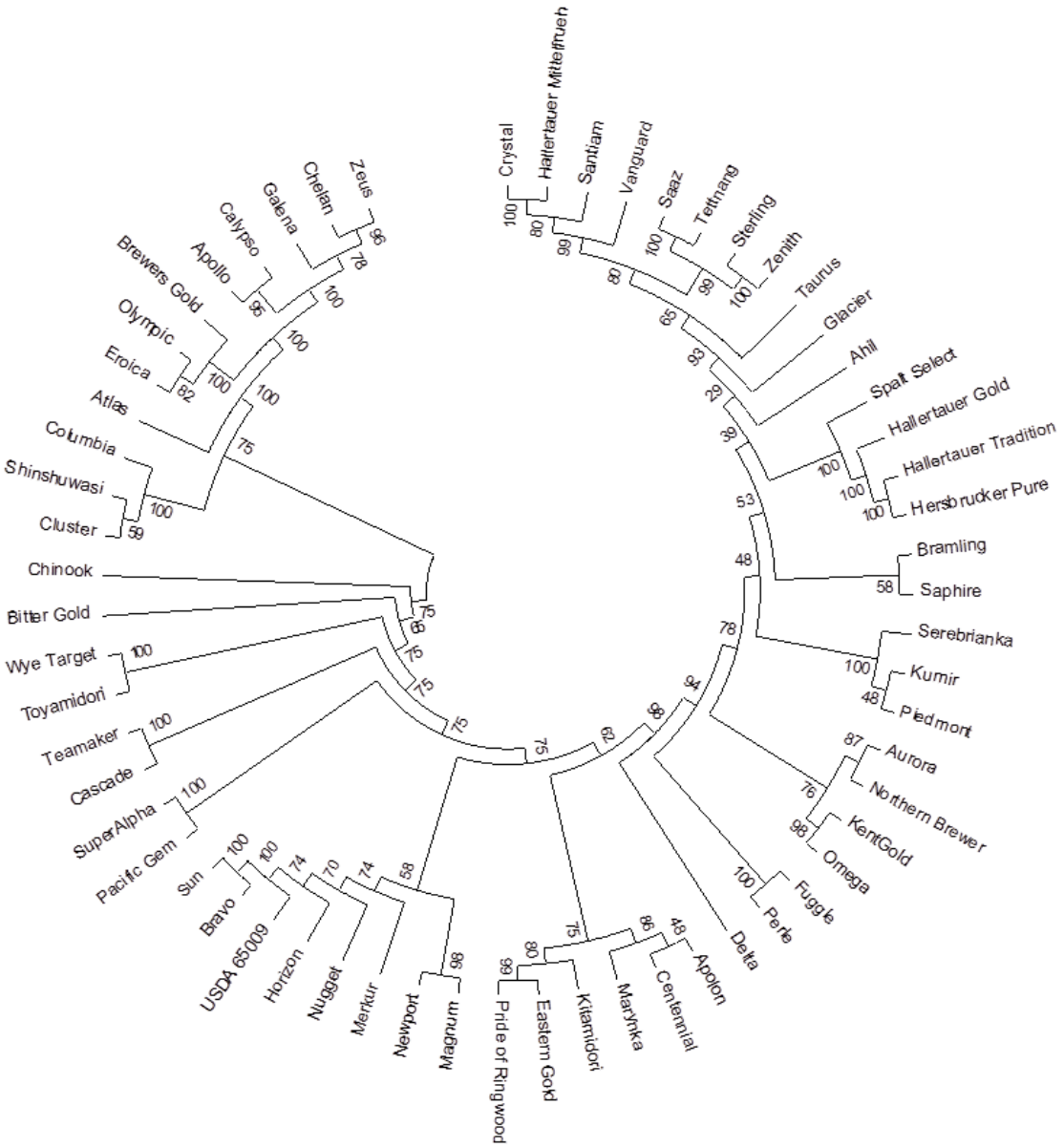

Fig. 4 Dendrogram of Nei's NJ hierarchical distance analysis. Nodes are labeled with support values from 10,000 bootstrap samplings implemented in MEGA5

conifers [9], switchgrass [21] and, now, hops. While such genotyping-by-sequencing marker technology is expected to replace hybridization-based microarray systems, previously developed genic SSR markers for hops [19], which can be assayed by next generation sequencing, are likely to complement GBS markers because of their hyper-variability and exonic distribution [36]. We characterized and then validated the hop SNPs with a number of well-defined methods, including read depth, call rate, coverage and segregation behavior within a large, full-sib family. Further validation was presented by application of the SNPs to standard genetic distance analysis among a set of commonly known cultivars. We showed the UNEAK pipeline with stringent settings produced a large number of markers that require little to no further filtering for applications to genetic distance analyses and, thus, also to true-ness-to-type determination. Bootstrap analysis with 17, 128 SNPs produced a tree of the cultivar's relatedness with unprecedented support.

Modifications to the GBS UNEAK TASSEL pipeline as implemented here are relatively easy to accomplish to match different sequencing budgets [5]. For example, choice of restriction enzyme for reduction can change the density, distribution and read depth; and pooling levels of 48-, 96-, 192-, 288- and 384-plex per lane effect sequencing costs. Comparison among GBS studies in other species suggests variation in sequencing strategy and in computing pipeline is easy to modify for a variety of study designs. The UNEAK pipeline obviates a sequence reference for SNP calling and may have some advantages in elimination of paralog-based calling errors. Nevertheless, more efficiency and power, due to imputation of missing data, are possible from GBS TASSEL application to detect referenced SNPs, using assembled transcriptomes and draft genomes from hops, when they become available.

Our demonstration and application of a flexible, next generation molecular markers system predisposes a variety of previously intractable studies. For example, haplotype mapping in diverse hop populations, mapping of recombination hot and cold spots, linkage disequilibrium mapping, detection of small-effect QTL, genomewide association and population-wide MAF-frequency finger-printing, polyploidy and aneuploidy determinations, and genome selection for powerful and efficient hop breeding. Many other applications are likely. Emergent, next generation technologies have provided new horizons for a special crop plant of great traditional value.

5 Acknowledgement

We especially acknowledge Cornell's Institute for Genomic Diversity GBS team (Sharon Mitchell, Charlotte Acharya and Theresa Fulton) for testing and establishment of GBS for hops. We thank Dr. Fei Lu, of Cornell, for guidance with UNEAK. Special thanks are extended to Dr. Edward Buckler (USDA-ARS, Cornell University) and his all his laboratory members for helpful advice, inspiration and suggestions that enabled this research. We also thank Dr. John Henning of USDA-ARS, Corvallis for helpful discussions on filtering SNP markers. This research was funded by Hopsteiner, S.S. Steiner, Inc.

6 References

1. Bassil, N.; Gilmore, B.; Oliphant, J. M.; Hummer, K. E. and Henning, J. A.: Genic SSRs for European and North American hop (*Humulus lupulus* L.), Genetic Resources and Crop Evolution **55** (2008), no. 7, pp. 959-969.
2. Bennett, M. D. and Leitch, I. J.: Plant DNA C-values database (release 6.0, Dec. 2012), (2012) http://www.kew.org/cvalues/
3. Bennetzen, J. L. and Kellogg, E. A.: Do plants have a one-way ticket to genomic obesity?, Plant Cell **9** (1997), no. 9, pp. 1509-1514.
4. 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.
5. Byrne, S.; Czaban, A.; Studer, B.; Panitz, F.; Bendixen, C. and Asp, T.: Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing, PLoS One **8** (2013), no. 3, pp. e57438.
6. 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.
7. Čerenak, A., Jakše, J., Luthar, Z., Javornik, B., Satovic, Z. and Carovic-Stanko, K.: Mapping of QTLS for yield and alpha acid content in hop, Acta Hort. (ISHS) **848** (2009), pp. 89-92.
8. Cerenak, A.; Satovic, Z. and Javornik, B.: Genetic mapping of hop (*Humulus lupulus* L.) applied to the detection of QTLs for alpha-acid content, Genome **49** (2006), no. 5, pp. 485-494.
9. Chen, C.; Mitchell, S.; Elshire, R.; Buckler, E. and El-Kassaby, Y.: Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform, Tree Genetics & Genomes (2013), pp. 1-8.
10. Elshire, R. J.; Glaubitz, J. C.; Sun, Q.; Poland, J. A.; Kawamoto, K.; Buckler, E. S. and Mitchell, S. E.: A robust,simple genotyping-by-sequencing (GBS) approach for high diversity species, PLoS ONE **6** (2011), no. 5, pp. e19379.
11. Hartl, L. and Seefelder, S.: Diversity of selected hop cultivars detected by fluorescent AFLPs, Theoretical and Applied Genetics **96** (1998), no. 1, pp. 112-116.
12. Henning, J.A. ; Townsend, M. S.; Gent, D. H. ; Bassil, N.; Matthews, P.D.; Buck, E. and Beatson, R.: QTL mapping of powdery mildew susceptibility in hop (*Humulus lupulus* L.), Euphytica (2011), no. 180, pp. 411-420.
13. Henning, J. A.; Townsend , S. M. and Kenny, S.: Potential heterotic crosses in hops as estimated by AFLP-based genetic diversity and coefficient of coancestry, J. Am. Soc. Brew. Chem. **62** (2004), no. 2, pp. 63-70.
14. Henning, J. H.; Townsend, M. S. and Matthews, P. D.: Predicting Offspring Performance in Hop (*Humulus lupulus* L.) Using AFLP Markers, J. Am. Soc. Brew. Chem. **68** (2010), no. 3, pp. 125-131.
15. Howard, E. L.; Whittock, S. P.; Jakse, J.; Carling, J.; Matthews, P. D.; Probasco, G.; Henning, J. A.; Darby, P.; Cerenak, A.; Javornik, B.; Kilian, A. and Koutoulis, A.: High-throughput genotyping of hop (*Humulus lupulus* L.) utilising diversity arrays technology (DArT), Theor. Appl. Genet. **122** (2011), no. 7, pp. 1265-80.
16. Ibarra-Laclette, E.; Lyons, E.; Hernandez-Guzman, G.; Perez-Torres, C. A.; Carretero-Paulet, L.; Chang, T.-H.; Lan, T.; Welch, A. J.; Juarez, M. J. A.; Simpson, J.; Fernandez-Cortes, A.; Arteaga-Vazquez, M.; Gongora-Castillo, E.; Acevedo-Hernandez, G.; Schuster, S. C.; Himmelbauer, H.; Minoche, A. E.; Xu, S.; Lynch, M.; Oropeza-Aburto, A.; Cervantes-Perez, S. A.; de Jesus Ortega-Estrada, M.; Cervantes-Luevano, J. I.; Michael, T. P.; Mockler, T.; Bryant, D.; Herrera-Estrella, A.; Albert, V. A. and Herrera-Estrella, L.: Architecture and evolution of a minute plant genome, Nature **498** (2013), no. 7452, pp. 94-98.
17. 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 (2013), pp. 1-13.
18. Kenny, M. P. S. T.: Chloroplast DNA differences between cultivated hop, *Humulus lupulus* and the related species *H. japonicus*, Theoretical and Applied Genetics **89** (1994), no. 2-3, pp. 372-378.
19. Koelling, J.; Coles, M.; Matthews, P. and Schwekendiek, A.: Development of new microsatellite markers (SSRs) for *Humulus lupulus*, Molecular Breeding **30** (2012), no. 1, pp. 479-484.
20. Koie, K., Inaba, A., Okada, Y., Kaneko, T. and Ito, K.: Construction of the genetic Linkage map and QTL analysis on hop (*Humulus lupulus* L.), Acta Hort. (ISHS) **668** (2005), pp. 59-66.
21. Lu, F.; Lipka, A. E.; Glaubitz, J.; Elshire, R.; Cherney, J. H.; Casler, M. D.; Buckler, E. S. and Costich, D. E.: Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel insights from a network-based SNP discovery protocol, PLoS Genetics **9** (2013), no. 1, pp. e1003215.
22. Majer, A.; Javornik, B.; Cerenak, A. and Jakse, J.: Development of novel EST-derived resistance gene markers in hop (*Humulus lupulus* L.), Molecular Breeding (2013), pp. 1-14.
23. Mantel, N.: The detection of disease clustering and a generalized regression approach, Cancer Research **27** (1967), no. 2, pp. 209-220.
24. Mardis, E. R.: Next-generation sequencing platforms, Annual Review of Analytical Chemistry (Palo Alto Calif) **6** (2013), pp. 287-303.
25. McAdam, E. L.; Freeman, J. S.; Whittock, S. P.; Buck, E. J.; Jakse, J.; Cerenak, A.; Javornik, B.; Kilian, A.; Wang, C. H.; Andersen, D.; Vaillancourt, R. E.; Carling, J.; Beatson, R.; Graham, L.; Graham, D.; Darby, P. and Koutoulis, A.: Quantitative trait loci in hop (*Humulus lupulus* L.) reveal complex genetic architecture underlying variation in sex, yield and cone chemistry, BMC Genomics **14** (2013), pp. 360.
26. Ming, R.; Bendahmane, A. and Renner, S.: Sex chromosomes in land plants, Annual Review of Plant Biology **62** (2011), pp. 485-514.
27. Nagel, J.; Culley, L. K.; Lu, Y.; Liu, E.; Matthews, P. D.; Stevens, J. F. and Page, J. E.: EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol, Plant Cell **20** (2008), no. 1, pp. 186-200.
28. Method for selecting hop line and breeding marker and primer set used for selecting hop line, EP2088194.
29. Patzak, J.: Comparison of RAPD, STS, ISSR and AFLP molecular methods used for assessment of genetic diversity in hop (*Humulus lupulus* L.), Euphytica **121** (2001), no. 1, pp. 9-18.
30. Patzak, J.; Henychova, A.; Krofta, K. and Nesvadba, V.: Study of molecular markers for xanthohumol and DMX Contents in hop (*Humulus lupulus* L.) by QTLs mapping analysis, BrewingScience – Monatsschrift für Brauwissenschaf **65** (2012), no. July/August, pp. 96-102.
31. Patzak, J.; Matousek, J.: Development and evaluation of expressed sequence tag-derived microsatellite markers for hop genotyping, Biologia Plantarum **55** (2011), no. 4, pp. 761-765.
32. Patzak, J.; Nesvadba, V.; Henychová, A. and Krofta, K.: Assessment of the genetic diversity of wild hops (*Humulus lupulus* L.) in Europe using chemical and molecular analyses, Biochemical Systematics and Ecology **38** (2010), no. 2, pp. 136-145.
33. Singh, S.; Gupta, S.; Mani, A. and Chaturvedi, A.: Mining and gene ontology based annotation of SSR markers from expressed sequence tags of *Humulus lupulus*, Bioinformation **8** (2012), no. 3, pp. 114-22.

34. Sokal, R. R. and Rohlf, F. J.: The comparison of dendrograms by objective methods, Taxon **11** (1962), no. pp. 33-40.
35. Sonah, H.; Bastien, M.; Iquira, E.; Tardivel, A.; Légaré, G.; Boyle, B.; Normandeau, É.; Laroche, J.; Larose, S.; Jean, M. and Belzile, F.: An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping, PLoS ONE **8** (2013), no. 1, pp. e54603.
36. Spindel, J.; Wright, M.; Chen, C.; Cobb, J.; Gage, J.; Harrington, S.; Lorieux, M.; Ahmadi, N. and McCouch, S.: Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations, Theoretical and Applied Genetics (2013), online.
37. Springer, N. M.; Ying, K.; Fu, Y.; Ji, T.; Yeh, C.-T.; Jia, Y.; Wu, W.; Richmond, T.; Kitzman, J.; Rosenbaum, H.; Iniguez, A. L.; Barbazuk, W. B.; Jeddeloh, J. A.; Nettleton, D. and Schnable, P. S.: Maizeinbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content, PLoS Genetics **5** (2009), no. 11, pp. e1000734.
38. Sucher, N. J.; Hennell, J. R. and Carles, M. C.: DNA fingerprinting, DNA barcoding, and next generation sequencing technology in plants, Methods Mol. Biol. **862** (2012), pp. 13-22.
39. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M. and Kumar, S.: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. **28** (2011), no. 10, pp. 2731-2739.
40. Temnykh, S.; DeClerck, G.; Lukashova, A.; Lipovich, L.; Cartinhour, S. and McCouch, S.: Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential, Genome Res. **11** (2001), no. 8, pp. 1441-52.
41. Timothy, M. B.; Candice, N. H.; Rajandeep, S. S.; Jillian, M. F.; James, M. J.; German, M.; Brieanne, V.; Buell, C. R.; Shawn, M. K. and Natalia de, L.: Marker density and read depth for genotyping populations using genotyping-by-sequencing, Genetics **193** (2013), no. 4, pp.224-243.
42. Trick, M.; Adamski, N. M.; Mugford, S. G.; Jiang, C. C.; Febrer, M. and Uauy, C.: Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat, BMC Plant Biol **12** (2012), pp. 14.
43. van Bakel, H.; Stout, J.; Cote, A.; Tallon, C.; Sharpe, A.; Hughes, T. and Page, J.: The draft genome and transcriptome of *Cannabis sativa*, Genome Biology **12** (2011), no. 10, pp. R102.
44. Wang, G.; Tian, L.; Aziz, N.; Broun, P.; Dai, X.; He, J.; King, A.; Zhao, P. X. and Dixon, R. A.: Terpene biosynthesis in glandular trichomes of hop, Plant Physiology **148** (2008), no. 3, pp. 1254-1266.
45. Whittock, S., Leggett, G., Jakše, J., Javornik, B., Carling, J., Kilian, A., Matthews, P.D., Probasco, G., Henning, J.A., Darby, P., Čerenak, A. and Koutoulis, A.: Use of diversity array technology (DART) for genotyping of *Humulus lupulus* L., Acta Hort. (ISHS) **848** (2009), pp. 59-64.

Received 22 September 2013, accepted 21 November 2013