

## **Non-Mendelian SNP inheritance and atypical meiotic configurations are prevalent in hop (*Humulus lupulus* L.)**

Dong Zhang<sup>1</sup>, Katherine A. Easterling<sup>1</sup>, Nicholi J. Pitra<sup>1</sup>, Mark C. Coles, Edward S. Buckler, Hank W. Bass, Paul D. Matthews<sup>2</sup>

Hopsteiner, S.S. Steiner, Inc., New York, New York, 10065 (DZ, KAE, NJP, MCC, PDM), Institute for Genomic Diversity, Cornell University, Ithaca, New York, 14853 (DZ, ESB), Agricultural Research Service, United States Department of Agriculture, Ithaca, New York, 14853, (ESB), Department of Biological Science, Florida State University, Tallahassee, Florida, 32306-4295 (KAE, HWB)

### Footnotes:

PDM created germplasm resources, devised and directed the studies; ESB provided guidance on statistical analyses and project design; MCC collected samples and prepared DNA extracts; NJP prepared the GBS sequencing libraries; NJP and DZ analyzed the GBS marker data and interpreted linkage results; the 3D cytogenetic data was collected by KAE, analyzed by KAE and HWB (ORCID ID 0000-0003-0522-0881), and interpreted by KAE, HWB, all authors contributed to the writing and editing, with major contributions from DZ, KAE, HWB, and PDM.

The study was funded by Hopsteiner, S.S. Steiner, Inc.

<sup>1</sup>These authors contributed to this work equally.

<sup>2</sup>Corresponding author: Email [pmatthews@hopsteiner.com](mailto:pmatthews@hopsteiner.com)

## ABSTRACT

Hop breeding programs seek to exploit genetic resources for bitter flavor, aroma and disease resistance. However, these efforts have been thwarted by segregation distortion including female-biased sex ratios. To better understand the transmission genetics of hop, we genotyped 4,512 worldwide accessions of hop, including cultivars, landraces, and over 100 wild accessions, using a genotyping-by-sequencing (GBS) approach. From the resulting ~1.2M single nucleotide polymorphisms, pre-qualified GBS markers were validated by inferences in population structures and phylogeny. Analysis of pseudo-testcross mapping data from F1 families revealed mixed patterns of Mendelian and non-Mendelian segregation. Three-dimensional cytogenetic analysis of late meiotic prophase nuclei from two wild and two cultivated hop revealed conspicuous and prevalent occurrences of multiple, atypical, non-disomic chromosome complexes, including autosomes. We used genome-wide association studies and  $F_{ST}$  analysis to demonstrate selection mapping of genetic loci for key traits, including sex, bitter acids, and drought tolerance. Among the possible mechanisms underlying the observed segregation distortion from the genomic data analysis, the cytogenetic analysis points to meiotic chromosome behavior as one of the contributing factors. The findings shed light on long-standing questions on the unusual transmission genetics and phenotypic variation in hop, with major implications for breeding, cultivation, and the natural history of *Humulus*.

## CORE IDEAS:

- GBS Pseudo-testcross data from F1 families reveal extensive segregation distortion.
- Cytogenetic analyses reveal atypical, non-disomic, meiotic configurations.
- Genetic loci associated with sex determination are mapped to the Linkage Group 4.
- Hotspots exhibiting unusual  $F_{ST}$  variance provide clues about signature of selection in

hops.

- Combined analyses implicate meiotic chromosome behavior in segregation distortion.

### Abbreviations:

GBS: genotyping-by-sequencing; SNP: single nucleotide polymorphism; GWAS: genome-wide association studies; SD: segregation distortion; NOR: nucleolus organizer region; CV: modern cultivars;  $F_{ST}$ : fixation index; Pt: pseudo-testcross; MAF: minor allele frequency; t-SNE: t-Distributed Stochastic Neighbor Embedding; IBS: identity by state; LLE: Locally Linear Embedding method; LG: linkage groups; MLM: mixed linear model.

### INTRODUCTION

The Cannabaceae family of flowering plants has a rich history of contributions to humanity, with the promise of still greater contributions as a result of new commercial values and invigorated research in two members, *Humulus lupulus* (hop) ( $2n=2X=20$ ) and *Cannabis sativa* (hemp, marijuana) ( $2n=2X=20$ ) (van Bakel et al., 2011), which diverged around 27.8 Myr (Laursen, 2015). The hop plant (*H. lupulus*) is a high-climbing dioecious bine and an herbaceous perennial with historic uses in brewing and nutraceutical medicine and modern uses as bio-fuel and animal fodder (Siragusa et al., 2008). Metabolic engineering and marker-directed breeding in hop recently increased as clinical studies identified hop-derived prenylflavonoids as therapeutic agents for treatment of cancer, dyslipidemia, and postmenopausal symptoms (Ososki and Kennelly, 2003; Stevens and Page, 2004; Nagel et al., 2008; Miranda et al., 2016). Despite the value of these traits and products, the hop plant has proven refractory to traditional breeding and conventional genomic strategies for genetic dissection of complex, quantitative traits. Several factors contribute to this difficulty, including its aspects of its reproductive

system such as dioecy and obligate outcrossing, high degree of heterozygosity, large genome size, and poorly understood sex-determination system (Neve, 1958).

Wild *H. lupulus* is represented by at least five extant taxonomic varieties: (1) var. *lupulus* for European wild hop, (2) var. *cordifolius* mainly distributed in Japan, (3) var. *neomexicanus* in the Southwestern U.S., (4) var. *pubescens* in the Eastern/Midwestern U.S. and (5) var. *lupuloides* throughout the northern Great Plains and spreading into other parts of North America. Asian and North American wild hop resemble each other morphologically, suggesting a genetically close relationship, while they differ more so from European hop (Murakami et al., 2006). Many contemporary cultivars are hybrids of North American and European genetic materials, in which North American hop have been characterized by their higher bitterness and aroma (Reeves and Richards, 2011) than European cultivars. In other crops, breeding programs have successfully exploited novel genetic variations from wild exotic germplasms into modern cultivars (Tanksley and McCouch, 1997; Bradshaw, 2016) to gain desirable traits such as desired flavors, drought tolerance, and disease resistance. Successes with wild resources and predictions of climate change have spurred resurgence in conservation biology of plant genetic resources (Castañeda-Álvarez et al., 2016; Gruber, 2016).

Molecular marker systems including non-referenced GBS markers (Matthews et al., 2013) and GWAS (Henning et al., 2015; Hill et al., 2016) have been developed and used for genetic mapping of disease resistance and sex determination. Despite these advances, understanding the genetic inheritance patterns in hop remains a major challenge. For example, significant distortion from Mendelian segregation expectations has been repeatedly reported in mapping populations, indicating that the segregation bias was due to genetic properties rather than genotyping errors (Seefelder et al., 2000; McAdam et al., 2013). Relatedly, female-biased sex ratios have been observed in most families (Neve

1991; Jakse et al., 2008). The segregation data for hop resemble to some extent those from segregation distortion systems that are well described in certain plants known to exhibit chromosomal rearrangements and meiotic drive (reviewed by Taylor and Ingvarsson, 2003). For instance, in *Clarkia*, *Oenothera*, *Viscum*, and *Calycadenia*, translocation heterozygosity and other chromosomal abnormalities can modify Mendelian segregation patterns and impact intraspecies fertility (Snow, 1960; Wiens and Barlow, 1975; Carr and Carr, 1983; Rauwolf et al., 2008; Golczyk et al., 2014).

With regard to the chromosomal composition of hop, classical cytogenetics has established that the species has heteromorphic sex chromosomes and occasional meiotic quadrivalents of unknown chromosomal composition (Sinotô, 1929; Neve, 1958; Haunold, 1991; Shephard and Parker, 2000). More recently, somatic hop karyotypes have been developed for several varieties, including FISH mapping of the locations of the NOR, 5S rDNA and the abundant *Humulus* subtelomeric repeats, HSR1 (Karlova et al., 2003; Divashuk et al., 2011). Functional genomics in hop has been advanced by detailed linkage analysis (Henning et al., 2017) and whole genome sequencing (Natsume et al., 2015), yet these data are not integrated into a single annotated reference genome, nor connected to the chromosome numbers of the published karyotypes.

To further characterize the genome of hop, we carried out next generation sequencing (NGS) of 4512 accessions, including 22 F1 families, genotyped with GBS SNP marker system, comprising 1.2 million SNPs. This study greatly extends the previous NGS GBS studies in hop (Matthews et al., 2013; Henning et al., 2015; Hill et al., 2016) with much larger association panels and marker sets, providing new population structure information. Instead of filtering out SNPs that show segregation distortion (SD), we included and exploited them in our analysis, strengthening the size and quality of candidate gene lists. We also examined several plants at the cytological level and found

peculiarities consistent with the marker segregation irregularities. These new findings advance our working knowledge of the genome of hop, and point to chromosome structure and recombination constraints as important aspects guiding future breeding strategies.

## **MATERIALS AND METHODS**

### **Plant materials**

The hop plants used in this study were grown under standard agronomic conditions at the Golden Gate Ranches, S.S. Steiner, Inc, Yakima, WA. The un-domesticated, exotic hop are from the National Clonal Germplasm Repository in Corvallis, Oregon (accession details in Table S1-S3). Fifty milligrams of young leaf tissues were extracted in a 96 well block using Qiagen Plant DNeasy Kits and was tested for quality, quantity, and purity, prior to library preparations, using an Agilent 2100 Bioanalyzer (Applied Biosystems, Foster City, CA) and Life Technologies (Carlsbad, CA) Qubit 3.0 Fluorometer. The GBS libraries were prepared using the *ApeK1* enzyme according to Elshire, et al. (Elshire et al., 2011). Pools of 96 accessions were sequenced on one lane of an Illumina HighSeq 2000 (Illumina, San Diego, CA)

### **3D cytogenetic analysis of male meiotic prophase nuclei**

*H. lupulus* panicles were harvested from the Hopsteiner male yard (Yakima, WA, USA) throughout the day, fixed in Carnoy's solution (3:1 ethanol:acetic acid) overnight, and exchanged into 70% ethanol for storage at -20C. For 3D microscopy, buds were equilibrated in meiocyte Buffer A [MBA, (Bass et al., 1997)] for 15 min at RT, repeated twice, then fixed in 2% formaldehyde in MBA at RT for 2h. After fixation, buds were washed twice in MBA, 15-min each, at RT, and stored in MBA at 4C. Anther lengths were recorded and meiotic cells were microdissected onto glass slides and mounted in VectaShield + DAPI (Vector Laboratories). Three-dimensional images were collected on

a DeltaVision deconvolution microscope, using a 60X lense and 0.2 micron Z-step optical sections [as summarized by (Howe et al., 2013)]. 3D datasets capturing entire nuclei at various stages of meiosis were collected. Deconvolved images were further processed using linear scaling of intensity and software programs (Volume Viewer, Copy Region, Projection, 3D Model) to allow for inspection from various angles.

Classification and quantification of meiotic chromosome configurations were made on diakinesis stage nuclei, using a combination of visual inspection methods, including paging back and forth through individual optical sections of the 3D data stacks along with inspection of through focus projections made from multiple angles as well as viewing of cropped sub-volumes. For this study, a nucleus determined to be in diakinesis had at least two bivalents less than 5 microns in length. The number of bivalents and non-bivalent complexes were counted for each plant using at least 20 diakinesis nuclei. The non-bivalent complexes were split into two sub-categories, *quadrivalents* (two bivalents joined into a ring of 4, or interlocked chain link structure) or *other complexes* (non-quadrivalents with variable number of chromosomes).

### **SNP calling and quality control**

The reference sequence refers to a draft haploid genome sequence of Shinshu Wase (SW) (Natsume et al., 2015), which is a modern cultivar bred from a seedling selection cross between Saazer and White Vine-OP. The draft genome, with a total size of 2.05 Gb, consists of ~130,000 scaffolds covering approximately 80% of the estimated genome size of hop (2.57 Gb).

Tassel 5 GBS v2 Pipeline (Glaubitz et al., 2014) was applied to identify tags with at least 10x total coverage, and to call SNPs. Tag sequences were mapped to the reference genome using BWA aligner.

One main source of erroneous SNP calling is misalignment caused by incomplete reference genome, gene duplication and low-complexity regions. To filter out erroneous SNPs due to misalignment, we used two criteria: (1) SNPs with an excessive coverage can be false positives. We observed that heterozygosity rates and MAF are significantly increased when read coverage exceeds 127 (Figure S1). (2) The orientation of paired reads of the cultivar Apollo (unpublished data), a highly used maternal line in our F1 families, was used to detect false positive SNPs caused by gene duplications. Paired-end alignment was generated by BWA Sampe. Identification of correctly aligned regions was based on SAM flags indicating reads mapped in proper pairs. Using criteria (2) was able to detect ~73% SNPs with the excessive coverage.

### **Pseudo-testcross**

Three F1 families were used to conduct pseudo-testcross (Pt) recombination mappings, including (1) “144” (N = 179) derived from a cross between Nugget (maternal line) and Male50 (paternal line); (2) “247” (N = 364) derived from two parental lines, Super Galena and Male15; (3) “265” (N = 95) derived from a cross between Chinook and Male57. Using markers heterozygous in the maternal line and homozygous in the paternal line, three genetic map sets were constructed, consisting of 3551 SNPs for “144”, 2369 SNPs for “247” and 4506 SNPs for “265”.

Our analyses followed the main steps in HetMappS pipelines (Hyma et al., 2015). Specifically, (1) to remove contaminants, identity by state (IBS) based distance matrices calculated by TASSEL (Bradbury et al., 2007) were used to identify outliers for each family; (2) SNPs having both parental genotypes (e.g. AA×Aa) with read depth  $\geq 4$  were retained for the next step; (3) in progeny, SNPs with average read depth  $\geq 4$  and with site coverage  $\geq 50\%$  were retained for the next step; (4) to eliminate the effect of under-



calling heterozygotes and sequencing errors, we masked progeny genotypes with depth=1, and converted genotypes aa to Aa because genotype aa cannot exist for parental genotypes AA×Aa in Pt; (5) after correction, SNPs with  $15\% \leq \text{MAF} \leq 35\%$  were selected to create linkage groups, and SNPs with  $5\% \leq \text{MAF} < 15\%$  were deemed the pronounced SD markers; (6) to cluster and order markers, an adjacency matrix with Spearman's correlation ( $\rho$ ) were derived from the remaining SNPs; (7) on the basis of absolute values of  $\rho$ , the Louvain method (Blondel et al., 2008) implemented in NetworkX (<http://networkx.github.io/>) was applied to detect communities (clusters). The Louvain method is an efficient algorithm for community detection in large networks. A similar method, modulated modularity clustering (MMC) (Stone and Ayroles, 2009), has been successfully applied to construct linkage groups. The clustering patterns of markers were cross-checked by the locally linear embedding method (LLE) (Roweis et al., 2000), a nonlinear dimensionality reduction method, implemented in Python scikit-learn; (8) to identify coupling phase from each “absolute  $\rho$ ” cluster, negative values of  $\rho$  were set to zero, and the Louvain method was applied to positive values of  $\rho$  (Hyma et al., 2015); (9) MSTmap (Wu et al., 2008) was used to provide a solution of genetic ordering within each linkage group.

Putative 10×2 linkage groups in coupling were obtained in each F1 family. As the karyotype has not been fully understood in hop, the linkage group ID numbers were arbitrarily assigned in “144”. Using the genetic map in “144” as a central reference, we assigned the ID numbers to linkage groups in other crosses. Linkage groups deriving from two grandparents are distinguished by suffix “.1” and “.2”. Linkage groups may or may not represent one chromosome due to pseudo-linkage resulting from chromosomal rearrangement, as discussed in Results.

### **Genome-wide association studies (GWAS)**

An association population includes 850 individuals, in which 837 (116 males and 721 females) are progeny in 6 F1 families and 13 are paternal lines. Male and female were encoded as '1' and '0' individually. A total of 356,527 SNPs with coverage  $\geq 50\%$  and MAF  $\geq 5\%$  were retained. The Mixed Linear Model (MLM) (Bradbury et al., 2007; Lipka et al., 2012) was used to assess genotype-phenotype association. The Bonferroni method was used to adjust the significance cutoff for an overall probability of 0.05 for type I error.

## RESULTS

### Phylogenetic relationships of modern cultivars and North American indigenous exotics

European var. *lupulus* is the ancestor of most commercial hop used today, thereby commercial cultivars retain a large proportion of var. *lupulus* genome. In addition, the genetic diversity of hop crop has been contributed by mostly male donors from North America and Asia. To understand the phylogenetic relatedness in hop races, we focused on a subset of 251 accessions, consisting of 183 modern cultivars (CV) consisting of all progenitors of F1 families in this study and 68 wild hop as summarized in Figure 1. The neighbor-joining tree (Figure 1a) shows three distinct clusters. The modern cultivars were clustered together, indicating a common derivation in domestication of hop. The other two clusters reflect geographical origins of North American wild hop (Figure 1b), in which one group (SW\_wild) includes 22 Southwestern U.S. wild hop (represented by var. *neomexicanus*), and the other group contains 20 wild hop (represented by var. *lupuloides*) from Northern U.S./Canada (N\_wild) and 3 (represented by var. *pubescens*) from Midwestern U.S. (MW\_wild). Seven wild individuals from Kazakhstan are intermediate among the modern cultivars, consistent with a previous inference (Murakami et al., 2006) of a close genetic relationship between wild hop from Europe and the Altai region (close to Western China, located on boundaries of Russia, Mongolia, Kazakhstan and China).

The level of population differentiation, fixation index ( $F_{ST}$ ), was measured across the three clusters. SW\_wild exhibits relatively close genetic relationship ( $F_{ST} = 0.1663$ ) with N\_wild, apparently supporting relatively close ancestry and geographical origins of the two wild populations. Genetic distinction between the modern cultivars and the North American wild hop is evident: [ $F_{ST}$  (CV vs. SW\_wild) = 0.31;  $F_{ST}$  (CV vs. N\_wild) = 0.295].

To demonstrate the population structure of F1 families and varieties clones ( $N \geq 60$ ) (Figure S2a) in our dataset, we used a nonlinear algorithm (implemented in Python scikit-learn), t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008), for dimension reduction of the identity by state (IBS)-based distance matrix. The F1 families derived from genetically divergent progenitors can be easily distinguished from one another, while the half-sibling families exhibit ambiguous clustering patterns (Figure S2).

### 3D cytogenetic analysis of meiotic chromosomes

Cytological analysis was performed using 3D imaging of nuclei from four different male *H. lupulus* hop plants that were obtained from either wild seed (var. *lupuloides* from Crooked Lake or var. *neomexicanus* from Chimney Rock) or produced as progeny from crosses within the Hopsteiner breeding program (cross 256, cross 255). Late meiotic prophase nuclei were stained with DAPI and imaged using 3D microscopy in order to survey the chromosome configurations. The hop meiocytes used in this study should have a chromosome constitution of  $2n=2X=20$ , including sex chromosomes (Sinotô, 1929; Winge, 1929) of unresolved constitution.

Typically, diploid nuclei from organisms with normal disomic inheritance exhibit diakinesis chromosomes in which each bivalent is distinct and spatially separate from other bivalents and distributed around the nuclear periphery. In striking contrast, hop diakinesis described here showed considerable deviation from a “typical” pattern of 10 well-separated bivalents, as summarized in Figure 2 for diakinesis-stage nuclei. A notable diversity of chromosome configurations was observed, including canonical bivalents (arrows, Figure 2a) and various other complexes. The average number of bivalents per nucleus is depicted for each plant (Figure 2a). None of the plants averaged more than 6 bivalents per nucleus, leaving at least four homolog pairs of chromosomes on average per nucleus that could be involved in other configurations. Chimney Rock (var. *neomexicanus*) contained an average of 2.2 bivalents per nucleus, by far the fewest of the four plants examined. The other wild plant, Crooked Lake (var. *lupuloides*) contained an average of 5.7 bivalents per nucleus. The F1 progeny from cross 265 and 255 contained an average of 4.7 and 6.0 bivalents, respectively. Nuclei with 10 bivalents were observed at a low frequency (~5%) in Crooked Lake and crosses 265 and 255, but so far not at all in Chimney Rock. Taken together, the findings from this 3D analysis reveal that complexes are not limited to heteromorphic sex chromosomes, but instead are both prevalent and heterogeneous within and among different plants.

In order to further classify the chromosome configurations, we carried out detailed analysis of sub-nuclear regions cropped in 3D from the full datasets (Figure 2b-d). Individual chromosomes or complexes were classified on the basis of their morphology and proximity into several categories, bivalents, quadrivalents, and other complexes. The bivalents (Figure 2b) were classified as three types: “Ring”, which appeared as pairs of chromosomes frequently in a ring configuration; “Sex (XY)”, which appeared as the only heteromorphic pair in the set; or “NOR)-linked”, which appeared to be attached to a nucleolus. The quadrivalents (Figure 2c) were defined as two pairs of non-homologous

chromosomes joined together by presumed chiasmata. The quadrivalents were classified as three types: “Ring of four” which appeared as two bivalents in an open ring; a “Double ring” which appeared as two bivalents in a chain-link pattern; or “NOR-linked plus X” which appeared as connected to both the nucleolus and the X chromosome of the sex bivalent. Quadrivalents of any type were found to occur with an average per-cell frequency of 1.4 for Crooked Lake, 0.2 for Chimney Rock, 1.3 for cross 265 hybrid, and 1.0 for cross 255 hybrid. The other complexes (Figure 2d) were heterogeneous and less readily classified, but referred to as “Multiple” which included non-quadrivalent complexes of variable composition, or “Long chain” which appeared as numerous interconnected series of chromosomes. The most common “Multiple” complexes involved more than two pairs of chromosomes, but occasional complexes of one bivalent plus one univalent were also observed. Combining all types, the average per-cell frequencies of complexes were 0.55 for Crooked Lake, 2.08 for Chimney Rock, 0.38 for cross 265 hybrid and 0.85 for cross 255 hybrid.

Overall, complexes were found in all plants, wilds and F1 progeny. The wild plant from Chimney rock (var. *neomexicanus*) exhibited an unexpectedly large number and variety of complexed chromosomes, including long chain arrangements (e.g. Figure 2d) and atypical configurations with more than one nucleolus. To the extent that the complexes are held together by crossovers, these findings may reflect translocation heterozygosity, segmental aneuploidy, or other atypical pairing regions resulting in the segregation distortion reported here and previously (Seefelder et al., 2000; McAdam et al., 2013).

### **Segregation distortion in progeny from F1 crosses**

Genetic markers that exhibit non-Mendelian inheritance frequencies can result from biological processes or technical errors. While genotyping errors are random, the biologically distorted markers typically exhibit pronounced correlation with Mendelian

segregation markers. On the basis of clustering of pairwise Spearman's correlation in pseudo-testcross (Pt) markers (exemplified in Figure 3) in three F1 families, we observed that the loci with 5-15% minor allele frequency (MAF), deviated significantly from the 25% allele frequency expected for pseudo-testcross (Pt) markers. These MAFs account for 28.3%, 49% and 48.3% in families "144", "247" and "265" respectively, in which proportions of the distorted loci correlated ( $\rho \geq 0.3$ ) to the Mendelian segregation markers (15-35% MAF) are 78.3%, 48.9% and 71.8% (Figure S4). This finding is consistent with a previous QTL study in hop using DArT markers (McAdam et al., 2013). These observations are consistent with two resulting hypotheses: (1) that large scale, genome-wide, and atypical meiotic chromosomal interactions occur in the progenitors of the three populations; and (2) that patterns of linkage can differ across the three populations.

Analyses of pseudo-testcross data from families "144" and "247" show multiple 'super' linkage groups in terms of their size and inter-marker correlation (Figure 4a,S3a). In family "265", linkage groups tend to have equal size (Figure S3b), but exhibit relatively high correlation to one another. Alignments across the three sets of maternal linkage maps, before phasing coupling groups, (Figure 4b,4c) show that most of the common/anchor markers were distinctly clustered. The clustering patterns of markers (exemplified in Figure 5) are cross-checked using the Louvain method (Blondel et al., 2008) and the locally linear embedding method (LLE) (Roweis et al., 2000) (see details in Methods).

Translocation heterozygosity can extend linkage beyond the limits of a single chromosome, resulting in segregation ratios distorted from Mendelian expectations. Severe SD is known to result from altered recombination and linkage that occurs near breakpoints, creating pseudolinkage, or suppressing crossovers, and complicating marker

ordering efforts in these regions (Taylor and Ingvarsson, 2003; Rauwolf et al., 2008; Farré et al., 2011). We used spatial coordinates calculated from LLE of Pt markers, in agreement with correlation heatmaps, to visualize genetic linkage patterns that emerge with and without inclusion of SD markers, as shown in Figure 5. The markers showing segregation distortion (yellow dots in Figure 5) appear to bridge the otherwise distinct linkage groups (red or blue dots, Figure 5). These intriguing marker behavior patterns could be related to the chromosome interactions observed at late prophase by 3D cytology (Figure 2). Together, these observations suggest that chromosome structural variation impacts hop transmission genetics.

The largest linkage group is from family “265”, shown in Figure 6. It appears as a major linkage complex that is derived from 5 interacting groups of well-linked markers with  $15\% \leq \text{MAF} \leq 35\%$  (Figure 6a). By plotting the normal and distorted markers in separate colors (Figure 6b, grey vs. cyan, respectively), a clear pattern emerges in which the SD markers predominate in the space bridging the non-distorted markers. Chromosome markers appear to change in their degree of distortion as they approach and enter the area of convergence. This may reflect a multitude of chromosomal phenomena superimposed over multiple individuals. Indeed, our cytogenetic analysis shows variable chromosomal interaction patterns for multiple nuclei from individual plants.

One linkage group (LG) in one family corresponding to multiple groups in the other family, suggests loci in common involved in recombination suppression and linkage disequilibrium, which is most likely influenced by the presence of chromosome rearrangements in the progenitor of the former family. One striking case (Figure 7,S5) in LG2.1 of family “144” corresponding to two coupling LGs (2.1 and 2.2) in “265”. Two additional correspondences (LG1.1-LG1.2 and LG3.1-LG3.1) were used as positive control of the clustering approaches. However, such one-to-multiple correspondence was

seldom observed across the three families. That may reflect the conservation of normally segregating chromosomal parts positioning in the heterozygotes complex and invariable occurrence of the translocation heterozygotes in the progenitors of the three families.

### **GWAS for sex determination**

Despite the prevalence of segregation distortion, the GBS linkage data should still be amenable to genetic analysis linking genotype to phenotype. To test this idea, we examined markers for sex determination in hop, a dioecious species with a chromosomal sex determination system (Shephard and Parker, 2000; Ming et al., 2011). We used a mixed linear model to assess evidence of phenotype-genotype association as shown in Figure 8. In families “247” ( $N = 364$ ,  $N_{\text{male}} = 30$ ) and “265” ( $N = 95$ ,  $N_{\text{male}} = 13$ ), LG4 consistently shows the most striking association with sex (Figure 8a,S6), even though “265” has a small effective population size. This signal was additionally supported by  $F_{\text{ST}}$  mapping in “247” (Figure 8b), but pseudo-testcross only accounts for part of association signals. To extend the analysis genome-wide, we assessed association between 356,527 markers and 850 individuals ( $N_{\text{male}} = 129$ ,  $N_{\text{female}} = 721$ ). A total of 588 SNPs with  $P \leq 10^{-7}$  were identified (Table S4 and Figure 8c), with LG4 and other LGs accounting for 38.6% and 0.0% of the association markers, respectively. The 588 SNPs were highly correlated (Figure 8d), as would be expected if the association markers derive primarily from one LD block. Adding up scaffolds showing association approximates ~9.75Mb of the mapping resolution accounting for ~0.38% of the hop genome. These results confirm the importance of our LG4 in sex determination in hop, suggesting that LG4 may be a sex chromosome. These findings establish the utility of the GBS data for linkage mapping and provide clues about specific genes and families involved in sex determination system in hop.

### **Genetic differences and phenotypic variation across populations**



To assess genetic contributions to between-population phenotypic differences, we used  $F_{ST}$  analysis (Table S5), plotted as linkage group-based pairwise  $F_{ST}$  heatmaps, for the population differentiation across var. *neomexicanus*, var. *lupuloides* and CV (Figure S7).  $F_{ST}$  values are a measure of allele frequency variance between populations, and they can be used to identify regions of domestication or targets for breeding. From this analysis, two notable patterns emerged. First, the degree of genetic variation, as expected, is much greater in CV vs. either of the wilds, *neomexicanus* or *lupuloides*, than in the wilds, *neomexicanus* vs. *lupuloides*. Regions of high  $F_{ST}$  in CV vs. *neomexicanus* are also found to exhibit high  $F_{ST}$  in CV vs. *lupuloides*. Second, the 5 largest linkage groups account for a large proportion of genetic variation between populations. Taken together, these results confirmed our suspicion that domestication traits should result in unusual  $F_{ST}$  values when comparing wilds to cultivars, but not between wilds, which have undergone different degrees of natural versus domestication-based selection for certain traits. The hotspots with unusually high  $F_{ST}$  values can be prioritized to identify genetic loci affecting certain traits, especially for chemical composition and drought tolerance.

## DISCUSSION

Hop crop acreage and usage is rapidly expanding and diversifying because of a burgeoning craft brewing industry. Hop breeding programs have a long history of attempting to exploit genetic resources for bitter flavor, aroma and disease resistance. However, a worsening drought and unseasonably hot weather pose major challenges to these efforts. For example, in Europe and the US, most hop farms experienced severe water shortage in 2015. Like many other crops, exploitation of novel genetic variation in response to drought stress is of paramount importance for a sustainable hop production system.

### Meiotic chromosome pairing interactions in wild and hybrid hop

Previous cytogenetic and genetic studies together with the current genomic findings prompted cytogenetic analysis for evidence of non-disomic meiotic chromosome configurations. Analysis of more than 100 diakinesis stage nuclei confirmed the presence of atypical meiotic chromosomal configurations in hop revealing additional complexities (Figure 2d). This study confirms the tendency for sex chromosomes to be involved in quadrivalent, or multiple associations (Sinotô, 1929; Winge, 1929). In addition, these new findings clearly implicate autosomes and possible structural heterozygosity as prevalent in hop. This idea is consistent with early speculations from Winge regarding autosomes being involved in tetrapartite/quadrivalent associations [reviewed by (Vyskot and Hobza, 2004)]. Here, only one set of heteromorphic sex chromosomes were observed in all plants, but this cannot be stated as certain without chromosome specific FISH probes. However, sex bivalent and NOR-bearing chromosomes are the only chromosomes that are morphologically distinct, and therefore, their interactions with each other, alone, or with other chromosomes were noted. Specifically, sex bivalents were observed to interact directly with the NOR-bearing chromosome in about 2 in 20 nuclei for each plant except Crooked Lake. In contrast to previous studies, the current study clearly documents autosome-only complexes in both wild and cultivated hybrid hop plants. For example, the ring of four (Figure 2c), double ring (Figure 2c), and the multiple complex (Figure 2d, 1st image) provide examples of non-sex chromosome multiples.

In considering why such observations may not have been reported, we considered several possible reasons. First, previous studies [as reported and reviewed by (Shephard and Parker, 2000)] focused primarily on somatic karyotypic analyses rather than meiotic pairing configurations in late prophase. Secondly, the 3D cytological analyses reported here likely affords a greater opportunity to detect interactions, given the ability to visualize single nuclei and subnuclear regions from multiple perspectives after imaging. We note that the nature of the “interactions” observed are not defined at the molecular

level, but likely represent crossovers. This interpretation is consistent with classical and modern cytology of chiasmata in cell staining preparations, and is supported by the GBS-based segregation data reported here.

### **Variable segregation patterns revealed by linkage analysis of GBS data**

The lack of detailed cytological evidence hinders the correspondence of our linkage groups with the exact meiotic configuration. Moreover, we can not rule out the possibility that rather than a single meiotic configuration, the clustering of markers may depict meiotic events occurring in many nuclei, which were captured by our GBS data. Non-distorted markers in one “super” linkage group may originate in a translocation complex derived from multiple chromosomes. Such complexes could lead directly to several segregation distortion patterns that involve SNP marker groups of variable sizes from small to large. For a multivalent with two terminal crossovers per chromosome, small linkage groups could reflect regions distal to sub-terminal crossovers. Medium linkage groups may reflect normal chromosomes or even partial co-segregation of groups of sub-terminal regions. Large linkage groups could reflect normal large chromosomes or even co-segregation of groups of chromosomes. For example, if chromosomes were arranged in a Renner complex or something comparable, then translocation heterozygous multivalents could assort by copolar cosegregation of every other centromere and the linked loci therein. And by extension, if multivalents are forming in different ways, as our cytogenetics show (Figure 2), then the co-segregation signals would be weaker, but detectable. One way that heterogeneity in complex formation could occur is that common chromosomal regions, such as the abundant sub-telomere repeats HSR1, could mediate synapsis. If they did and also could recombine, that could shuffle the distal segments with loci expected to display some degree of co-segregation instead of independent assortment.

We have observed diverse meiotic configurations in two F1 progeny and two wild hop plants. This raises the possibility that a complex involves a large number of chromosomes, and perhaps the whole genome. Indeed, a complex of at least 4 pairs of chromosomes has been observed in Chimney Rock hop. In addition, a large linkage group was simulated in F1 family “265” by our clustering model. There is a need for additional cytogenetic studies to answer the intriguing question of the largest chromosomal complexes in hop.

Notable in these findings is the fact that we seldom observed one-to-multiple correspondence across the three families. That may reflect the conservation of normally segregating chromosomal parts positioning in the chromosomal complexes and invariable occurrence of the structural translocation heterozygosity in the progenitors of the three families.

**Structural polymorphism and variable meiotic chromosome interactions may contribute to segregation distortion in hop**

At least 57 species of flowering plants are characterized by permanent translocation heterozygotes (Holsinger and Ellstrand, 1984). For instance, in *Clarkia* ( $2n = 18$ ) chromosomal polymorphisms, such as structural heterozygosity, has been observed in nearly half of the 34 known species (Snow, 1960). Similarly, natural and distinct cytotypes or chromosome races have been well described in Asteraceae family (Carr and Carr, 1983). In that study, structurally heterozygous individuals were found to occur within natural populations with various cytotypes or chromosome races being characterized on the basis of meiotic pairing configurations (Carr, 1977). Given that segregation distortion is a ubiquitous phenomenon in hop [(Seefeldt et al., 2000; McAdam et al., 2013) and extended by our study], together with multiple examples of naturally occurring chromosomal polymorphism in other plant genera, we favor the idea

that phenomena such as structural heterozygosity and segmental aneuploidy might play important roles in the population dynamics of hop.

Translocation heterozygosity may have an important connection to the significantly distorted sex-ratio in favor of females in hop. Likewise, female-biased sex ratios have been found in Mistletoe, another notable dioecious case of translocation heterozygosity. To maintain heterozygosity, *Oenothera*, a notable monoecious case of translocation heterozygosity, utilizes a system of balanced lethal to purge the lethal homozygotes (Steiner, 1956; Harte, 1994), which is referred to as “recessive lethals”. In the context of XY system, heteromorphism of sex chromosomes dictates that males are more severely affected than females by “X-linked recessive lethals”, because males only have one copy of the X chromosome. Hence, *H. lupulus* may use a system of balanced lethals at the expense of male offspring to preserve genetic heterozygosity.

Our results are compelling for translocation heterozygosity studies in light of high-density molecular markers in many other biota. For example, such large scale recombination suppression is also presented in at least 10 species of termite, some types of centipede, and perhaps all of the monotremes (Holsinger and Ellstrand, 1984; Rowell, 1987; Rens et al., 2004). Beyond homologous crossover, translocation heterozygosity has shown considerable evolutionary interest and selective advantage in its own right.

In future studies, it will be important to further characterize these genomes for evidence of structural polymorphisms and to explore the mechanistic underpinnings and biological consequences of these phenomena in hop. Investigations should include FISH to track specific loci through meiosis and into both post-meiotic daughter cells and the next generation; pollen-based assays expected to reveal aspects of grain viability and fertility; and continued structural and comparative genomics to directly resolve presumed points of

chromosomal breakage, which could uncover specific deletions, duplications, inversions, or translocations. Given the potential genetic and genomic complexities within and between hop species, future progress and investigation of questions from this and prior studies is a significant challenge requiring the integration of multiple disciplines and lines of evidence from a variety of different experiments in domestic and wild hops.

### **Perspectives of breeding strategies in hop**

Understanding genetic recombination is essential for speed and accuracy of plant breeding. Indeed, it is generally difficult to breed new commercial hop varieties through mass selection and crossing. Our findings show that a large scale, perhaps genome-wide, atypical meiotic chromosome behavior may be common in hop. Translocation heterozygosity can extend linkage to nonhomologous chromosomes, and favor severe segregation distortion accumulated near the translocation breakpoints (Taylor and Ingvarsson, 2003; Farré et al., 2011). Such a high degree recombination of suppression may hinder effective selection of desired allele combinations that make use of marker-assisted selection (MAS) based on Mendelian segregation patterns.

We found intriguing differences in population differentiation in the surrounding regions of the two *CHS* homologues (Figure 9), possibly resulting from genetic introgression and differential allele selection from domestication towards higher alpha acid yields.

Hence, targeted resequencing and mapping the “consensus” genomic regions that segregate appropriately may deserve emphasis in hop. A “normal” reference genome may be essential to elucidate structural differences arising from rearrangement events. *In silico* screening of primers/enzymes to avoid the regions with the tendency of segregation distortion may fulfill the purpose of cost-effective genotyping platforms in hop’s breeding programs.

**Additional files**

**Supplementary Figures.** The file contains supplementary figure S1-S7.

**Supplementary Table S1** Pedigrees of genotyped F1 populations.

**Supplementary Table S2** Cultivar and landrace accessions.

**Supplementary Table S3** Wild exotic accessions.

**Supplementary Table S4** 588 sex association ( $P \leq 10^{-10}$ ) SNPs. Scaffold, position, P value and MAF are indicated.

**Supplementary Table S5** SNPs with  $F_{ST} \geq 0.5$  in pairwise comparisons of var. *neomexicanus*, var. *lupuloides* and CV.

HapMap SNPs can be accessed at

<https://hopsteiner.app.box.com/s/r0tzqpdzcagvmxtxducy21lrdukuhbd1>.

**Acknowledgments**

We thank Buckler lab and Qi Sun's group at Cornell for helpful discussions. We thank the growers at Golden Gate ranches for cultivation of experimental plants.

## References

- van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE** (2011) The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* **12**: R102
- Bass HW, Marshall WF, Sedat JW, Agard DA, Cande WZ** (1997) Telomeres Cluster De Novo before the Initiation of Synapsis: A Three-dimensional Spatial Analysis of Telomere Positions before and during Meiotic Prophase. *J Cell Biol* **137**: 5–18
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E** (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* **10008**: 6
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES** (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–5
- Bradshaw JE** (2016) Use of Sexual Reproduction in Base Broadening and Introgression. *Plant Breed. Past, Present Futur*. Springer. pp 483–527
- Carr GD** (1977) A Cytological Conspectus of the Genus *Calycadenia* (Asteraceae ): An Example of Contrasting Modes of Evolution. *Am J Bot* **64**: 694–703
- Carr RL, Carr GD** (1983) Chromosome Races and Structural Heterozygosity in *Calycadenia ciliosa* Greene (Asteraceae). *Am J Bot* **70**: 744–755
- Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood RJ, Guarino L, Harker RH, Jarvis A, Maxted N, et al** (2016) Global conservation priorities for crop wild relatives. *Nat Plants* **2**: 1–6
- Divashuk MG, Alexandrov OS, Kroupin PY, Karlov GI** (2011) Molecular cytogenetic mapping of *Humulus lupulus* sex chromosomes. *Cytogenet Genome Res* **134**: 213–219
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE** (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**: 1–10



- Farré A, Benito IL, Cistué L, de Jong JH, Romagosa I, Jansen J** (2011) Linkage map construction involving a reciprocal translocation. *Theor Appl Genet* **122**: 1029–1037
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES** (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One*. doi: 10.1371/journal.pone.0090346
- Golczyk H, Massouh A, Greiner S** (2014) Translocations of chromosome end-segments and facultative heterochromatin promote meiotic ring formation in evening primroses. *Plant Cell* **26**: 1280–93
- Gruber K** (2016) Re-igniting the green revolution with wild crops. *Nat Plants* **2**: 1–4
- Harte C** (1994) *Oenothera Contributions of a Plant to Biology*. Monogr Theor Appl Genet. doi: 10.1017/CBO9781107415324.004
- Haunold A** (1991) Cytology and cytogenetics of Hops. In: T Tsuchiya and PK Gupta (Eds.), *Chromosome Engineering in Plants; Genetics, Breeding, Evolution*. Elsevier, New York, pp. 551-563.
- Henning J, Hill S, Darby P, Hendrix D** (2017) QTL examination of a bi-parental mapping population segregating for “short-stature” in hop (*Humulus lupulus* L.). *Euphytica* **213**: 77
- Henning JA, Gent DH, Twomey MC, Townsend MS, Pitra NJ, Matthews PD** (2015) Precision QTL mapping of downy mildew resistance in hop (*Humulus lupulus* L.). doi: 10.1007/s10681-015-1356-9
- Hill ST, Coggins J, Liston A, Hendrix D, Henning JA** (2016) Genomics of the hop pseudo-autosomal regions. *Euphytica*. doi: 10.1007/s10681-016-1655-9
- Holsinger KE, Ellstrand NC** (1984) The Evolution and Ecology of Permanent Translocation Heterozygotes. *Am Nat* **124**: 48–71
- Howe ES, Murphy SP, Bass HW** (2013) Three-Dimensional Acrylamide Fluorescence In Situ Hybridization for Plant Cells. *Methods Mol. Biol.* pp 53–66

- Hyma KE, Barba P, Wang M, Londo JP, Acharya CB, Mitchell SE, Sun Q, Reisch B, Cadle-Davidson L** (2015) Heterozygous Mapping Strategy (HetMappS) for High Resolution Genotyping-By-Sequencing Markers: A Case Study in Grapevine. *PLoS One*. doi: 10.1371/journal.pone.0134880
- Jakse J, Stajner N, Kozjak P, Cerenak A, Javornik B** (2008) Trinucleotide microsatellite repeat is tightly linked to male sex in hop (*Humulus lupulus* L.). *Mol Breed* **21**: 139–148
- Karlov GI, Danilova T V., Horlemann C, Weber G** (2003) Molecular cytogenetics in hop (*Humulus lupulus* L.) and identification of sex chromosomes by DAPI-banding. *Euphytica* **132**: 185–190
- Laursen L** (2015) Botany: The cultivation of weed. *Nature* **525**: S4–S5
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z** (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**: 2397–9
- Maaten L Van Der, Hinton G** (2008) Visualizing Data using t-SNE. *J Mach Learn Res* **9**: 2579–2605
- Matthews PD, Coles MC, Pitra NJ** (2013) Next Generation Sequencing for a Plant of Great Tradition: Application of NGS to SNP Detection and Validation in Hops (*Humulus lupulus* L.). *BrewingScience* **66**: 185–191
- McAdam EL, Freeman JS, Whittock SP, Buck EJ, Jakse J, Cerenak A, Javornik B, Kilian A, Wang C-H, Andersen D, et al** (2013) Quantitative trait loci in hop (*Humulus lupulus* L.) reveal complex genetic architecture underlying variation in sex, yield and cone chemistry. *BMC Genomics* **14**: 360
- Ming R, Bendahmane A, Renner SS** (2011) Sex Chromosomes in Land Plants. *Annu Rev Plant Biol* **62**: 485–514
- Miranda CL, Elias VD, Hay JJ, Choi J, Reed RL, Stevens JF** (2016) Xanthohumol improves dysfunctional glucose and lipid metabolism in diet-induced obese

C57BL/6J mice. *Arch Biochem Biophys* 1–9

**Murakami A, Darby P, Javornik B, Seigner E, Lutz A, Svoboda P** (2006) Molecular phylogeny of wild Hops , *Humulus lupulus* L . *Heredity (Edinb)* **97**: 66–74

**Nagel J, Culley LK, Lu Y, Liu E, Matthews PD, Stevens JF, Page JE** (2008) EST Analysis of Hop Glandular Trichomes Identifies an O-Methyltransferase That Catalyzes the Biosynthesis of Xanthohumol. *Plant Cell* **20**: 186–200

**Natsume S, Takagi H, Shiraishi A, Murata J, Toyonaga H, Patzak J, Takagi M, Yaegashi H, Uemura A, Mitsuoka C, et al** (2015) The Draft Genome of Hop (*Humulus lupulus*), an Essence for Brewing. *Plant Cell Physiol* **0**: 1–14

**Neve RA** (1958) Sex Chromosomes in the Hop *Humulus lupulus*. *Nature* **181**: 1084 – 1085

**Neve RA** (1991) Hops. London Chapman Hall

**Ososki AL, Kennelly EJ** (2003) Phytoestrogens: a review of the present state of research. *Phyther Res* **17**: 845–869

**Rauwolf U, Golczyk H, Meurer J, Herrmann RG, Greiner S** (2008) Molecular Marker Systems for *Oenothera* Genetics. *Genetics* **180**: 1289–1306

**Reeves PA, Richards CM** (2011) Species Delimitation under the General Lineage Concept: An Empirical Example Using Wild North American Hops (*Cannabaceae*: *Humulus lupulus*). *Syst Biol* **60**: 45–59

**Rens W, Grützner F, O'brien PCM, Fairclough H, Graves JAM, Ferguson-Smith MA** (2004) Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *Proc Natl Acad Sci U S A* **101**: 16257–16261

**Roweis ST, Saul LK, Roweis ST** (2000) Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **290**: 2323–2326

**Rowell DM** (1987) Complex sex-linked translocation heterozygosity: Its genetics and biological significance. *Trends Ecol Evol* **2**: 242–246

- Seefeldler S, Ehrmaier H, Schweizer G, Seigner E** (2000) Male and female genetic linkage map of hops *Humulus lupulus*. *Plant Breed* **119**: 249–255
- Shepherd H, Parker J** (2000) Sexual development and sex chromosomes in hop. *New Phytol* **148**: 397–411
- Sinotô Y** (1929) Chromosome Studies in Some Dioecious Plants, with Special Reference to the Allosomes. *Cytologia (Tokyo)* **1**: 109–191
- Siragusa GR, Haas GJ, Matthews PD, Smith RJ, Buhr RJ, Dale NM, Wise MG** (2008) Antimicrobial activity of lupulone against *Clostridium perfringens* in the chicken intestinal tract jejunum and caecum. *J Antimicrob Chemother* **61**: 853–858
- Snow R** (1960) Chromosomal Differentiation in *Clarkia dudleyana*. *Am J Bot* **47**: 302–309
- Steiner E** (1956) New aspects of the balanced lethal mechanism in oenothera. *Genetics*
- Stevens JF, Page JE** (2004) Xanthohumol and related prenylflavonoids from hops and beer: To your good health! *Phytochemistry* **65**: 1317–1330
- Stone E a., Ayroles JF** (2009) Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet.* doi: 10.1371/journal.pgen.1000479
- Tanksley SD, McCouch SR** (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**: 1063–1066
- Taylor DR, Ingvarsson PK** (2003) Common features of segregation distortion in plants and animals. *Genetica* **117**: 27–35
- Vyskot B, Hobza R** (2004) Gender in plants: Sex chromosomes are emerging from the fog. *Trends Genet* **20**: 432–438
- Wiens D, Barlow BA** (1975) Permanent Translocation Heterozygosity and Sex Determination in East African Mistletoes. *Science* **187**: 1208–1209
- Winge O** (1929) Critical remarks to Y. Sinoto's paper on a tetrapartite sex chromosome complex in *Humulus*. *Hereditas* **12**: 269–270
- Wu Y, Bhat PR, Close TJ, Lonardi S** (2008) Efficient and accurate construction of

genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. doi: 10.1371/journal.pgen.1000212

### Figure Legends

**Figure 1 Population structure of 251 hop accessions and geographic origins of the U.S. wild hop.** 183 modern cultivars are indicated by red color. 68 wild hop are color-coded by geographic origins. (a) Neighbor-joining tree of the 251 hop accessions. (b) The state names are followed by sample counts. Three state groups (“MT, ND, SD, NE, IA, KS, MO”, “CO, AZ, NM” and “MA”) are color-coded to distinguish from one another.

**Figure 2 3D cytology of hop chromosomes from pollen mother cells at diakinesis.** For cytogenetic analysis of hop meiotic chromosomes, male panicles were fixed in Carnoy’s solution then formaldehyde. Meiocytes were extruded from anthers, placed on glass slides, stained with DAPI, and imaged by 3D deconvolution microscopy. Through-focus maximum intensity projections are shown for whole nuclei in panel A or 3D-cropped chromosomes in panels B-D. (a) Representative diakinesis nuclei are shown for two wild and two hybrid plants. Plant IDs and bivalent frequencies per nucleus are shown under each panel along with the total number of full 3D nuclei analyzed. Examples of ring bivalents are shown (arrows) and scale bars are indicated in microns. (b) Bivalent examples are shown and classified into types (“Ring”, “Sex (XY)” or “NOR-linked”) listed under each panel and the nucleolus (n) is indicated in the NOR-linked example. Examples of presumed chiasmata (crossovers) are indicated (arrowheads) and they show the typical appearance as small gaps or spaces. (c) Quadrivalent examples are shown and classified into types (“Double ring”, “Ring of four” or “NOR-linked plus X”) listed under each panel. An interpretive tracing of the NOR-linked plus X quadrivalent shows the nucleolus (blue), NOR-linked bivalent

(green), the sex chromosome X (yellow) and the sex chromosome Y (red). (d) Examples of other complexes involving multiple chromosomes of unknown composition are shown along with general descriptions (“Multiple” or “Long chain”) under each panel. The first three images show 3D-cropped regions that capture entire complexes. The last panel shows an entire nucleus with a long chain (LC) configuration of complex that winds around in space, along with two separate nucleoli.

**Figure 3 Pseudo-testcross (Pt) schema.** (a) SNP sites used in the testcross are color coded. Minor alleles are segregated either from grandparent1 (GP1) (green), or from GP2 (red). In other words, linkage groups of grandparents are joined with their phases in repulsion. Two phases are indicated by colors of green and red individually. Markers in coupling and repulsion are distinguished by positive and negative correlation individually. (b) Correlation coefficient-based clustering and spatial coordinates of Pt markers. We used two methods, Louvain modularity and locally linear embedding, to cross-check the clustering patterns of markers without and with inclusion of segregation distortion (SD). Mendelian segregation markers are enclosed by blue and red frames, and SD markers are enclosed in a yellow frame. See Methods for more details.

**Figure 4 Linkage groups for the maternal line of family “144” and correspondence across three genetic map sets.** The degrees of Spearman’s correlation ( $\rho$ ) are color-coded. (a) Unphased and phased (linkage for grandparents) groups are bounded by white and black frames individually. Alignment of unphased groups (b) between “144” and “247” and (c) between “144” and “265”. The markers in alignments are indexed by pseudo-chromosomal positions. The alignments demonstrate the consistency of clustering patterns of the common markers across “144”, “247” and “265”.

**Figure 5 Linkage of Mendelian ( $15\% \leq \text{MAF} \leq 30\%$ ) and non-Mendelian Pt markers ( $5\% \leq \text{MAF} < 15\%$ ), based on Spearman's correlation ( $\rho$ ).** In each sub-figure, clustering patterns without (left) and with (right) inclusion of segregation distortion are presented by LLE (top) and the Louvain Modularity (bottom). Mendelian markers in two linkage groups are indicated by blue and red colors individually. Segregation distortion (SD) markers are indicated by yellow color. Correlation map (a) of LG1.1 and LG4.1 in maternal linkage of cross “144”, (b) of LG2.1 and LG8.1 in maternal linkage of cross “247”, (c) of LG10.1 and LG10.2 in maternal linkage of cross “265”, and (d) of LG2.1 and LG2.2 in maternal linkage of cross “265”.

**Figure 6 Linkage patterns of the 5 largest linkage groups in family “265”, based on spatial coordinates defined by LLE.** (a) Linkage groups are color-coded. (b) Markers with non-Mendelian frequencies (cyan, for  $0.15 \leq \text{MAF} < 0.2$ ) versus Mendelian frequencies (grey, for  $0.2 \leq \text{MAF} \leq 0.3$ ) are co-plotted.

**Figure 7 One-to-two genetic correspondence between “144” and “265”.** (a) LG2.1 in “144” corresponds to LG2.1 and LG2.2 in “265”. Two instances of one-to-one correspondence (LG1.1-LG1.2 and LG3.1-LG3.1) are added for control. Spatial representations (XYZ coordinates) of linkage groups in (b) “265” and (c) “144” were derived from LLE.

**Figure 8 Association studies and  $F_{ST}$  mapping of sex determination in hop.** (a) Linkage group-based Manhattan-plot of MLM for sex determination in family “247” ( $N = 364$ ,  $N_{\text{male}} = 30$ ). Light and deep colors are used to distinguish two phases (linkage for

grandparents) in coupling. (b) Manhattan-plot of  $F_{ST}$  in females vs. males in “247”. (c) Log Quantile-Quantile (QQ) plot of 356,526 association tests (SNPs) for sex determination in 850 individuals ( $N_{\text{male}} = 129$ ,  $N_{\text{female}} = 721$ ). (d) Correlation among 588 association ( $P \leq 10^{-7}$ ) markers, the proportions of 588 markers in LG4, other LGs and unmapped data set.



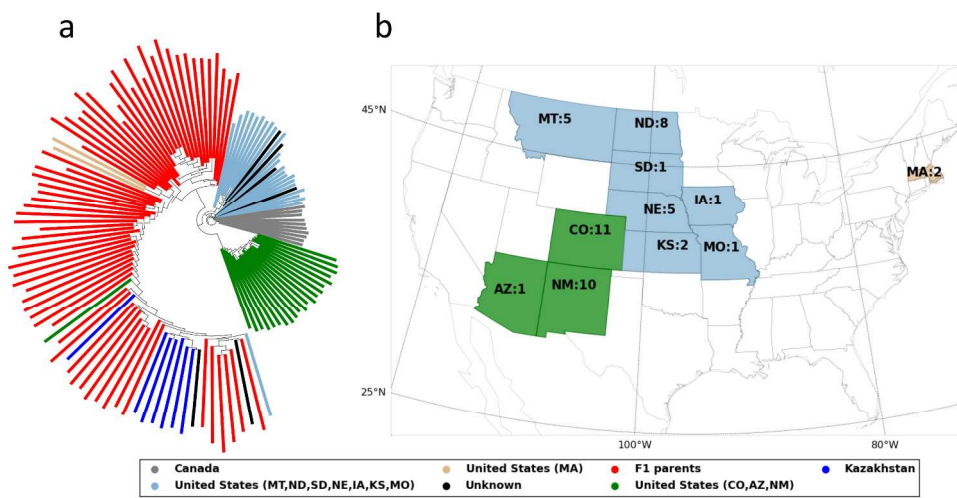


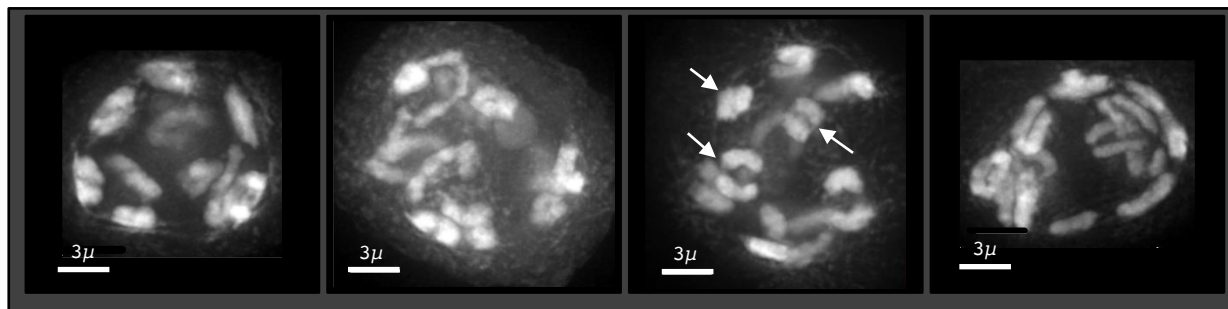
Figure 1 Population structure of 251 hop accessions and geographic origins of the U.S. wild hop.

233x123mm (300 x 300 DPI)

**Whole nucleus projections, male diakinesis****a**

Wilds

Hybrids



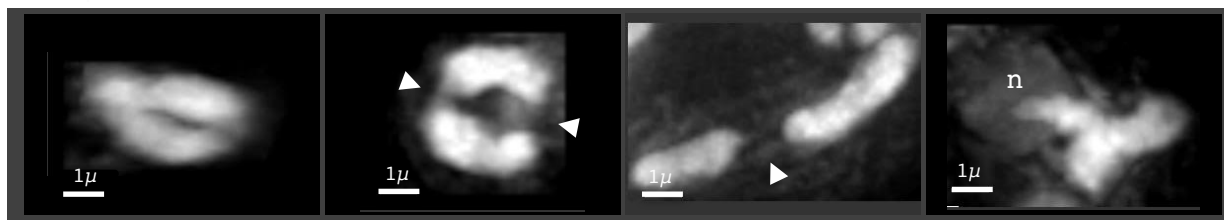
**Plant ID**  
**Variety**  
**Bivalents**

Crooked Lake  
*lupuloides*  
avg. 5.7 (n=20)

Chimney Rock  
*neomexicanus*  
avg. 2.2 (n=25)

Cross 265  
Hybrid  
avg. 4.7 (n=21)

Cross 255  
Hybrid  
avg. 6.0 (n=20)

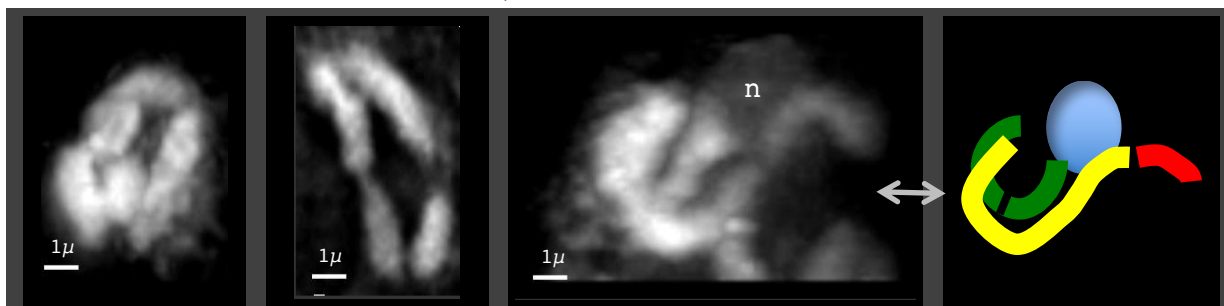
**b****Bivalents****Type:**

Ring

Ring

Sex (XY)

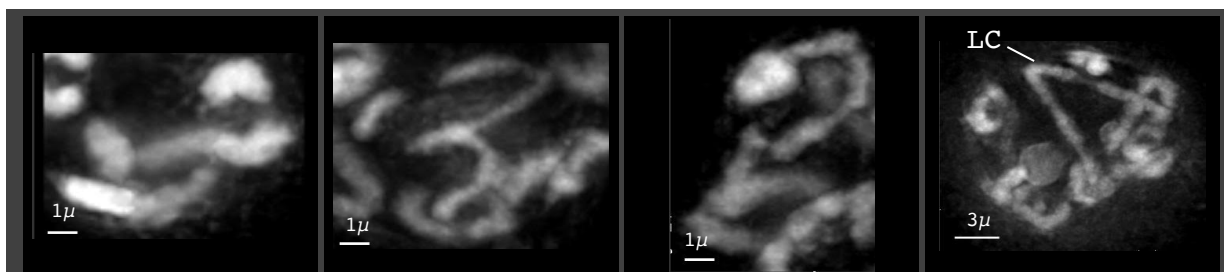
NOR-linked

**c****Quadrivalents****Type:**

Double ring

Ring of four

NOR-linked plus X

Interpretive drawing of  
NOR-linked plus X**d****Other complexes, variable composition****Type:**

Multiple

Multiple

Multiple

Long chain

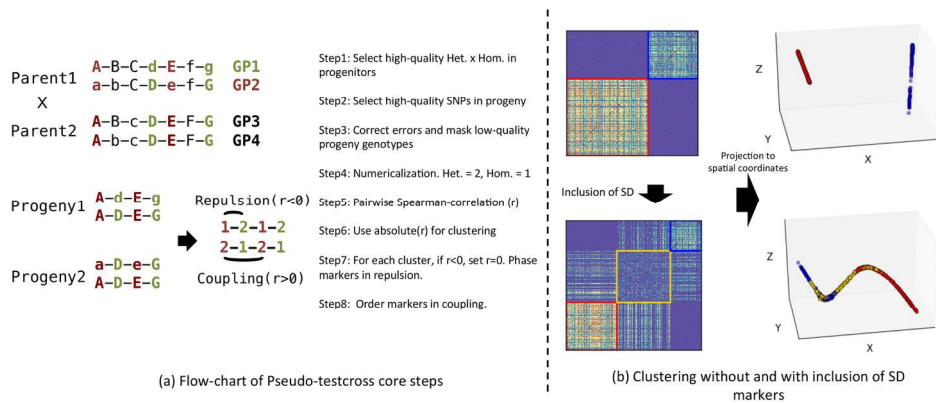


Figure 3 Pseudo-testcross (Pt) schema.

175x73mm (300 x 300 DPI)

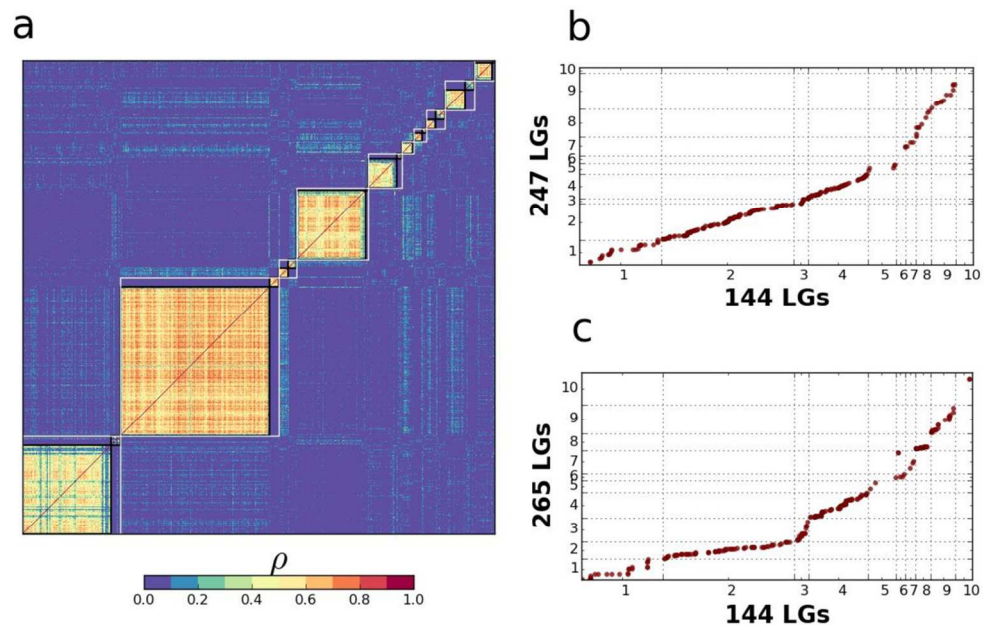


Figure 4 Linkage groups for the maternal line of family "144" and correspondence across three genetic map sets.

121x77mm (300 x 300 DPI)

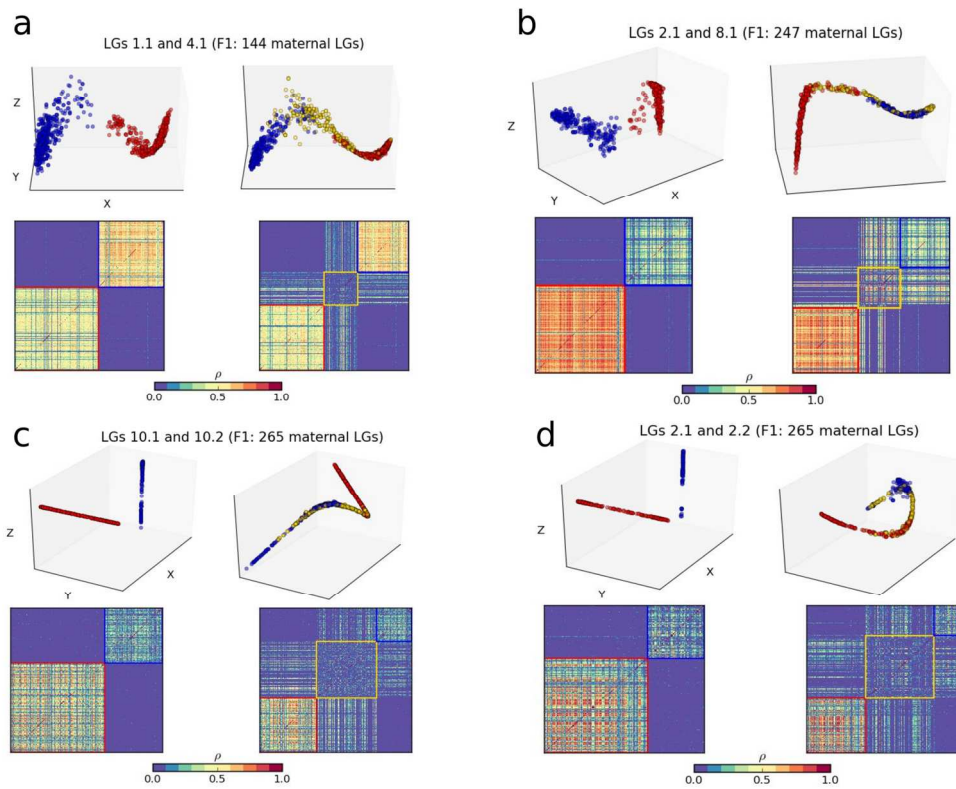


Figure 5 Linkage of Mendelian ( $15\% \leq \text{MAF} \leq 30\%$ ) and non-Mendelian Pt markers ( $5\% \leq \text{MAF} < 15\%$ ), based on Spearman's correlation ( $\rho$ ).

151x120mm (300 x 300 DPI)

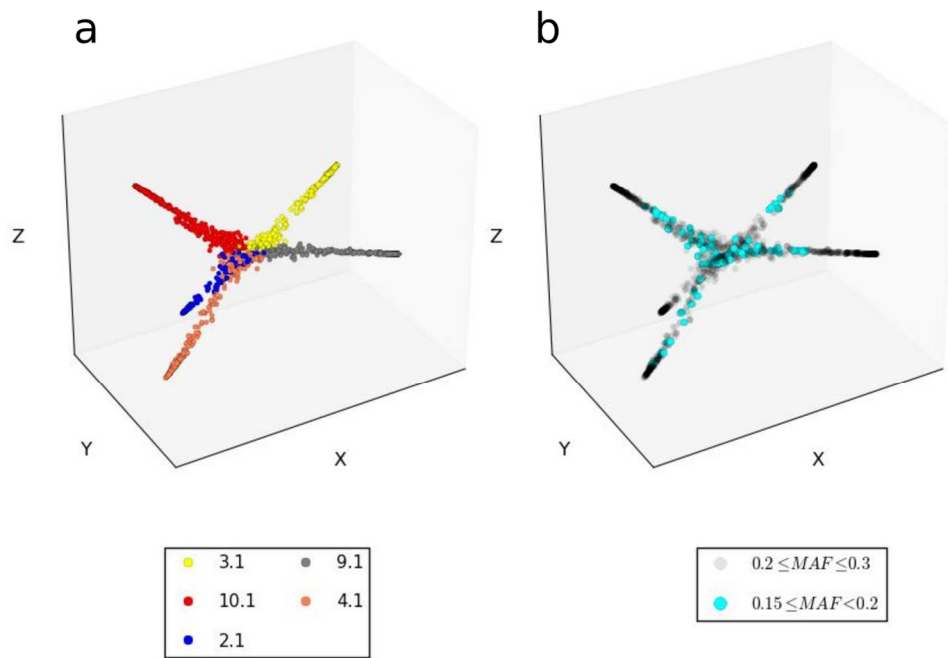


Figure 6 Linkage patterns of the 5 largest linkage groups in family "265", based on spatial coordinates defined by LLE.

130x89mm (300 x 300 DPI)

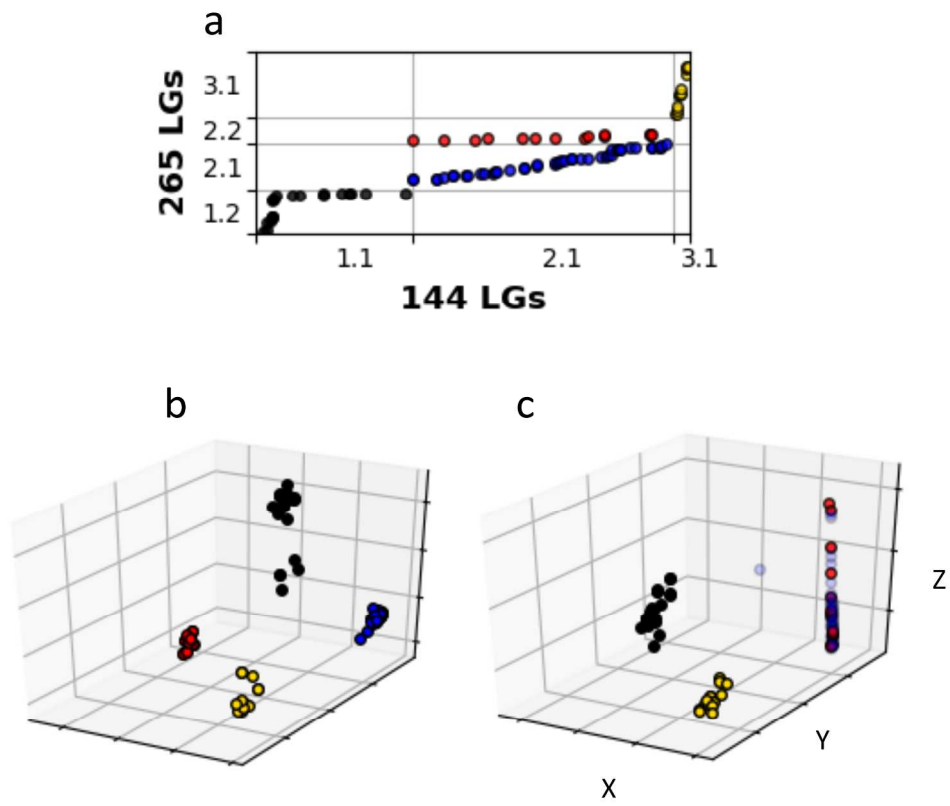


Figure 7 One-to-two genetic correspondence between "144" and "265".

216x180mm (300 x 300 DPI)

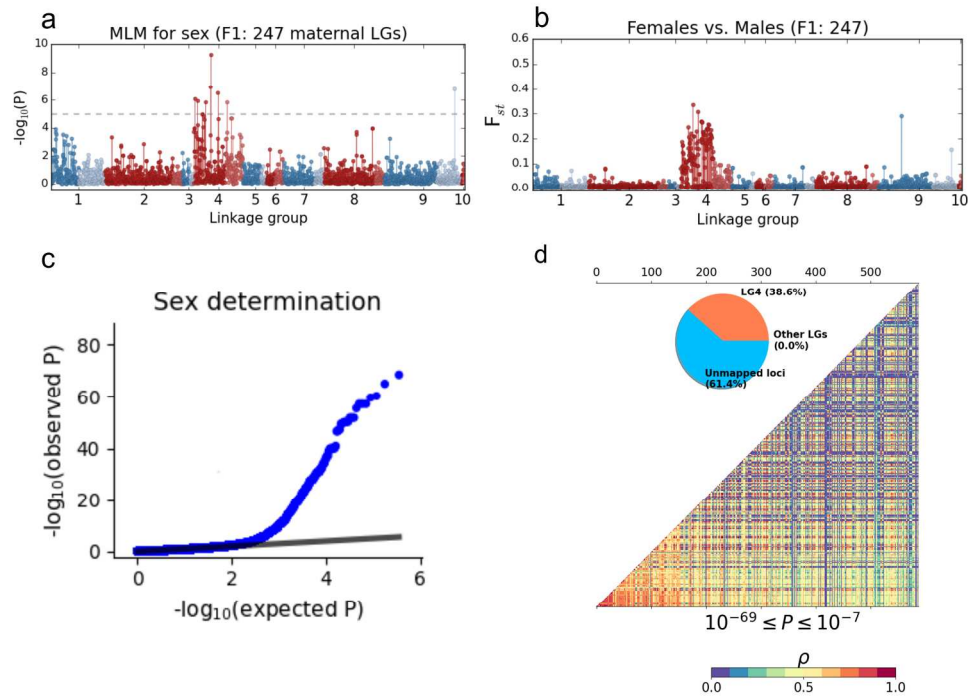
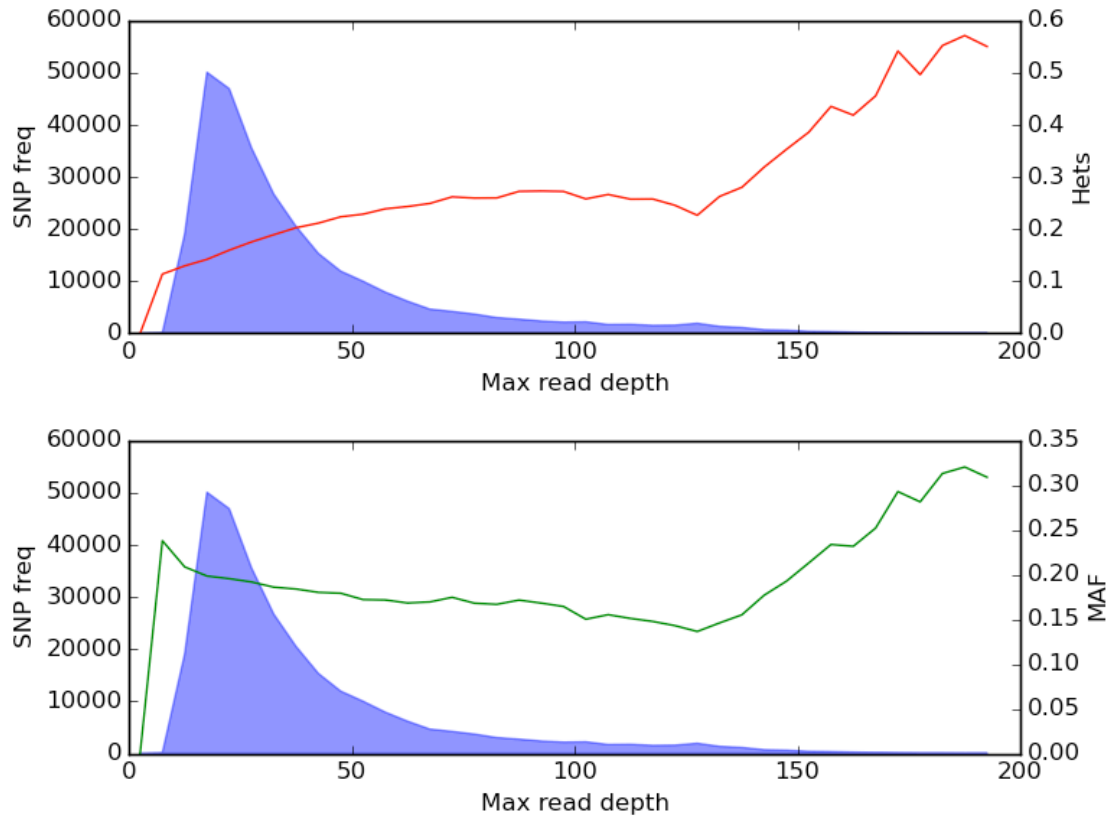


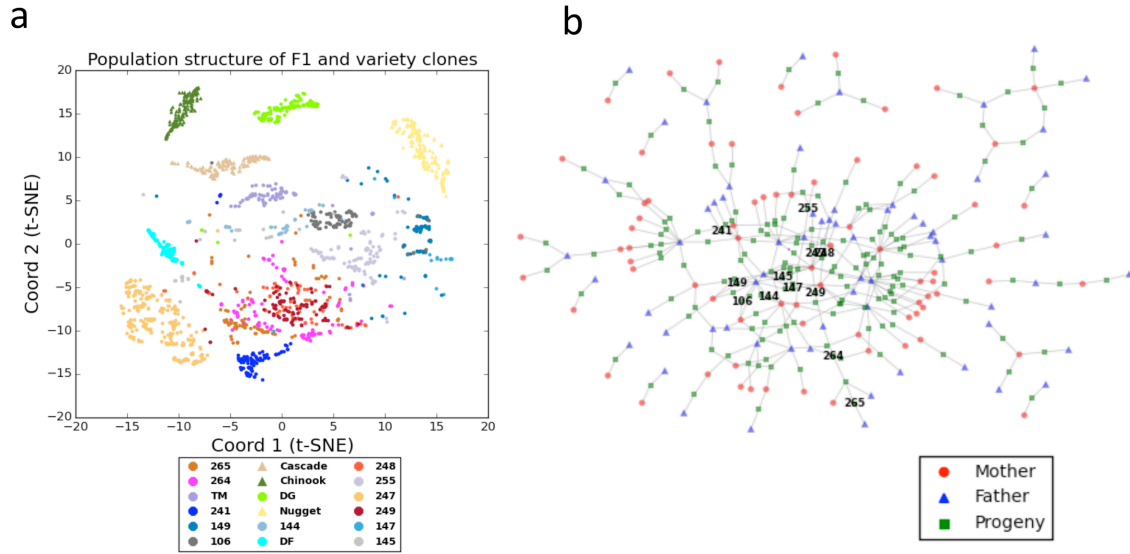
Figure 8 Association studies and FST mapping of sex determination in hop.

189x137mm (300 x 300 DPI)

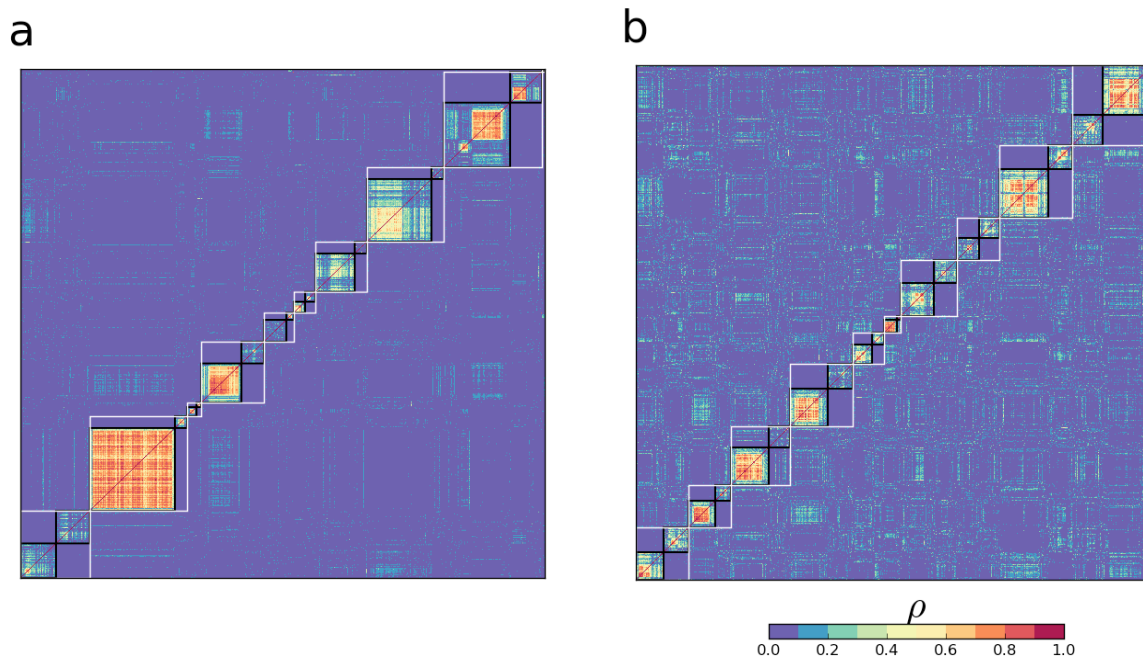




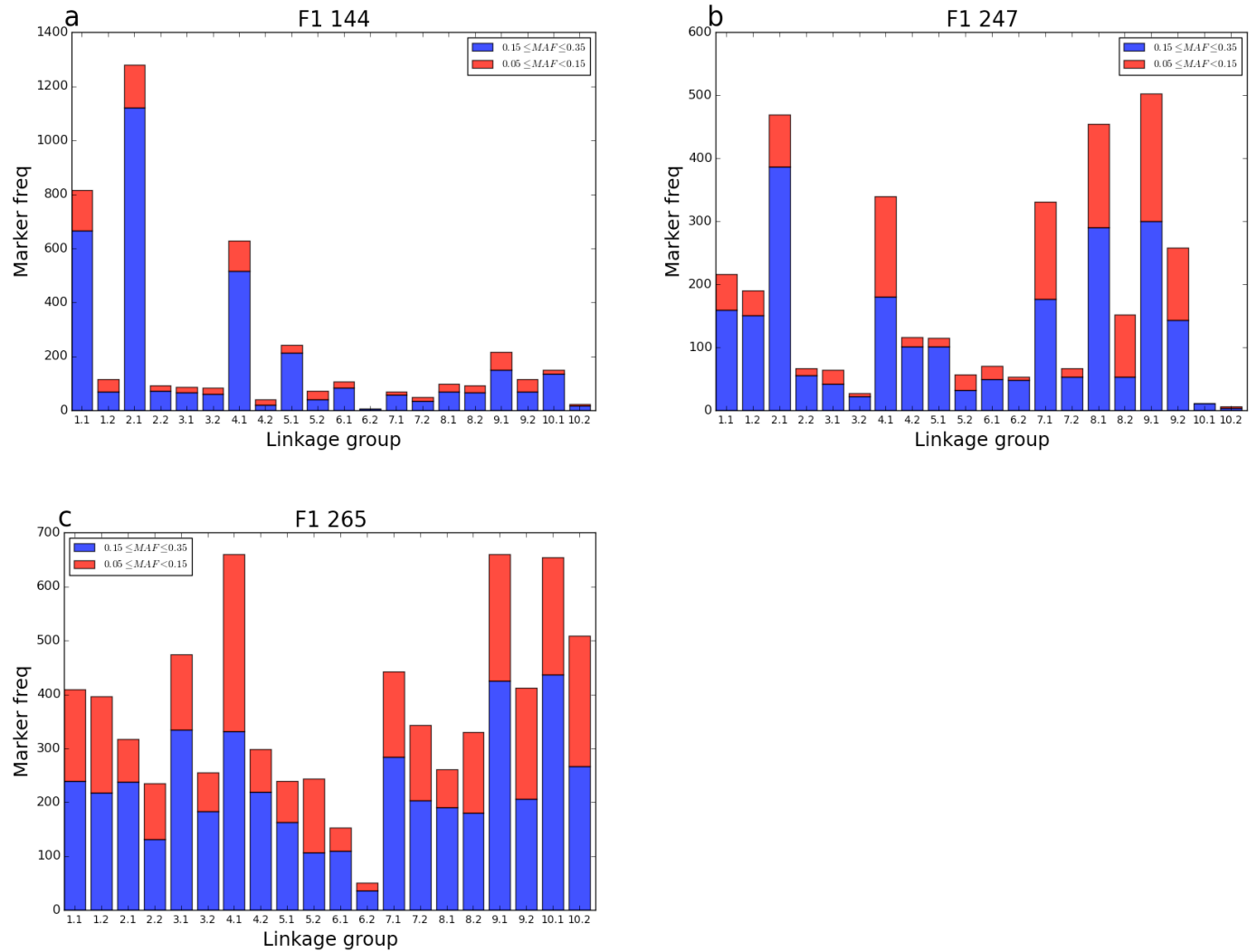
**Figure S1 Correlation between the max read depth of one SNP site with the heterozygosity ratio (denoted by red curve) and MAF (denoted by green curve). Blue filled curves show the correlation between max read depths with SNP frequencies.**



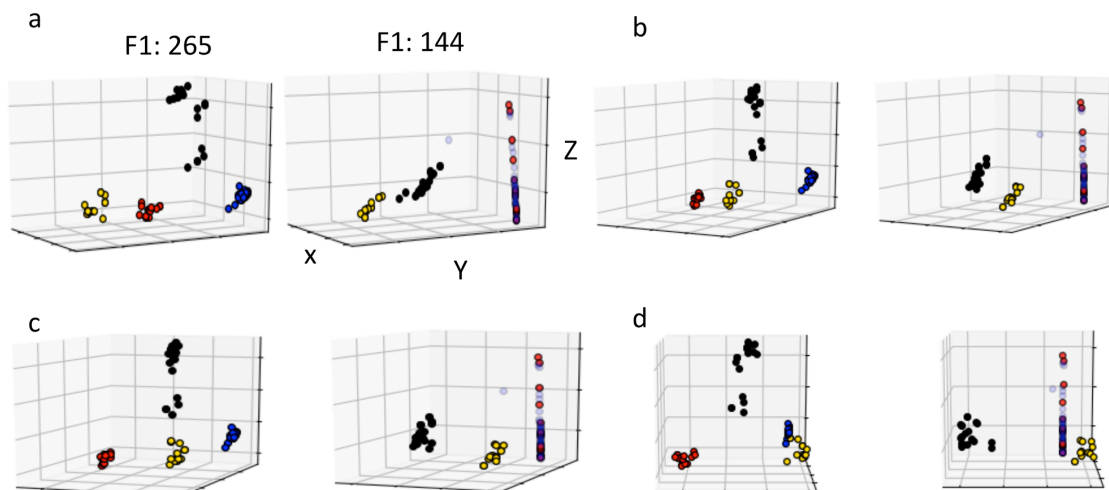
**Figure S2 Population structure and pedigree network of GBS data.** (a) t-SNE plot for F1 families (circle) and variety clones (triangle) ( $N \geq 60$ ). (b) The overview of pedigree for GBS data. F1 families ( $N \geq 60$ ) are denoted.



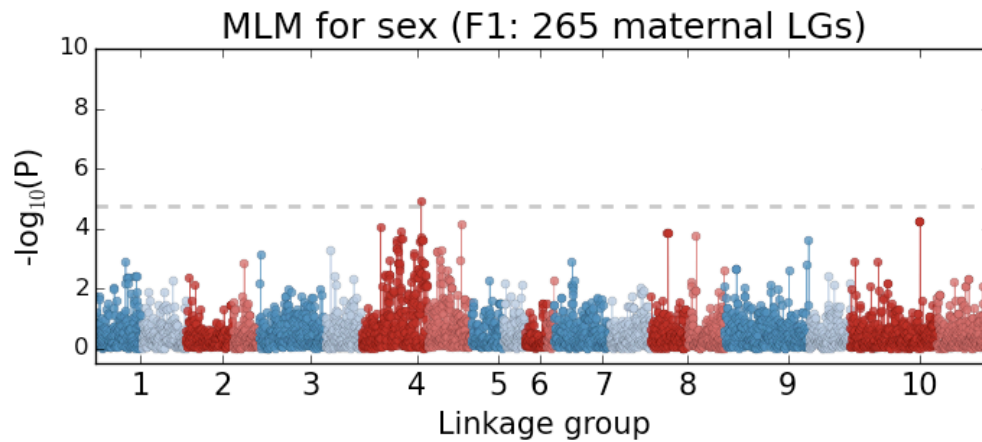
**Figure S3 Linkage groups for the maternal lines of families (a) “247” and (b) “265”.** Unphased and phased (linkage for grandparents) groups are bounded by white and black frames individually. The degrees of Spearman’s correlation ( $\rho$ ) are color-coded.



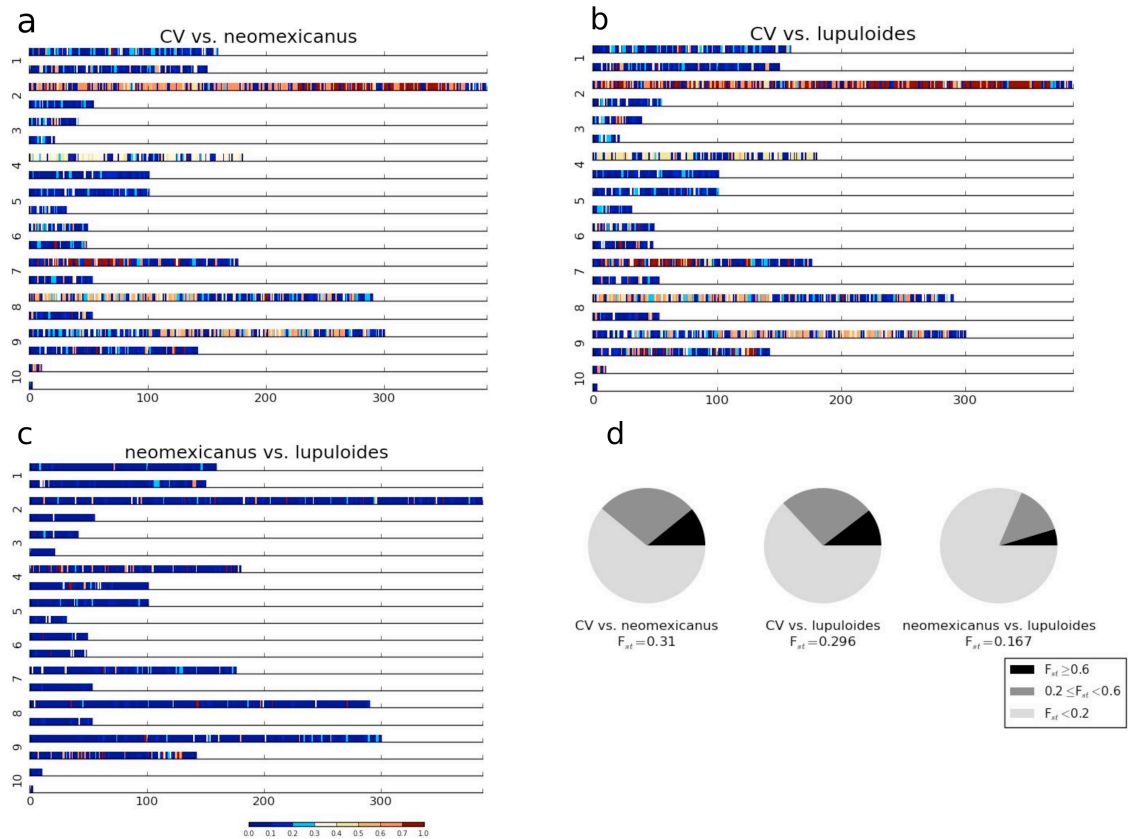
**Figure S4 Genome-wide views of the segregation distortion in three F1 families.** Mendelian markers ( $15\% \leq MAF \leq 35\%$ ) and correlated ( $\rho \geq 0.3$ ) segregation distortion ( $5\% \leq MAF < 15\%$ ) are represented by blue and red bars individually. F1 families (a) “144”. (b) “247”. (c) “265”.



**Figure S5 Clustering patterns (derived from Locally-Linear Embedding method) of linkage group (LG) 1.2 (black), 2.1 (blue), 2.2 (red), 3.1 (yellow) in cross “265” (left axes) and linkage group LG 1.1 (black), 2.1 (blue+red), 3.1(yellow) in cross “144” (right axes). The initial azimuth (XY plane) angles are (a) 30, (b) 60, (c) 70 and (d) 90.**



**Figure S6** Association studies of sex determination in the F1 family “265” ( $N = 95$ ,  $N_{\text{male}} = 13$ ). Linkage group-based Manhattan-plot of MLM. Light and dark colors are used to distinguish two phases (linkage for grandparents) in coupling.



**Figure S7 Linkage group (in family "247")-based  $F_{st}$  heatmaps and the overall  $F_{st}$  distribution.** Population differentiation (a) between modern cultivars (CV) and var. *neomexicanus*; (b) between CV and var. *lupuloides*; (c) between var. *neomexicanus* and var. *lupuloides*. (d) Spectrum of the overall  $F_{st}$  distribution.

**Pedigrees of genotyped crosses**

<b>Cross</b>	<b>Mother</b>	<b>Father</b>
1	Cascade	ZoV_0441&2M
104	Cascade	Male62
105	Serebrianka	Male50
106	Zenith	Male50
108	Apollo	ZgY_0449M
109	Toyamidori	ZgY_0449M
110	Saaz	WillaM._Male_(Toppenish)
111	Tettnang	WillaM._Male_(Toppenish)
112	Hallertauer_N	WillaM._Male_(Toppenish)
113	Hallertauer_G	WillaM.Male_(Sunnyside)
114	Kent_Golding	WillaM.Male_(Sunnyside)
115	Cascade	WillaM.Male_(Sunnyside)
117	_04201	(B)04F1/46
118	_04201	(E)03O1/91
119	_04201	(F)04M4/19
120	_04217	(B)04F1/46
121	_04222	(A)03N2/40
122	_04222	(B)04F1/46
123	_04222	(C)04G38/10
125	_04222	(E)03O1/91
132	ZgZa_dwf	ZgY_0449M
133	SZL	ZgY_0449M
134	_04175	ZgY_0449M
135	Hallertauer_T	Male56
136	Hallertauer_T	Male56
137	Kent_Golding	Male56
138	Cascade	Male47
139	Delta	Male47
140	_04190	Male47
141	Wye_Target	Male50
143	Cascade	Male50
144	Nugget	Male50
145	Super_Galena	Male50
147	Wye_Target	_035_0648M
148	Zenith	_035_0648M
149	Serebrianka	_035_0648M
15	Nugget	DAM



150	Cascade	_035_0648M
151	Super_Galena	_035_0648M
152 1-05	Super_Galena	_035_0648M
152 1-13	Super_Galena	_035_0648M
152 1-18	Super_Galena	_035_0648M
159	_05235	_070_0752M
16	Nugget	Male54
165	_05237	Male54
167	_05237	_052_0754M
178	_04229	Male54
179	Galena	MV_0437M
18	Newport	DAM
190	Apollo	133_0864M
211	Merkur	Male50
212	Merkur	Male20
216	Serebrianka	Male66
217	Apollo	187C_1075M
218	Bravo	Male22
219	Galena	179B_1074M
220	Hallertauer_M	Male47
221	Hersbrucker_I	Male47
222	Saphir	Male47
223	Calypso	Male47
224	Centennial	Male47
234	Apollo	19/93/18
235	Nugget	19/93/18
236	Cascade	19/93/18
238	Delta	19/93/18
24	Merkur	Canada_2M
241	Cascade	Male11
247	Super_Galena	Male15
248	Super_Galena	_075_0778M
249	Super_Galena	_075_0779M
25	Merkur	IZdM
255	_07270	19/93/18
26	Merkur	DAM
264	Chinook	_075_0779M
265	Chinook	Male57
28	Eastern_Gold	IZdM
31	Eastern_Gold	Male54
35	Wye_Target	Male54
37	Newport	_02_P3/49
39	_0180	_02_F9/60
47	Apollo	MV_0437M

48	_03142	MV_0437M
52	Apollo	ZgZg_0438M
56	Wye_Target	ZgZg_0438M
57	Apollo	Male48
62	Apollo	Male51
65	_03120	Male51
67	Apollo	Si_0219M
68	_03142	Si_0219M
70	Apollo	Male54
71	_03142	Male54
74	Wye_Target	Male54
75	Apollo	Male65
78	_03118	RJbeta_0227M
79	_01076	DAM
80	_03124	DAM
82	Cluster	ZsZc_0545M
83	Northern_Bre	ZsZc_0545M
84	Wye_Target	ZsZc_0545M
85	USDA_21055	ZsZc_0545M
86	Toyamidori	ZsZc_0545M
87	_98005	ZsZc_0545M
88	_03140	ZsZc_0545M
89	_02085	ZsZc_0545M
91	_03129	ZsZc_0545M
92	_03153	ZsZc_0545M
93	Bravo	ZsZc_0545M
94	Apollo	ZsZc_0545M
95	Super_Galena	ZsZc_0545M
97	Serebrianka	_035_0648M
98	Wye_Target	_035_0648M
Aurora_x_te	Aurora	USDA_21087M
C1324-001		
C1324-002		
Califorina_Cl	California_Clu	Father_of_Calicross
Cascade	USDA_19124	
Chinook	Petham_Gold	USDA_63012M
DF	Zeus	EA_99M
DG	Zeus	FA_99M
DZp	Zeus	DT_0330M
EA	_98001	USDA_19058M
Fi	_98004	FA_9901M
FK	_98004	FA_0003M
french_land		
FW	_98004	FA_9914M

IE	_98005	FA_9901M
ii	_98005	FA_9901M
KA	Cascade	FA_9901M
MA	Super_Galena	USDA_19058M
Mi	Super_Galena	FA_9901M
MS	Super_Galena	DG_0111M
MZc	Super_Galena	Si_0219M
Ni	_99010	FA_9901M
Northern_Br	Northern_Bre	1196
NS	_99010	DG_0111M
Nugget	USDA_65009	
Pi	_00016	FA_9901M
Ri	KitaMidori	FA_9901M
Si	Toyamidori	FA_9901M
SZL	Toyamidori	DG_0126M
TeaMaker x	TeaMaker	USDA_21422M
Tetraploid_S	New_Zealand	53-5-61
Tetraploid_L	USDA_21049	105/58
TK	Wye_Target	FA_0003M
XaZM	_03116	RJbeta_0227M
XR	Taurus	DG_0110M
YI	Merkur	FA_9901M
YR	Merkur	DG_0110M
Yugoslavian_land		
ZbM	_00022	FC_0005M
ZdR	_00034	DG_0110M
ZdZi	_00034	Ei_0223M
ZeT	_01037	DG_0112M
Zeus_1		
Zeus_2		
ZeX	_01037	FA_9915M
ZgM	_01041	FC_0005M
ZgR	Bravo	DG_0110M
ZgV	_01041	Male47
ZgY	_01041	DG_0116M
ZgZa	_01041	Male50
ZhM	_01042	FC_0005M
ZhR	_01042	DG_0110M
ZjT	_01045	DG_0112M
ZLR	Zenith	DG_0110M
ZnV	Hallertauer_T	Male47
ZnZh	Hallertauer_T	Saazer_M
ZoV	Fuggle	Male47
ZoZh	Fuggle	Saazer_M

ZqZh	Glacier	Saazer_M
ZsZb	_00022	Si_0219M
ZsZc	Apollo	Si_0219M
ZsZL	Apollo	DG_0126M
ZsZn	Apollo	Male20
ZuZM	_02082	RJbeta_0227M
ZvZb	_02096	Male51
ZvZc	_02096	Si_0219M
ZvZn	_02096	Male20
ZZ	A-unk	BA_M