Non-Mendelian inheritance of SNP markers reveals extensive chromosomal translocations in dioecious hops (*Humulus lupulus* L.)

Dong Zhang, Nicholi J. Pitra, Mark C. Coles, Edward S. Buckler, Paul D. Matthews

Hopsteiner, S.S. Steiner, Inc., New York, New York, 10065 (M.C.C, P.D.M, N.J.P, D.Z.), Institute for Genomic Diversity, Cornell University, Ithaca, New York, 14853 (E.S.B, D.Z), Agricultural Research Service, United States Department of Agriculture, Ithaca, New York, 14853, (E.S.B.)

Footnotes:

E.S.B. provided guidance on statistical analyses and project design, N.J.P. prepared sequencing libraries, N.P. and D.Z. analyzed data and interpreted results, M.C.C. collected samples and prepared DNA extracts, P.D.M created germplasm resources, devised and directed the studies, D.Z., N.J.P. wrote and P.D.M edited the manuscript.

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

Corresponding author emails: zhangdong20046@gmail.com, pmatthews@hopsteiner.com

Abstract

Genome-wide meiotic recombination structures, sex chromosomes, and candidate genes for sex determination were discovered among *Humulus* spp. by application of a novel, high-density molecular marker system: ~1.2M single nucleotide polymorphisms (SNPs) were profiled with genotyping-by-sequencing (GBS) among 4512 worldwide accessions, including 4396 cultivars and landraces and 116 wild accessions of hops. Pre-qualified GBS markers were validated by inferences on families, population structures and phylogeny. Candidate genes discovered for several traits, including sex and drought stress-resistance, demonstrate the quality and utility of GBS SNPs for genome-wide association studies (GWAS) and Fst analysis in hops. Most importantly, pseudo-testcross mappings in F1 families delineated non-random linkage of Mendelian and non-Mendelian markers: structures that are indicative of unusual meiotic events which may have driven the evolution and cultivation of hops.

Introduction

The Cannabaceae family of flowering plants has a rich history of contributions to humanity, with the promise of still greater contributions as result new commercial values and invigorated research in two members, *Humulus lupulus* (hop) (2n = 20) and *Cannabis sativa* (hemp, marijuana) (2n = 20) (van Bakel et al., 2011), which diverged around 27.8 Myr (Laursen, 2015). Hop (*H. lupulus*) is a high-climbing, herbaceous perennial, dioecious vine, and has a long history of use as flavoring and stability agent in beer as well as nutraceutical medicine, bio-fuel fermentations and animal fodder (Siragusa et al., 2008). For example, studies of specific hop-derived prenylflavonoids in prevention of cancer, dyslipidemia, and postmenopausal systems spawn interest in metabolic engineering and marker-directed breeding in hop (Ososki and Kennelly, 2003; Stevens and Page, 2004; Nagel et al., 2008; Miranda et al., 2016). The properties of its reproductive system, such as dioecy and obligate outcrossing, high heterozygosity and a large genome size (~2.6Gb), and complex sex-determination system (Neve, 1958), render challenges of genetic dissection of complex traits in hops.

Wild *H. lupulus* is represented by at least five extant species: (1) var. *lupulus* for

63 European wild hops; (2) var. *cordifolius* mainly distributed in Japan, and vars. (3)
64 *neomexicanus* (in the Southwestern U.S.), (4) *pubescens* (in the Eastern/Midwestern U.S.)
65 and (5) *lupuloides* (throughout the northern Great Plains); spreading throughout North
66 America. Asian and North American wild hops resemble each other morphologically,
67 suggesting a genetically closer relationship, while they differ more so from European
68 hops (Murakami et al., 2006). Many contemporary cultivars are hybrids of North
69 American and European genetic materials, in which North American hops have been
70 characterized by their higher bitterness and aroma (Reeves and Richards, 2011) than
71 European cultivars. In other crops, breeding programs have successfully exploited novel
72 genetic variations from wild exotic germplasms into modern cultivars (Tanksley and
73 McCouch, 1997; Bradshaw, 2016) to gain desirable traits such as favored flavors, drought
74 tolerance, and disease resistance. Successes with wild resources and predictions of
75 climate change have spurred resurgence in conservation biology of plant genetic
76 resources (Castañeda-Álvarez et al., 2016; Gruber, 2016).
77
78 Molecular marker systems have been developed for hops and applied in genetic mapping
79 of families (reviewed in Henning et al., 2015), including non-referenced GBS markers
80 (Matthews et al., 2013) and GWAS applied to disease resistance (Henning et al., 2015)
81 and sex determination (Hill et al., 2016), using whole genome-referenced GBS markers.
82 However, genetic phenomenon in hops still is under-explained. For example: (1)
83 significant segregation distortion from Mendelian segregation expectations have been
84 repeatedly reported in mapping populations, indicating that the segregation bias was due
85 to genetic properties rather than genotyping errors (Seefelder et al., 2000; McAdam et al.,
86 2013); (2) Unusual female-biased sex ratios have been observed in controlled crosses,
87 where high pollen loads were applied (Jakse et al., 2008). In other genetic systems,
88 segregation distortion is a result of various patterns of meiotic drive and chromosomal
89 (re)arrangements (Taylor and Ingvarsson, 2003). Examples of meiotic drive include the B
90 chromosomes observed in insects (Fontana and Vickery, 1973), the X-linked meiotic
91 drive in Drosophila species (Lyttle, 1993) and the t-haplotype in mice (Silver, 1993).
92 Some well-known examples of chromosomal (re)arrangements in plants include the
93 neocentromeres (knobs) of maize (Buckler et al., 1999), the translocation heterozygosity

in some species of the genus *Clarkia* (Snow, 1960), *Oenothera* (Rauwolf et al., 2008; Golczyk et al., 2014) and *Viscum* (mistletoe) (Wiens and Barlow, 1975; Rauwolf et al., 2008).

One classical cytogenetic analysis (Sinotô, 1929) suggested that at least 4 chromosomes in the male of var. *cordifolius* are involved in reciprocal translocation, detected by a chromosome chain in a shape of 'zigzag' during the first meiotic metaphase. Our characterization with an unprecedented system of high density genome-wide SNP markers allows detailed examination of translocation heterozygosity, which predisposes rediscovery of the mode of inheritance in this species. Cytogenetic investigations in *Clarkia* (Snow, 1960) and *Oenothera* (Rauwolf et al., 2008; Golczyk et al., 2014) may provide comparative insight on putative translocation heterozygosity in *Humulus*.

However, except for Japanese wild hops (var. *cordifolius*), heterozygotes in males were reported neither in European wild types (var. *lupulus*) nor in North American wild types (var. *neomexicanus*) (Winge, 1932; Jacobsen, 1957; Shephard and Parker, 2000). To date, size ratio of sex chromosomes in wild hops is not fully explained: a description of sex-chromosome variability across subspecies in hops is widely accepted (Shephard and Parker, 2000). Moreover, lack of cytogenetic studies in modern hybrids confuses the mode of genetic inheritance in hops. New insight into genetic parameters in *Humulus* is herein offered by a next generation sequencing (NGS) platform applied to an unprecedented 4512 accessions, including 22 sibling and halfsibling families, genotyped with GBS SNP marker system, comprising 1,235,148 SNPs. Previously reported NGS GBS studies in hop (Matthews, 2013, Henning et al., 2015; Hill et al., 2016) have focused on smaller (511) association panels and reduced marker sets, have not addressed population structure. Furthermore, filtering against SNPs in segregation distortion (SD), thus, ignoring rather than characterizing chromosome regions in SD, resulted in a low-power genetic system that dubiously supported candidate gene discoveries. We present a qualified, high-density marker system applied across a structured association panel and 22 families, analysis of population structure among exotic vs. cultivated accessions within the panel, and a set of strongly supported candidates associated with

125 sex determination. Inclusion, rather than ignorance, of markers in strong SD, has led to
126 testable hypotheses of chromosome structure and recombination constraints in hop, which
127 requires re-assessment of breeding strategies. The need for new, classical cytogenetic
128 studies is implicated by NGS exploration.
129
130 **Results**
131 **Phylogenetic relationships of modern cultivars and North American indigenous**
132 **exotics**
133 European var. *lupulus* is the ancestor of most commercial hops used today, thereby
134 commercial cultivars retain a large proportion of var. *lupulus* genome. In addition, the
135 genetic diversity of hop crop has been contributed by mostly male donors from North
136 America and Asia. To understand the phylogenetic relatedness in hop races, we focused
137 on a subset of 251 accessions, consisting of 183 modern cultivars (CV) including all
138 progenitors of F1 families in this study and 68 wild hops. The neighbor-joining tree
139 (Figure 1a) shows three distinct clusters. The modern cultivars were clustered together,
140 indicating a common derivation in domestication of hops. The other two clusters reflect
141 geographical origins of North American wild hops (Figure 1b), in which one group
142 (SW_wild) includes 22 Southwestern U.S. wild hops (represented by var. *neomexicanus*),
143 and the other group contains 20 wild hops (represented by var. *lupuloides*) from Northern
144 U.S./Canada (N_wild) and 3 (represented by var. *pubescens*) from Midwestern U.S.
145 (MW_wild). Seven wild individuals from Kazakhstan are intermediate among the modern
146 cultivars, consistent with a previous inference (Murakami et al., 2006) of a close genetic
147 relationship between wild hops from Europe and the Altai region (close to western China,
148 located on boundaries of Russia, Mongolia, Kazakhstan and China).
149
150 The level of population differentiation, fixation index (Fst), was measured across the
151 three clusters. SW_wild exhibits relatively close genetic relationship (Fst = 0.1663) with
152 N_wild, apparently supporting relatively close ancestry and geographical origins of the
153 two wild populations. Genetic distinction between the modern cultivars and the North
154 American wild hops is evident: [Fst (CV vs. SW_wild) = 0.31; Fst (CV vs. N_wild) =
155 0.295].

156

157 To demonstrate the population structure of F1 families (N ≥ 60) (Figure 2a) in our dataset,
158 we used a nonlinear algorithm (implemented in Python scikit-learn), t-Distributed
159 Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008), for dimension
160 reduction of the IBS-based distance matrix. The F1 families derived from genetically
161 divergent progenitors can be easily distinguished from one another, while the half-sibling
162 families exhibit ambiguous clustering patterns. A network of pedigree (Figure 2b) reflects
163 that the F1 families in the current dataset were mostly derived from genetically related
164 varieties.

165

166 **Segregation distortion in hybrids**

167 Ubiquitous presence of non-Mendelian factors results from inherent mechanisms in *H.*
168 *lupulus* or from genotyping errors. While genotyping errors are random, the genuinely
169 distorted markers can exhibit pronounced correlation with Mendelian segregation
170 markers. On the basis of clustering of pairwise Spearman's correlation in
171 pseudo-testcross (Pt) markers in three F1 families, we hypothesize that (1) severe SD
172 tends to occur near breakpoints to favor translocation complexes; (2) patterns of linkage
173 can differ across the three populations; (3) a large scale, perhaps genome-wide, meiotic
174 chromosomal complex might occur in the progenitors of the three populations; and (4)
175 translocation heterozygosity may be a ubiquitous phenomenon in hybrids of hops,
176 implicating its genetics and biological significance throughout the cultivation history of
177 hops.

178

179 Pseudo-testcross in families "144" and "247" shows multiple 'super' linkage groups in
180 terms of their size and inter-marker correlation (Figure 3a,S2a). In family "265", linkage
181 groups tend to have equal size (Figure S2b), but exhibit relatively high correlation to one
182 another. Alignments across the three sets of clusters (before phasing coupling groups)
183 (Figure 3b,3c) show most of anchor (common) markers were distinctly clustered. Using a
184 nonlinear algorithm (implemented in Python scikit-learn), locally linear embedding (LLE)
185 (Roweis et al., 2000), the consistent clustering patterns (Figure 4) were implicated by
186 projection of genetic maps into spatial coordinates. Moreover, we observed that the

Louvain method performs extremely well to phase groups in coupling. The effectiveness of the clustering methods indicates the correlation across linkage groups was caused by real meiotic events.

The loci with 5% ≤ MAF < 15%, deviated significantly from the 1:3 expectation for Pt markers, account for 28.3%, 49% and 48.3% in families "144", "247" and "265" respectively, in which proportions of the distorted loci correlated (rho ≥ 0.3) to the Mendelian segregation markers (15% ≤ MAF ≤ 35%) are 78.3%, 48.9% and 71.8%.

Spatial coordinates of Pt markers, in accordance with correlation heatmaps, offer a complete picture (Figure 4) of genetic linkage with and without inclusion of SD markers, which appear to play a role in bridging Mendelian markers. Such linkage patterns were constantly observed in the three families (Figure 4, S3).

We show the correlation across the 5 largest linkage groups (Figure 5a) in family "265" based on the spatial representation of the positions of the markers with 15% ≤ MAF ≤ 35%. The model (Figure 5b) clearly depicts distortion decreases, when the markers are more distal to the convergence areas, and indicates that the progenitors of family "265" experienced translocation to form a large chromosomal complex.

One linkage group (LG) in one family corresponding to multiple groups in the other family, would suggest corresponding loci were involved in chromosomal rearrangement in progenitor of the former family. One striking case (Figure 6) can be found in LG2.1 in "144" corresponding to two coupling LGs (2.1 and 2.2) in "265". Two control correspondences (LG1.1-LG1.2 and LG3.1-LG3.1) were used to illustrate the power of the clustering approaches. However, such one-to-multiple correspondence was seldom observed across the three families. That may reflect the conservation of chromosomes positioning in the heterozygotes complex and invariable occurrence of the translocation heterozygotes in the progenitors of the three families.

GWAS for sex determination

218 Sexual phenotype regulation is a particularly important problem in dioecious plants,
219 herein exemplified by hop (*H. lupulus*). Male and female hop flowers can be easily
220 distinguished; the male flower closely resembles a typical perfect flower, while female
221 inflorescence meristems produce flowers arranged in 'cones' (Shephard and Parker,
222 2000).
223
224 We used a mixed linear model to assess evidence of phenotype-genotype association. In
225 families "247" (N = 364, N_{male} = 30) and "265" (N = 95, N_{male} = 13), linkage group (LG)
226 4 (Figure 7a,S4) consistently shows the most striking association with sex, even though
227 "265" has a small effective population size. This signal was additionally supported by Fst
228 mapping in "247" (Figure 7b). However, pseudo-testcross accounts for part of association
229 signals. To perform genome-wide scan, we assessed association between 356,527
230 markers and 850 individuals (N_{male} = 129, N_{female} = 721), as described in Methods.
231
232 A total of 588 SNPs with $P \leq 10^{-7}$ were identified (Figure 7c,7d). We noted that LG4 and
233 other LGs respectively account for 38.6% and 0.0% of the association markers,
234 reinforcing the importance of LG4 in sex expression in hops. On the basis of pairwise
235 Spearman's rho, we observed high correlation among the 588 SNPs, implicating that the
236 association markers mostly come from one linkage disequilibrium (LD) block. Adding up
237 scaffolds showing association approximates ~9.75Mb of the mapping resolution
238 accounting for ~0.38% of the hop genome.
239
240 To identify gene candidates for hop's sex, we aligned by BLAST the top 100 scaffolds
241 against UniProt and NCBI protein databases to search for genes with known function
242 involving in the mechanism of sex determination (Table S4). A total of 8 gene candidates
243 (Ruegger et al., 1998; Ishiguro et al., 2002; Koizuka et al., 2003; Hála et al., 2008;
244 Shimizu et al., 2008; Zhang et al., 2008; Chen et al., 2010; Zhang et al., 2011) were
245 obtained (Figure 7d), in which 7 blast hits have 81%-99% length matching to the target
246 proteins and 7 hits encompass ≥ 1 the significant association site ($p \leq 10^{-25}$).
247
248 In hop, a significant deformation of the apical meristem, the producer of flower primordia

249 cells, has been observed during the transition to the reproductive phase, resulting in the
250 apparent morphological differences between male and female flowers (Shephard and
251 Parker, 2000). Two notable examples, relating to the floral structures, are (1) a
252 glucose-regulated protein 94 (GRP94)-like protein on scaffold LD152823 that is known
253 in Arabidopsis affecting shoot apical meristems, floral meristems and pollen tube
254 elongation (Ishiguro et al., 2002); and (2) a Squamosa-like protein, identified on scaffold
255 LD147778, has essential roles in vegetative phase change and flower development in
256 multiple plants (Chen et al., 2010).
257
258 **Genetic differences and phenotypic variation across populations**
259 To assess genetic contributions to between-population phenotypic differences, we used
260 Fst analysis to characterize genetic variations across var. *neomexicanus*, var. *lupuloides*
261 and CV. Previously cloned genes in *Humulus* were highlighted to suggest that with
262 respect to essential chemical composition and drought tolerance, hotspots with unusually
263 high or low Fst values deserve a great deal of attention.
264
265 A linkage map-based view of Fst highlights two notable patterns (Figure 8). First, the
266 degree of genetic variation, as expected, is much greater in CV vs.
267 *neomexicanus*/*lupuloides* than in *neomexicanus* vs. *lupuloides*. Regions that exhibit
268 above-average population differentiation in CV vs. *neomexicanus* typically also exhibit
269 above-average population differentiation in CV vs. *lupuloides*. Second, the 5 largest
270 linkage groups account for a large proportion of genetic variation between populations.
271 The 90th percentile of Fst values (Figure 8d), referred to as 0.63 in CV vs. *neomexicanus*
272 and 0.62 in CV vs. *lupuloides*, were used to define the significant genetic difference.
273 In beer brewing, the chemical composition, resins imparting bitterness and essential oils
274 contributing to flavor and aroma, determine the quality and flavor of hops. The key
275 brewing resins include alpha-acids and beta-acids, also referred to as humulones and
276 lupulones respectively. It is generally accepted that North American hops have high
277 content of alpha acids, rendering bitter beers, while European hops have lower content of
278 resins and essential oil. Enhancement of alpha acids content in the contemporary cultivars
279 was accomplished by crossing European hops (var. *lupulus*) to North American donors,

280 one of which resembles var. *lupuloides* from Canada (Neve, 1991; Reeves and Richards,
281 2011). Genes affecting resins and essential oil can be found in the regions presenting
282 genetic divergence between North American hops and adapted varieties.
283
284 In hop, two homologues of chalcone synthase (*CHS*) in bitter acid biosynthesis have been
285 cloned (Figure S5), referred to as *CHS_VPS* (GenBank: AB015430.1) (Okada and Ito,
286 2001) and *CHS_H1* (GenBank: CAC19808.1) (Matoušek et al., 2002) respectively. Both
287 genes are highly active in lupulin glands, to serve as catalysts of the synthesis reactions of
288 alpha-acid and beta-acid. Significant genetic differences (Fst $\geq$ 0.6) between cultivars
289 and North American wild types emerge in a region adjacent to (~40Kb away from)
290 *CHS_H1*. In contrast, a surrounding region (~15Kb away from) of *CHS_VPS* harbors
291 moderate Fst (= ~0.4) in CV vs. *neomexicanus*, and low Fst (= ~0.1) in CV vs. *lupuloides*.
292 The differences of population differentiation in the surrounding regions of the two *CHS*
293 homologues may reflect genetic introgression and differential allele selection resulting
294 from domestication towards higher alpha acid yields.
295 Var. *neomexicanus* is traditionally grown in rain-fed or supplementary irrigated areas,
296 and has capability to withstand periods of water deficit and to yield an economic return to
297 farmers. In contrast, hops frequently occurring in the high latitude of the northern North
298 America favor plenty of water and sunlight. Likewise, heavy irrigation is required in
299 modern hop crops. Hence, genic regions exhibiting the significant population
300 differentiation in CV vs. *neomexicanus*, but not in CV vs. *lupuloides* can be prioritized to
301 identify gene candidates affecting drought tolerance.
302
303 Two intriguing proteins were identified on scaffold "LD161390" (Figure 9a), which are a
304 *H. lupulus* basic leucine zipper transcription factor Long Hypocotyl 5 (*bZIP HY5*)
305 (GenBank: CBY88800.1) (Matousek et al., 2010) and a light-harvesting chlorophyll
306 (*LHCII*) a/b binding protein (*LHCB*) (NCBI RefSeq: XP_002307004.1).
307
308 *HY5* has a notable relationship with the phytohormone abscisic acid (ABA), a plant
309 hormone in response to environmental stress, such as high salinity, drought and low
310 temperature (Nakagawa H, Ohmiya K, 1996). Via binding and promoting *ABI5*, encoding

311 a bZIP transcription factor, *HY5* can mediate *ABA* sensitivity (Chen et al., 2008) to induce
312 stomatal closure, thus reducing transpiration and preventing water loss from leaves.
313
314 The putative *LHCB*, physically close to the *HY5*-like gene, embeds a mutation with
315 striking Fst in CV vs. *neomexicanus*, but not in CV vs. *lupuloides*. *LHCB* involves in the
316 steps of photosynthesis by capturing sunlight and balancing excitation energy between
317 photosystems I (PSI) and II (PSII). It is clear that over-expression of a *LHCB* is a key to
318 enhance stomatal sensitivity to ABA in green alga (*Dunaliella salina*) (Liu and Shen,
319 2004).
320
321 A previously *C. sativa* (marijuana) *LHCII* protein, *CP47* (NCBI RefSeq:
322 YP_009143579.1), is present on scaffold "LD140230" (Figure 9b), near markers showing
323 the pronounced population divergence between *neomexicanus* and CV/*lupuloides*. In
324 sorghum, *CP47* serves as a connection between the main light harvesting complex LHCII
325 and reaction center of PSII. The characteristic adaptation of sorghum to drought may be
326 partly related to downregulation of *CP47* during drought stress and high irradiance
327 (Masojídek et al., 1991).
328
329 **Discussion**
330 Hop crop acreage and usage is rapidly expanding and diversifying because of a
331 bourgeoning craft brewing industry. Hop breeding programs have been long attempting to
332 exploit genetic resources for bitter flavor, aroma and disease resistance. However, a
333 worsening drought and unseasonably hot weather may decentralize the targets of
334 breeding programs soon. For example, in Europe and the US, most hop farms
335 experienced severe water shortage in 2015. Like many other crops, exploitation of novel
336 genetic variation in response to drought stress is of paramount importance for a
337 sustainable hop production system.
338
339 Understanding genetic recombination is essential for speed and accuracy of plant
340 breeding. Indeed, it is generally difficult to breed new commercial hop varieties through
341 mass selection and crossing. Our findings show that a large scale, perhaps genome-wide,

342 chromosomal rearrangement may occur in the progenitors of F1 families. Translocation
343 heterozygosity can extend linkage to nonhomologous chromosomes, and favor severe
344 segregation distortion accumulated near the translocation breakpoints (Taylor and
345 Ingvarsson, 2003; Farré et al., 2011). Such high degree recombination suppression may
346 hinder effectively selection of desired, novel allele combinations. Inasmuch, breeding
347 strategies favoring introgression of diversity deserve re-emphasis in addition to extended
348 attention to recombination and segregation phenomena in traits. Manipulation of genetic
349 recombination in hops also deserves further focus.
350
351 At least 57 species of flowering plants were characterized by permanent translocation
352 heterozygotes (Holsinger and Ellstrand, 1984), in which *Clarkia* (2n = 18) may provide a
353 comparative system to hypothesize meiotic configurations in *Humulus*. Translocation
354 polymorphism has been observed in at least 14 of the 34 known species in *Clarkia*
355 (Snow, 1960). Judging from cytogenetic analyses in 9 natural populations of *Clarkia*
356 *dudleyana*, none or few translocation heterozygotes occurred within individual colonies,
357 while extensive translocation heterozygotes invariably arose from hybrids derived from
358 cytologically distinct races (Snow, 1960). The largest heterozygotes complex consists of
359 18 chromosomes. For *H. lupulus* wild types, until now, researchers have still been at odds
360 about the size ratios of sex chromosomes. Except for a tetravalent in var. *cordifolius*, no
361 meiotic chromosome associations were reported in other wild types, suggesting that the
362 five wild *H. lupulus* species are cytologically differentiated. Segregation distortion
363 presenting in cultivated mapping populations may stem from historical hybridization
364 across isolated populations of hops.
365
366 Based on the projection of SNPs into a 3D coordinate system using LLE, the
367 nondisjunction of the 5 largest LGs in "265" provides an example of funnel-shaped
368 representation of a heterozygotes complex. Specifically, chromosome segments distal to
369 and close to translocation breakpoints can be represented by wider and narrower end of
370 the funnel-shaped structure respectively. Apparently, two ends of the funnel also
371 characterize clusters of SNPs with similar allele frequencies. There is a need for
372 cytogenetic studies for drawing conclusions regarding the exact meiotic configuration in

373 *Humulus* hybrids for both males and females. Whether the translocation heterozygosity is
374 sex-linked deserves further investigation.
375
376 Translocation heterozygosity may have an important connection to the significantly
377 distorted sex-ratio in favor of females in hops. Likewise, female-biased sex ratios have
378 been found in Mistletoe, another notable dioecious case of translocation heterozygosity.
379 Known that, to maintain heterozygosity, *Oenothera*, a notable monoecious case of
380 translocation heterozygosity, utilizes a system of balanced lethal to purge the lethal
381 homozygotes (Steiner, 1956; Harte, 1994), which is referred to as "recessive lethals". In
382 the context of XY system, heteromorphism of sex chromosomes dictates that males are
383 more severely affected than females by "X-linked recessive lethals", because males only
384 have one copy of the X chromosome. Hence, *H. lupulus* may use a system of balanced
385 lethals at the expense of male offspring to preserve genetic heterozygosity in hybrids.
386
387 Our results are compelling for translocation heterozygosity studies in light of
388 high-density molecular markers in many other biota. For example, such large scale
389 recombination suppression is also presented in at least 10 species of termite, some types
390 of centipede, and perhaps all of the monotremes (Holsinger and Ellstrand, 1984; Rowell,
391 1987; Rens et al., 2004). Beyond homologous crossover, translocation heterozygosity has
392 shown considerable evolutionary interest and selective advantage in its own right.
393
394 **Materials and Methods**
395 **Plant materials**
396 Hops used in this study were grown under standard agronomic conditions at the Golden
397 Gate Ranches, S.S. Steiner, Inc, Yakima, WA. The un-domesticated, exotic hops are from
398 the National Clonal Germplasm Repository in Corvallis, Oregon (accession details in
399 Table S1-S3). Fifty milligrams of young leaf tissues were extracted in a 96 well block
400 using Qiagen Plant DNeasy Kits and was tested for quality, quantity, and purity, prior to
401 library preparations, using a Agilent 2100 Bioanalyzer (Applied Biosystems, Foster City,
402 CA) and Life Technologies (Carlsbad, CA) Qubit 3.0 Fluorometer. The GBS libraries
403 were prepared using the *Ape*K1 enzyme according to Elshire, et al. (Elshire et al., 2011).

404 Pools of 96 accessions were sequenced on one lane of an Illumina HighSeq 2000
405 (Illumina, San Diego, CA)
406
407 **SNP calling and quality control**
408 The reference sequence refers to a draft haploid genome sequence of Shinshu Wase (SW)
409 (Natsume et al., 2015), which is a modern cultivar bred from a seeding selection cross
410 between Saazer and White Vine-OP**.** The draft genome, with a total size of 2.05 Gb,
411 consists of ~130,000 scaffolds covering approximately 80% of the estimated genome size
412 of hop (2.57 Gb).
413
414 Tassel 5 GBS v2 Pipeline (Glaubitz et al., 2014) was applied to identify tags with at least
415 10x total coverage, and to call SNPs. Tag sequences were mapped to the reference
416 genome using BWA aligner.
417
418 One main source of erroneous SNP calling is misalignment caused by incomplete
419 reference genome, gene duplication and low-complexity regions. To filter out erroneous
420 SNPs due to misalignment, we used two criteria: (1) SNPs with an excessive coverage
421 (e.g. read depth > 127) can be false positives. For GBS data, the maximum read depth for
422 one genotype is unlikely to exceed 127 (Glaubitz et al., 2014). Indeed, we observed that
423 heterozygosity rates and MAF are significantly increased when read coverage exceeds
424 127 (Figure S1). (2) The orientation of paired reads of the cultivar Apollo (unpublished
425 data), a highly used maternal line in our F1 families, was used to detect false positive
426 SNPs caused by gene duplications. Paired-end alignment was generated by BWA Sampe.
427 Identification of correctly aligned regions was based on SAM flags indicating reads
428 mapped in proper pairs. Using criteria (2) was able to detect ~73% SNPs with the
429 excessive coverage.
430
431 **Pseudo-testcross**
432 Three F1 families were used to conduct pseudo-testcross (Pt) recombination mappings,
433 including (1) "144" (N = 179) derived from a cross between Nugget (maternal line) and
434 Male50 (paternal line); (2) "247" (N = 364) derived from two parental lines, Super

435 Galena and Male15; (3) “265” (N = 95) derived from a cross between Chinook and
436 Male57. Using markers heterozygous in the maternal line and null in the paternal line,
437 three genetic map sets were constructed, consisting of 3551 SNPs for “144”, 2369 SNPs
438 for “247” and 4506 SNPs for “265”.
439
440 Our analyses followed the main steps in HetMappS pipelines (Hyma et al., 2015).
441 Specifically, (1) to remove contaminants, identity by state (IBS) based distance matrices
442 calculated by TASSEL (Bradbury et al., 2007) were used to identify outliers for each
443 family; (2) SNPs having both parental genotypes (e.g. AA×Aa) with read depth ≥ 4 were
444 retained for the next step; (3) in progeny, SNPs with average read depth ≥ 4 and with site
445 coverage ≥ 50% were retained for the next step; (4) to eliminate the effect of
446 under-calling heterozygotes and sequencing errors, we masked progeny genotypes with
447 depth=1, and converted genotypes aa to Aa because genotype aa cannot exist for parental
448 genotypes AA×Aa in Pt; (5) after correction, SNPs with 15% ≤ MAF ≤ 35% were
449 selected to create linkage groups, and SNPs with 5%≤ MAF <15% were deemed the
450 pronounced SD markers; (6) to cluster and order markers, an adjacency matrix with
451 Spearman’s correlation (rho) were derived from the remaining SNPs; (7) on the basis of
452 absolute values of rho, the Louvain method (Blondel et al., 2008) implemented in
453 NetworkX (http://networkx.github.io/) was applied to detect communities (clusters). The
454 Louvain method is an efficient algorithm for community detection in large networks. A
455 similar method, modulated modularity clustering (MMC) (Stone and Ayroles, 2009), has
456 been successfully applied to construct linkage groups; (8) to identify coupling phase from
457 each “absolute rho” cluster, negative values of rho were set to zero, and the Louvain
458 method was applied to positive values of rho (Hyma et al., 2015); (9) MSTmap (Wu et
459 al., 2008) was used to provide a sub-optimal solution of genetic ordering within each
460 linkage group.
461
462 Putative 10×2 linkage groups in coupling were obtained in each F1 family. Linkage
463 groups may not represent one chromosome due to pseudo-linkage resulting from
464 chromosomal rearrangement, as discussed in Results.
465

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

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

474

475 **Additional files**

476 **Supplementary Figures.** The file contains supplementary figure S1-S5.

477 **Supplementary Tables.** The file contains supplementary tables. (**Table S1** Pedigrees of
478 genotyped F1 populations. **Table S2** Cultivar and landrace accessions. **Table S3** Wild
479 exotic accessions. **Table S4** BLASTX hits for scaffolds encompassing sex association (P
480 ≤ 10^{-10}) SNPs. Eight strongly supported gene candidates are highlighted.).

481

482 **Acknowledgments**

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

485

486 **Literature Cited**

487 Bakel H V, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The
488 draft genome and transcriptome of Cannabis sativa The draft genome and
489 transcriptome of Cannabis sativa. Genome Biol 12: R102

490 Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of
491 communities in large networks. J Stat Mech Theory Exp 10008: 6

492 Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007)
493 TASSEL: software for association mapping of complex traits in diverse samples.
494 Bioinformatics 23: 2633–5

495 Bradshaw JE (2016) Use of Sexual Reproduction in Base Broadening and Introgression.
496 Plant Breed. Past, Present Futur. Springer. pp 483–527

497 Buckler ES, Phelps-Durr TL, Buckler CS, Dawe RK, Doebley JF, Holtsford TP (1999)

498 Meiotic drive of chromosomal knobs reshaped the maize genome. Genetics 153:
499 415–26

500 Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood
501 RJ, Guarino L, Harker RH, Jarvis A, Maxted N, et al (2016) Global conservation
502 priorities for crop wild relatives. Nat Plants 2: 1–6

503 Chen H, Zhang J, Neff MM, Hong S-W, Zhang H, Deng X-W, Xiong L (2008)
504 Integration of light and abscisic acid signaling during seed germination and early
505 seedling development. Proc Natl Acad Sci U S A 105: 4495–4500

506 Chen X, Zhang Z, Liu D, Zhang K, Li A, Mao L (2010) SQUAMOSA promoter-binding
507 protein-like transcription factors: Star players for plant growth and development. J
508 Integr Plant Biol 52: 946–951

509 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011)
510 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity
511 species. PLoS One 6: 1–10

512 Farré A, Benito IL, Cistué L, de Jong JH, Romagosa I, Jansen J (2011) Linkage map
513 construction involving a reciprocal translocation. Theor Appl Genet 122: 1029–1037

514 Fontana PG, Vickery VR (1973) Segregation-distortion in the B-chromosome system of
515 Tettigidea lateralis (say) (Orthoptera: Tetrigidae). Chromosoma 43: 75–98

516 Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014)
517 TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS
518 One. doi: 10.1371/journal.pone.0090346

519 Golczyk H, Massouh A, Greiner S (2014) Translocations of chromosome end-segments
520 and facultative heterochromatin promote meiotic ring formation in evening
521 primroses. Plant Cell 26: 1280–93

522 Gruber K (2016) Re-igniting the green revolution with wild crops. Nat Plants 2: 1–4

523 Hála M, Cole R, Synek L, Drdová E, Pecenková T, Nordheim A, Lamkemeyer T,
524 Madlung J, Hochholdinger F, Fowler JE, et al (2008) An exocyst complex functions
525 in plant cell growth in Arabidopsis and tobacco. Plant Cell 20: 1330–1345

526 Harte C (1994) Oenothera Contributions of a Plant to Biology. Monogr Theor Appl Genet.
527 doi: 10.1017/CBO9781107415324.004

528 Henning JA, Gent DH, Twomey MC, Townsend MS, Pitra NJ, Matthews PD (2015)
529 Precision QTL mapping of downy mildew resistance in hop (Humulus lupulus L .).
530 doi: 10.1007/s10681-015-1356-9

531 Hill ST, Coggins J, Liston A, Hendrix D, Henning JA (2016) Genomics of the hop
532 pseudo-autosomal regions. Euphytica. doi: 10.1007/s10681-016-1655-9

533 Holsinger KE, Ellstrand NC (1984) The Evolution and Ecology of Permanent
534 Translocation Heterozygotes. Am Nat 124: 48–71

535 Hyma KE, Barba P, Wang M, Londo JP, Acharya CB, Mitchell SE, Sun Q, Reisch B,
536 Cadle-Davidson L (2015) Heterozygous Mapping Strategy (HetMappS) for High
537 Resolution Genotyping-By-Sequencing Markers: A Case Study in Grapevine. PLoS
538 One. doi: 10.1371/journal.pone.0134880

539 Ishiguro S, Watanabe Y, Ito N, Nonaka H, Takeda N, Sakai T, Kanaya H, Okada K (2002)
540 SHEPHERD is the Arabidopsis GRP94 responsible for the formation of functional
541 CLAVATA proteins. EMBO J 21: 898–908

542 Jacobsen P (1957) The sex chromosomes in Humulus. Hereditas 43: 357–370

543 Jakse J, Stajner N, Kozjak P, Cerenak A, Javornik B (2008) Trinucleotide microsatellite
544 repeat is tightly linked to male sex in hop (Humulus lupulus L.). Mol Breed 21:
545 139–148

546 Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T,
547 Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide
548 repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile
549 Kosena radish. Plant J 34: 407–415

550 Laursen L (2015) Botany: The cultivation of weed. Nature 525: S4–S5

551 Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang
552 Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics
553 28: 2397–9

554 Liu X-D, Shen Y-G (2004) NaCl-induced phosphorylation of light harvesting chlorophyll
555 a/b proteins in thylakoid membranes from the halotolerant green alga, Dunaliella
556 salina. FEBS Lett 569: 337–40

557 Lyttle TW (1993) Cheaters sometimes prosper: distortion of mendelian segregation by
558 meiotic drive. Trends Genet 9: 205–210

559 Maaten L Van Der, Hinton G (2008) Visualizing Data using t-SNE. J Mach Learn Res 9:
560 2579–2605

561 Masojídek J, Trivedi S, Halshaw L, Alexiou A, Hall DO (1991) The synergistic effect of
562 drought and light stresses in sorghum and pearl millet. Plant Physiol 96: 198–207

563 Matousek J, Kocábek T, Patzak J, Stehlík J, Füssy Z, Krofta K, Heyerick A, Roldán-Ruiz

564 I, Maloukh L, De Keukeleire D (2010) Cloning and molecular analysis of HlbZip1
565 and HlbZip2 transcription factors putatively involved in the regulation of the lupulin
566 metabolome in hop (Humulus lupulus L.). J Agric Food Chem 58: 902–12

567 Matoušek J, Novák P, Bříza J, Patzak J, Niedermeierová H (2002) Cloning and
568 characterisation of chs-specific DNA and cDNA sequences from hop (*Humulus*
569 *lupulus* L.). Plant Sci 162: 1007–1018

570 Matthews PD, Coles MC, Pitra NJ (2013) Next Generation Sequencing for a Plant of
571 Great Tradition : Application of NGS to SNP Detection and Validation in Hops
572 (*Humulus lupulus* L .). BrewingScience 66: 185–191

573 McAdam EL, Freeman JS, Whittock SP, Buck EJ, Jakse J, Cerenak A, Javornik B, Kilian
574 A, Wang C-H, Andersen D, et al (2013) Quantitative trait loci in hop (Humulus
575 lupulus L.) reveal complex genetic architecture underlying variation in sex, yield
576 and cone chemistry. BMC Genomics 14: 360

577 Miranda CL, Elias VD, Hay JJ, Choi J, Reed RL, Stevens JF. Xanthohumol improves
578 dysfunctional glucose and lipid metabolism in diet-induced obese C57BL/6J mice.
579 Archives of biochemistry and biophysics. 2016 Jun 1;599:22-30.

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

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

585 Nakagawa H, Ohmiya K HT (1996) A rice bZIP protein, designated OSBZ8, is rapidly
586 induced by abscisic acid. Plant J 217–227

587 Natsume S, Takagi H, Shiraishi a., Murata J, Toyonaga H, Patzak J, Takagi M,
588 Yaegashi H, Uemura a., Mitsuoka C, et al (2014) The Draft Genome of Hop
589 (Humulus lupulus), an Essence for Brewing. Plant Cell Physiol 0: 1–14

590 Neve RA (1958) Sex Chromosomes in the Hop Humulus lupulus. Nature 181: 1084 –
591 1085

592 Neve RA (1991) Hops. London Chapman Hall

593 Okada Y, Ito K (2001) Cloning and analysis of valerophenone synthase gene expressed
594 specifically in lupulin gland of hop (Humulus lupulus L.). Biosci Biotechnol
595 Biochem 65: 150–155

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

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

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

603 Rens W, Grützner F, O'brien PCM, Fairclough H, Graves JAM, Ferguson-Smith MA
604 (2004) Resolution and evolution of the duck-billed platypus karyotype with an
605 X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. Proc Natl Acad
606 Sci U S A 101: 16257–16261

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

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

611 Ruegger M, Dewey E, Gray WM, Hobbie L, Turner J, Estelle M (1998) The TIR1 protein
612 of Arabidopsis functions in auxin response and is related to human SKP2 and yeast
613 Grr1p. Genes Dev 12: 198–207

614 Seefelder S, Ehrmaier H, Schweizer G, Seigner E (2000) Male and female genetic linkage
615 map of hops Humulus lupulus. Plant Breed 119: 249–255

616 Shephard H, Parker J (2000) Sexual development and sex chromosomes in hop. New
617 Phytol 148: 397–411

618 Shimizu KK, Ito T, Ishiguro S, Okada K (2008) MAA3 (MAGATAMA3) helicase gene
619 is required for female gametophyte development and pollen tube guidance in
620 Arabidopsis thaliana. Plant Cell Physiol 49: 1478–1483

621 Silver LM (1993) The peculiar journey of a selfish chromosome: mouse t haplotypes and
622 meiotic drive. Trends Genet 9: 250–254

623 Sinotô Y (1929) Chromosome Studies in Some Dioecious Plants, with Special Reference
624 to the Allosomes. Cytologia (Tokyo) 1: 109–191

625 Siragusa GR, Haas GJ, Matthews PD, Smith RJ, Buhr RJ, Dale NM, Wise MG. (2008)
626 Antimicrobial activity of lupulone against Clostridium perfringens in the chicken
627 intestinal tract jejunum and caecum. Journal of antimicrobial chemotherapy. Apr
628 1;61(4):853-8.

629 Snow R (1960) Chromosomal Differentiation in Clarkia dudleyana. Am J Bot 47:
630 302–309

631 Steiner E (1956) New aspects of the balanced lethal mechanism in oenothera. Genetics

632 Stevens JF, Page JE (2004) Xanthohumol and related prenylflavonoids from hops and
633 beer: To your good health! Phytochemistry 65: 1317–1330

634 Stone E a., Ayroles JF (2009) Modulated modularity clustering as an exploratory tool for
635 functional genomic inference. PLoS Genet. doi: 10.1371/journal.pgen.1000479

636 Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic
637 potential from the wild. Science 277: 1063–1066

638 Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants
639 and animals. Genetica 117: 27–35

640 Wiens D, Barlow BA (1975) Permanent Translocation Heterozygosity and Sex
641 Determination in East African Mistletoes. Science (80-) 187: 1208–1209

642 Winge O (1932) The Nature of Sex Chromosomes. Proc Sixth Int Congr Genet 343–355

643 Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of
644 genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. doi:
645 10.1371/journal.pgen.1000212

646 Zhang Y, Feng S, Chen F, Chen H, Wang J, McCall C, Xiong Y, Deng XW (2008)
647 Arabidopsis DDB1-CUL4 ASSOCIATED FACTOR1 forms a nuclear E3 ubiquitin
648 ligase with DDB1 and CUL4 that is involved in multiple plant developmental
649 processes. Plant Cell 20: 1437–1455

650 Zhang Z, Zhang Y, Tan H, Wang Y, Li G, Liang W, Yuan Z, Hu J, Ren H, Zhang D
651 (2011) RICE MORPHOLOGY DETERMINANT encodes the type II formin FH5 and
652 regulates rice morphogenesis. Plant Cell 23: 681–700

653

654

655

656

657

658

659

660

661

662

663 **Figure Legends**

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

669 **Figure 2** Population structure and pedigree network of F1 families. (a) t-SNE plot for 20
670 F1 families (N $\geq$ 60). (b) The overview of pedigree for genotyped F1 families.

671 **Figure 3** Linkage groups for the maternal line of family "144" and correspondence across
672 3 genetic map sets. The degrees of Spearman's correlation (rho) are color-coded. (a)
673 Unphased and phased (linkage for grandparents) groups are bounded by white and black
674 frames individually. (b) Alignment of unphased groups between "144" and "247". (c)
675 Alignment of unphased groups between "144" and "265".

676 **Figure 4** Linkage of Mendelian (15% $\leq$ MAF $\leq$ 30%) and non-Mendelian Pt markers
677 (5% $\leq$ MAF $<$ 15%), based on Spearman's correlation (rho). In each sub-figure,
678 clustering patterns without (left) and with (right) inclusion of segregation distortion are
679 presented by LLE (top) and the Louvain Modularity (bottom). Mendelian markers in two
680 linkage groups are indicated by blue and red colors individually. Segregation distortion
681 (SD) markers are indicated by yellow color.

682 **Figure 5** Linkage patterns of the 5 largest linkage groups in family "265", based on
683 spatial coordinates defined by LLE. (a) Linkage groups are color-coded. (b) Markers with
684 0.15 $\leq$ MAF $<$ 0.2 and 0.2 $\leq$ MAF $\leq$ 0.3 are distinguished by cyan and gray colors.

685 **Figure 6** One-to-two genetic correspondence between "144" and "265". LG2.1 in "144"
686 corresponds to LG2.1 and LG2.2 in "265" (top sub-figure). Two instances of one-to-one
687 correspondence (LG1.1-LG1.2 and LG3.1-LG3.1) are added for control. Spatial
688 representations of linkage groups (bottom sub-figure) in the two families were derived
689 from LLE.

690 **Figure 7** Association studies and Fst mapping of sex determination in hop. (a) Linkage
691 group-based Manhattan-plot of MLM for sex determination in family "247" (N = 364,

692 N_{male} = 30). Light and deep colors are used to distinguish two phases (linkage for
693 grandparents) in coupling. (b) Manhattan-plot of Fst in females vs. males in "247". (c)
694 Log Quantile-Quantile (QQ) plot of 356,526 association tests (SNPs) for sex
695 determination in 850 individuals (N_{male} = 129, N_{female} = 721). (d) Correlation among 588
696 association ($P \leq 10^{-7}$) markers, the proportions of 588 markers in LG4, other LGs and
697 unmapped data set, and 8 gene candidates for sex determination in hop.

698 **Figure 8** Linkage group (in family "247")-based Fst heatmaps and the overall Fst
699 distribution. Population differentiation (a) between modern cultivars (CV) and var.
700 *neomexicanus*; (b) between CV and var. *lupuloides*; (c) between var. *neomexicanus* and
701 var. *lupuloides*. (d) Spectrum of the overall Fst distribution.

702 **Figure 9** Three gene candidates for drought tolerance in hops. Between-population Fst
703 values are indicated on the scaffolds where the three candidates are located. (a) The
704 approximate positions of *bZIP HY5* (GenBank: CBY88800.1) (Matousek et al., 2010) and
705 *LHCB* (NCBI RefSeq: XP_002307004.1) on scaffold "LD161390". (b) The approximate
706 position of *CP47* (NCBI RefSeq: YP_009143579.1) on scaffold "LD140230"

707

708

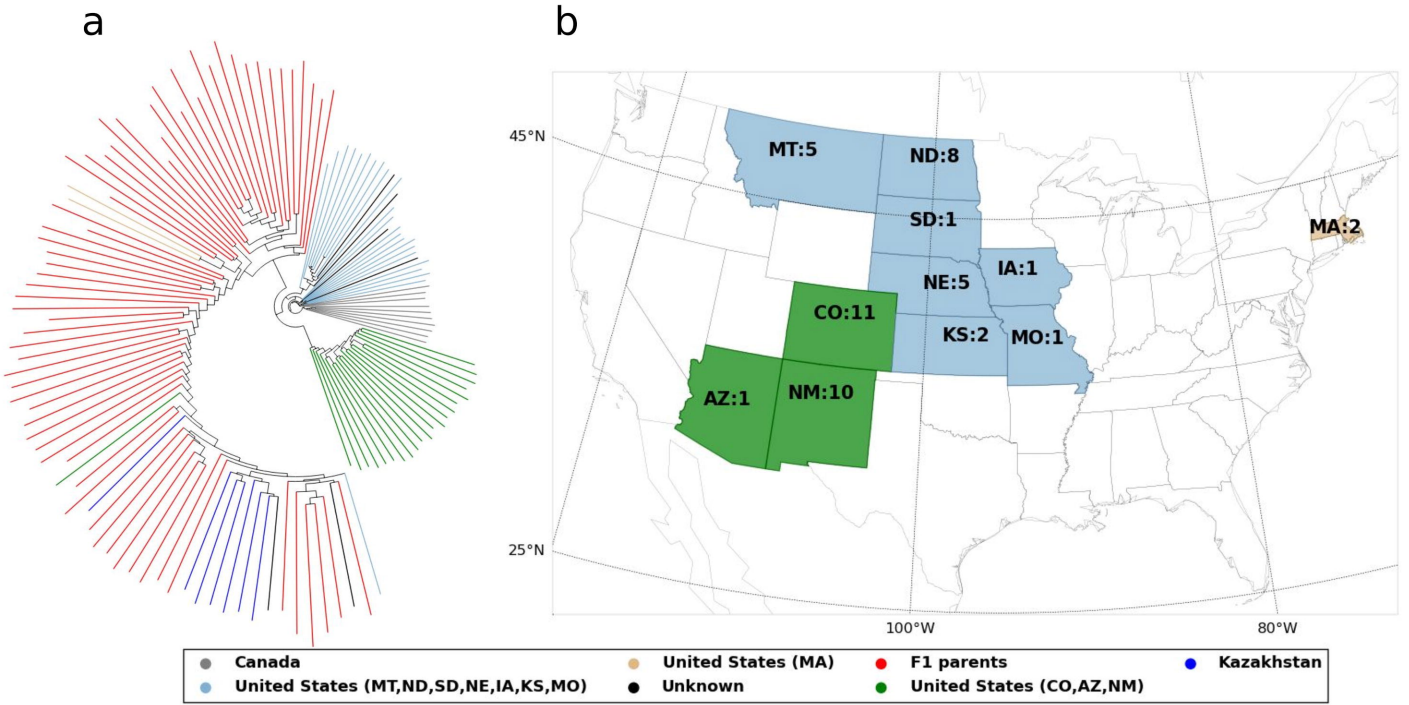
709

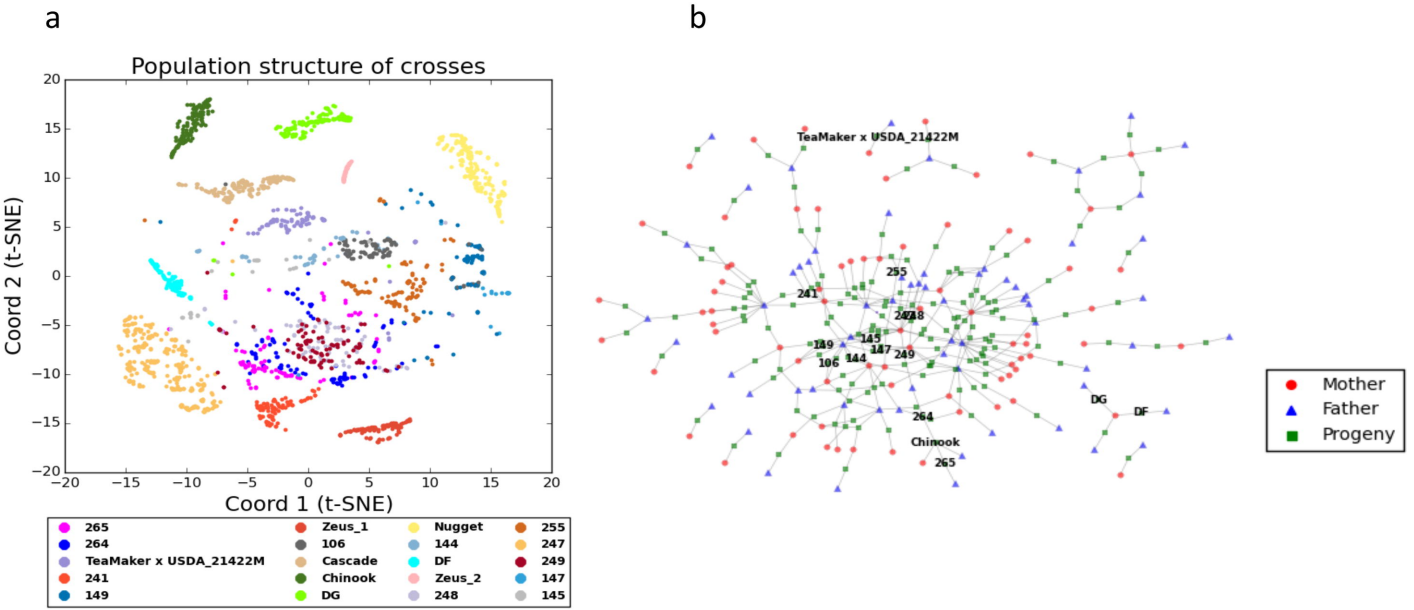
710

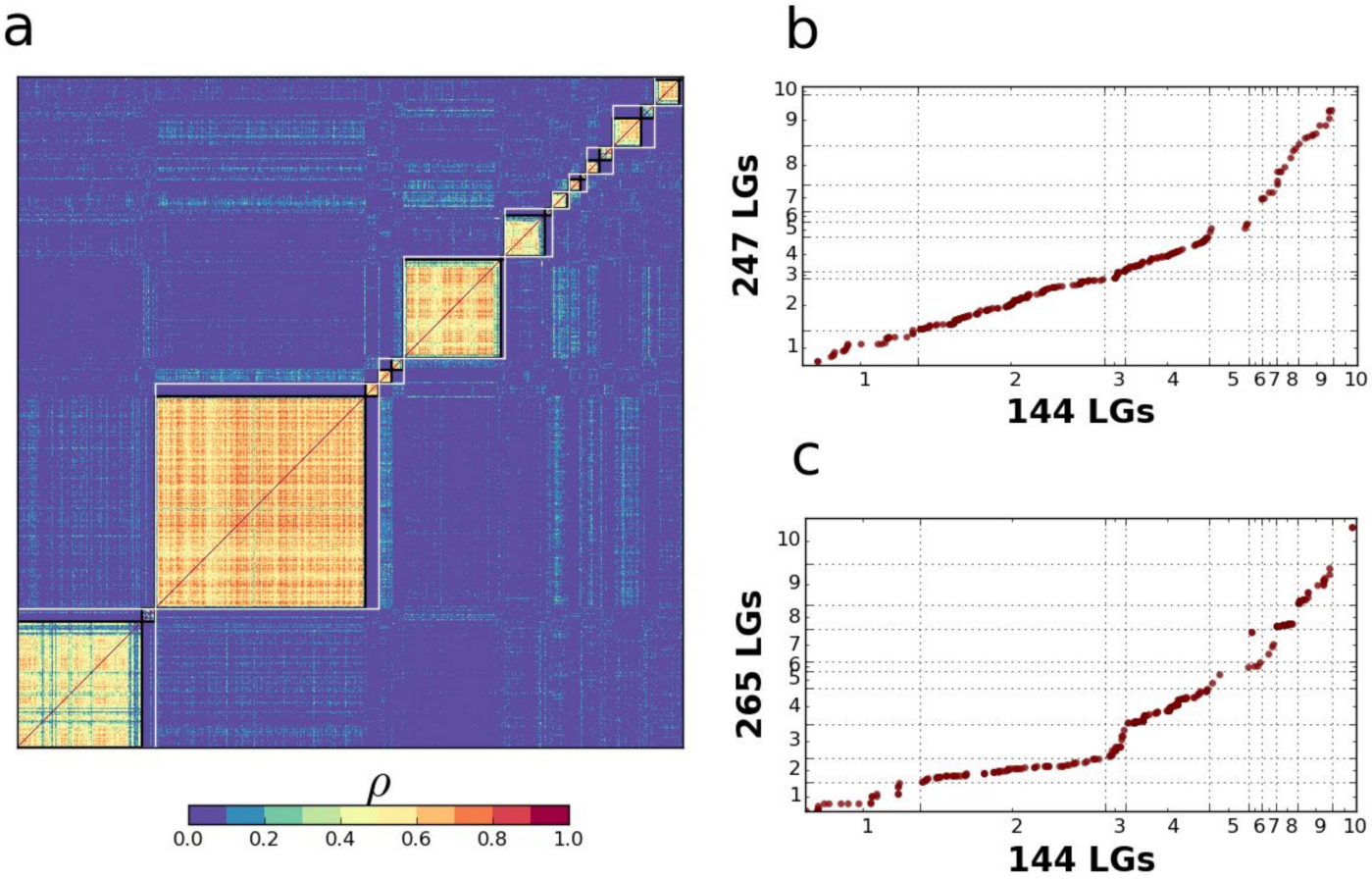
711

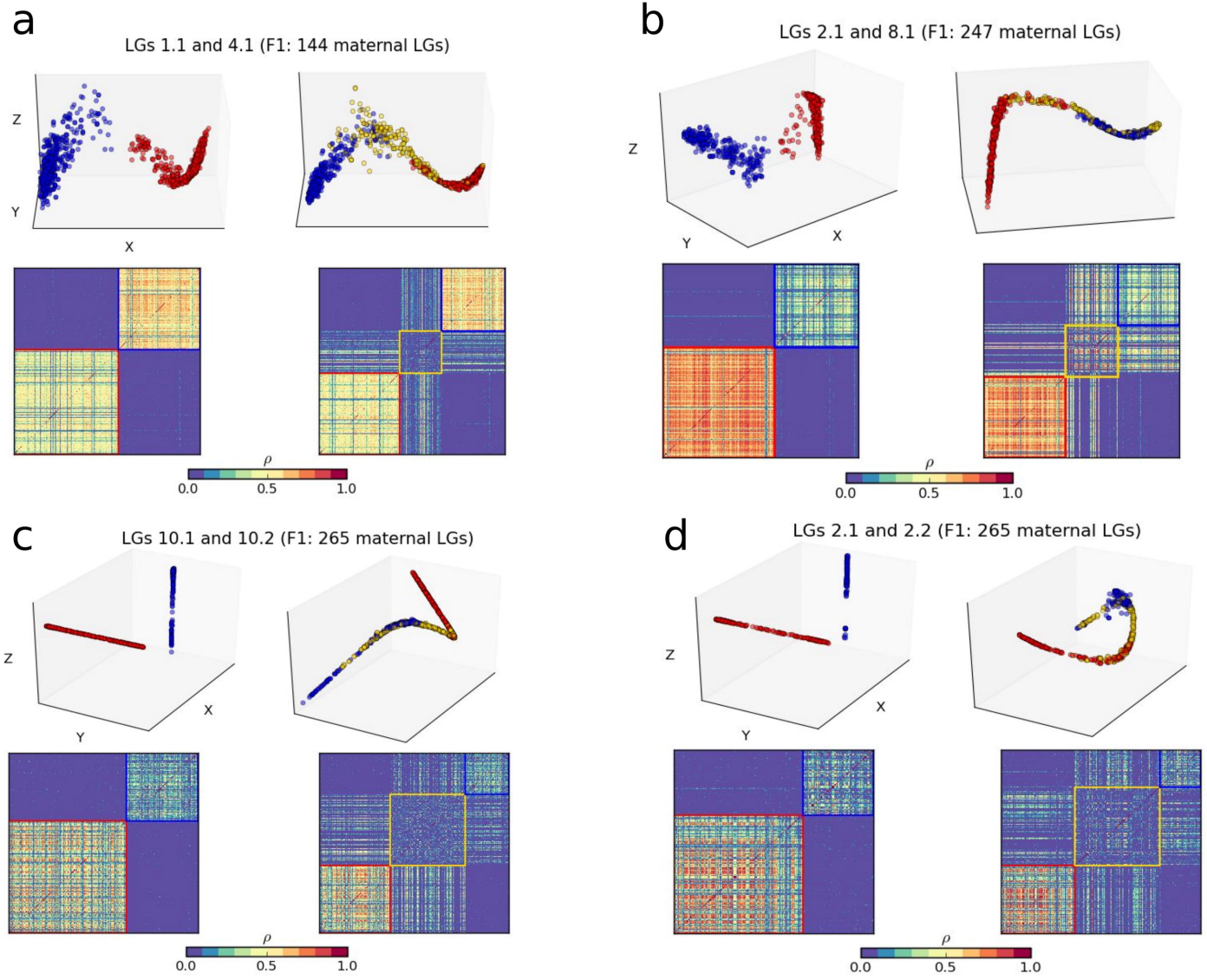
712

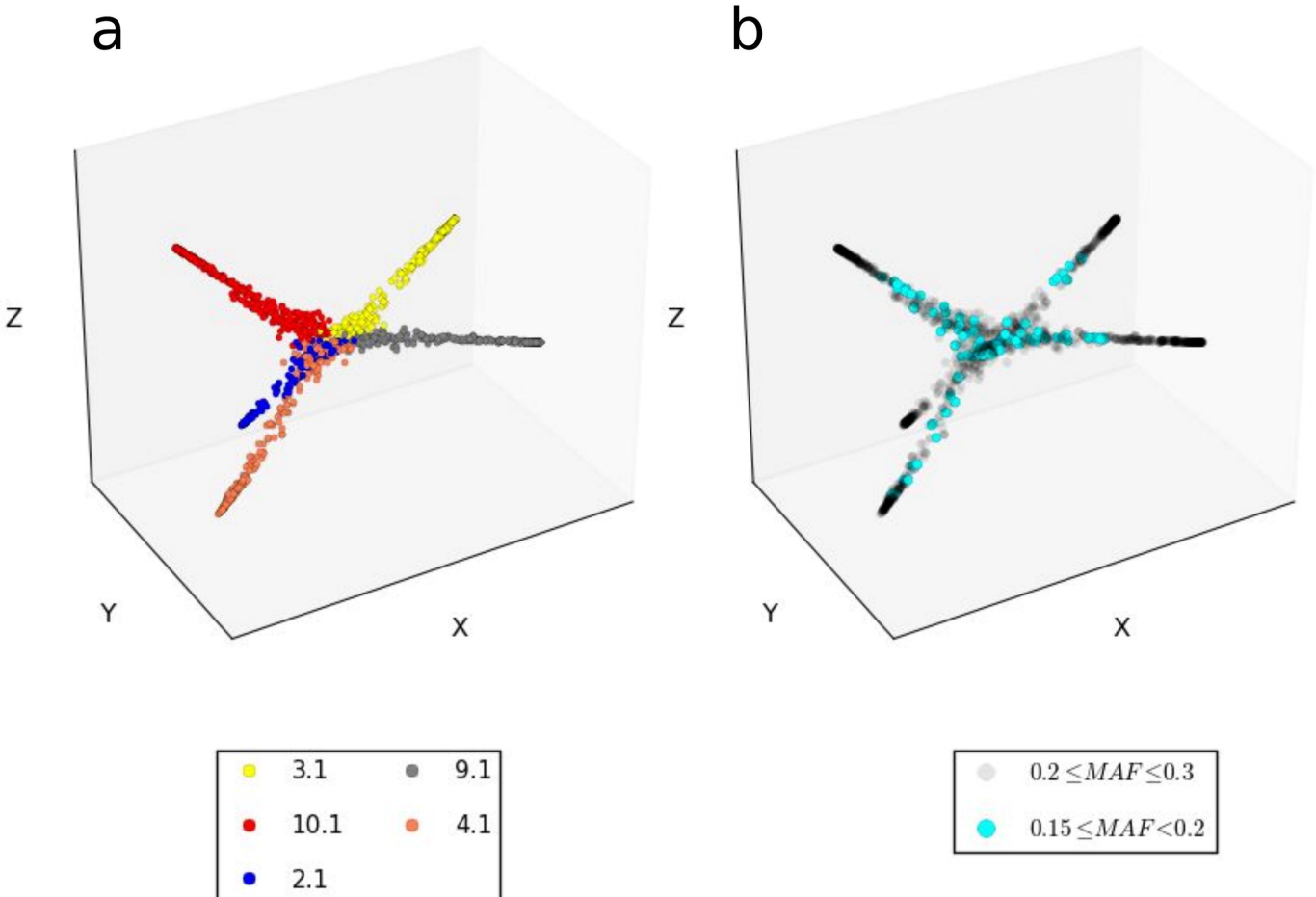
713

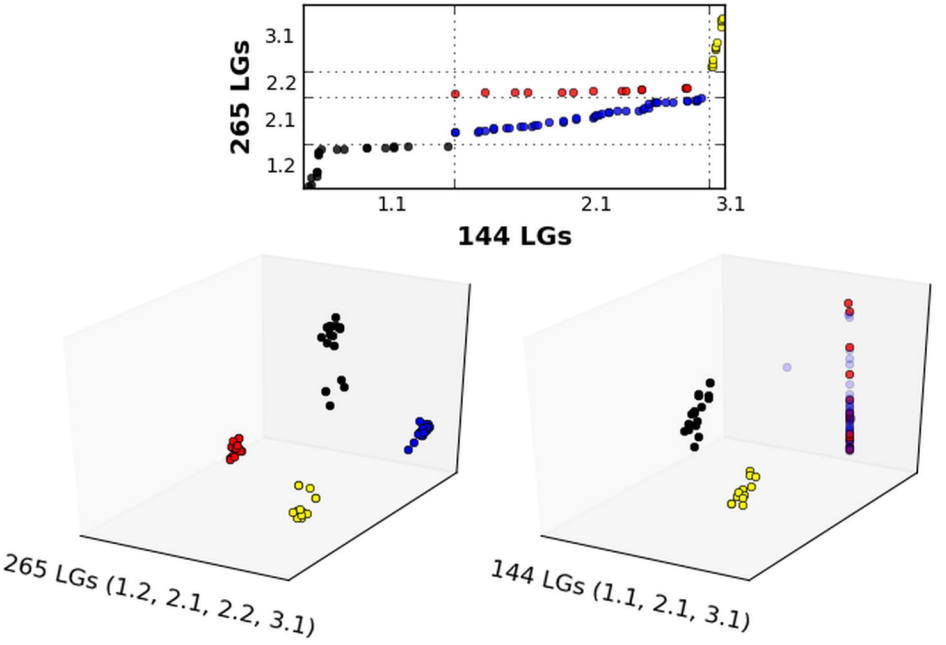

a
b
45°N
25°N
100°W
80°W
MT:5
ND:8
SD:1
MA:2
IA:1
NE:5
CO:11
KS:2
MO:1
AZ:1
NM:10
Canada
United States (MT,ND,SD,NE,IA,KS,MO)
United States (MA)
Unknown
F1 parents
United States (CO,AZ,NM)
Kazakhstan

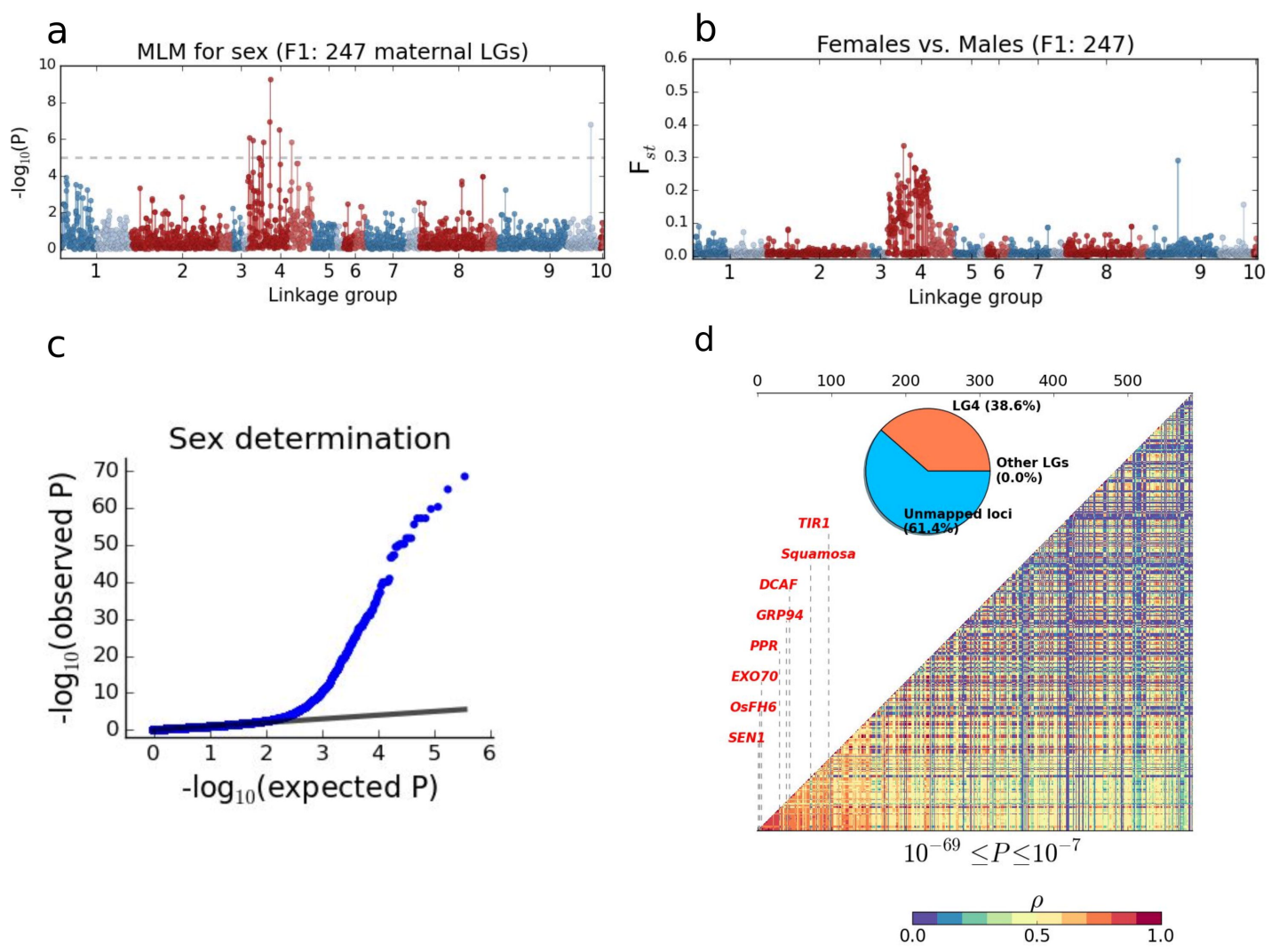

a
Population structure of crosses
Coord 2 (t-SNE)
Coord 1 (t-SNE)
265
264
TeaMaker x USDA_21422M
241
149
Zeus_1
106
Cascade
Chinook
DG
Nugget
144
DF
Zeus_2
248
255
247
249
147
145
b
TeaMaker x USDA_21422M
255
241
248
145
149
147
249
106
144
DG
DF
264
Chinook
265
Mother
Father
Progeny

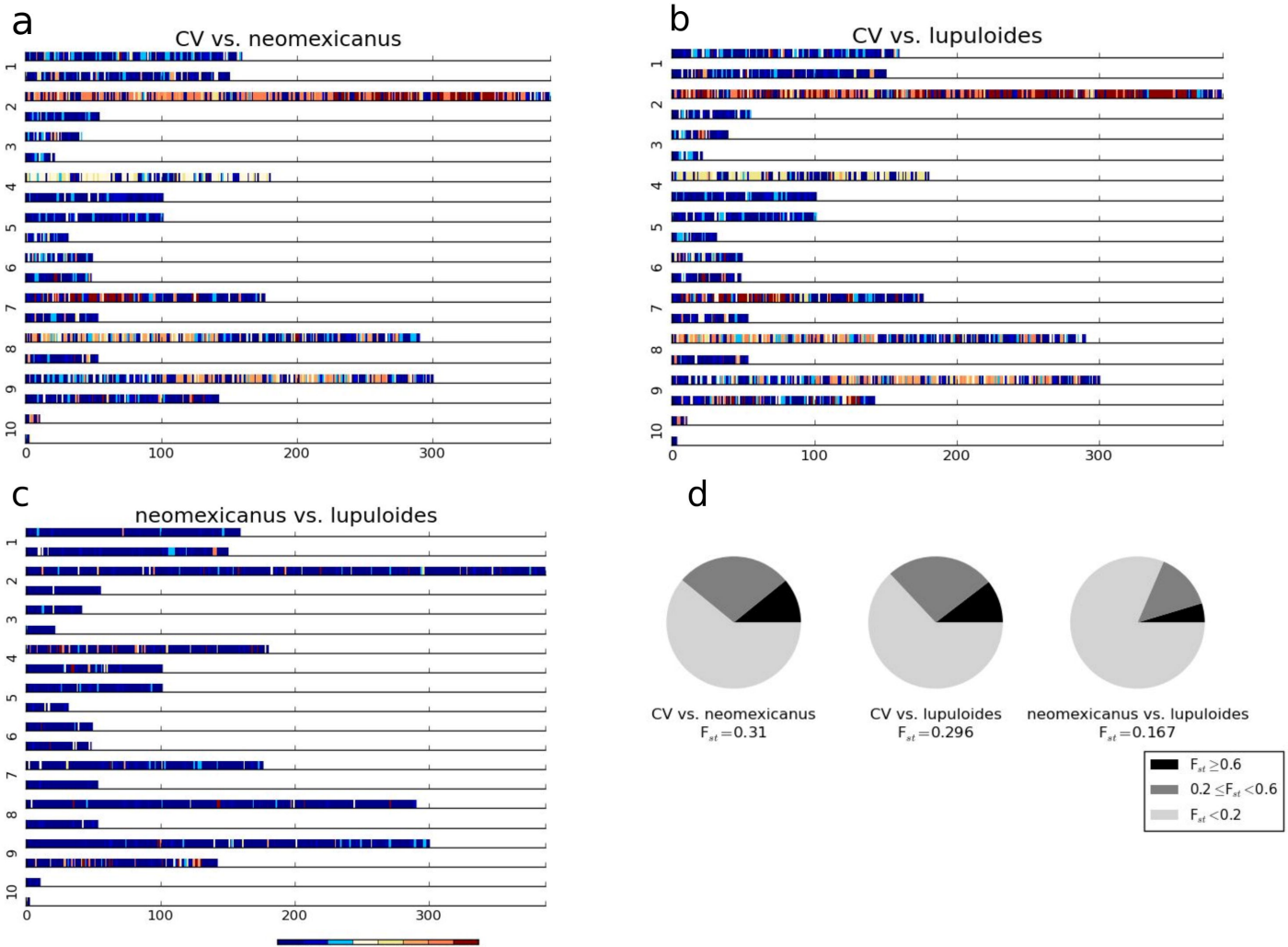

a
ρ
0.0
0.2
0.4
0.6
0.8
1.0
b
247 LGs
144 LGs
1
2
3
4
5
6
7
8
9
10
c
265 LGs
144 LGs

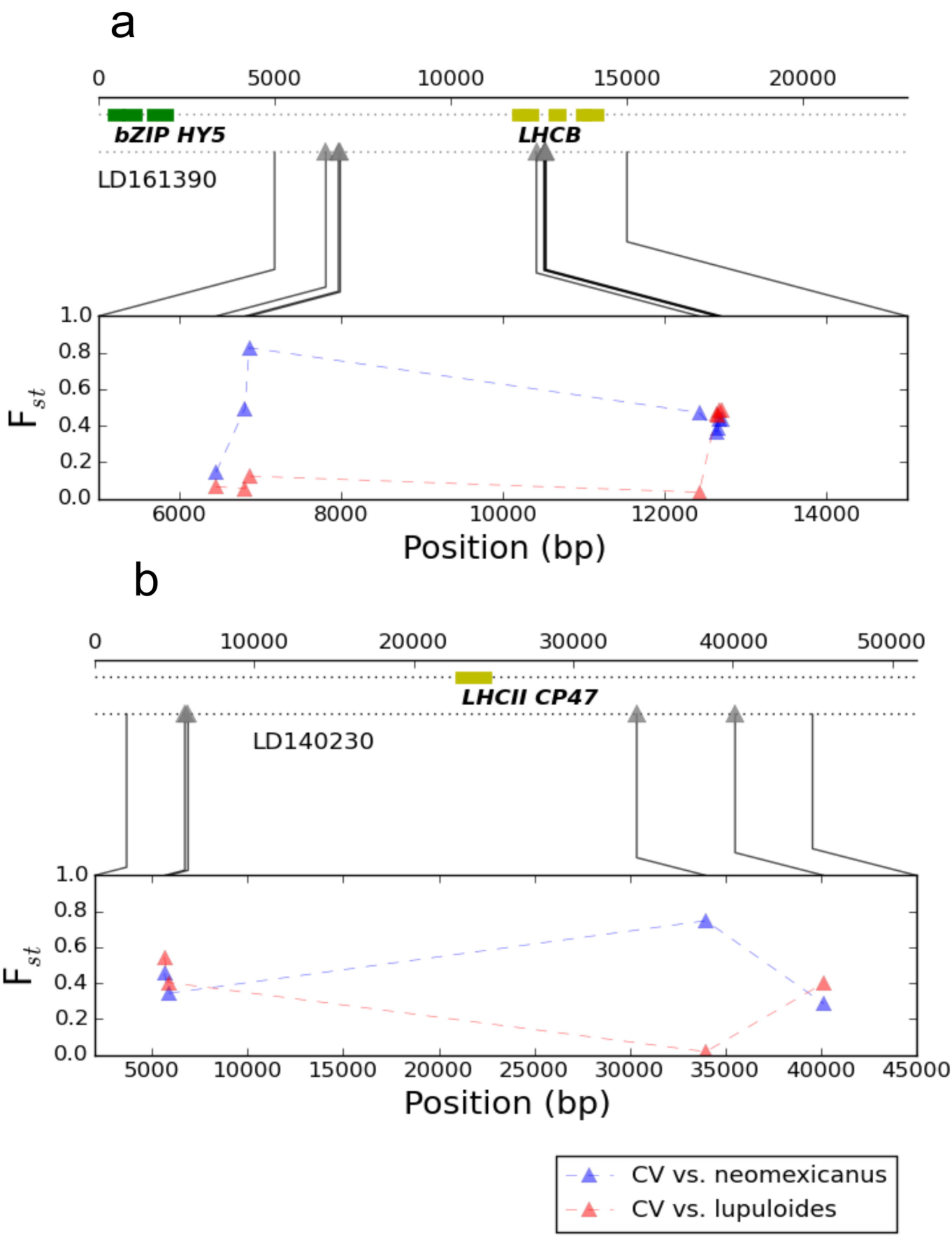

a
LGs 1.1 and 4.1 (F1: 144 maternal LGs)
Z
Y
X
ρ
0.0
0.5
1.0
b
LGs 2.1 and 8.1 (F1: 247 maternal LGs)
Z
Y
X
ρ
0.0
0.5
1.0
c
LGs 10.1 and 10.2 (F1: 265 maternal LGs)
Z
Y
X
ρ
0.0
0.5
1.0
d
LGs 2.1 and 2.2 (F1: 265 maternal LGs)
Z
Y
X
ρ
0.0
0.5
1.0


a
Z
Y
X
3.1
9.1
10.1
4.1
2.1
b
Z
Y
X
$0.2 \leq MAF \leq 0.3$
$0.15 \leq MAF < 0.2$


265 LGs
3.1
2.2
2.1
1.2
1.1
2.1
3.1
144 LGs
265 LGs (1.2, 2.1, 2.2, 3.1)
144 LGs (1.1, 2.1, 3.1)


a
MLM for sex (F1: 247 maternal LGs)
-log10(P)
Linkage group
b
Females vs. Males (F1: 247)
Fst
Linkage group
c
Sex determination
-log10(observed P)
-log10(expected P)
d
LG4 (38.6%)
Other LGs (0.0%)
Unmapped loci (61.4%)
TIR1
Squamosa
DCAF
GRP94
PPR
EXO70
OsFH6
SEN1
10^-69 ≤P≤10^-7
ρ


a
CV vs. neomexicanus
b
CV vs. lupuloides
c
neomexicanus vs. lupuloides
d
CV vs. neomexicanus
F_{st}=0.31
CV vs. lupuloides
F_{st}=0.296
neomexicanus vs. lupuloides
F_{st}=0.167
$F_{st} \geq 0.6$
$0.2 \leq F_{st} < 0.6$
$F_{st} < 0.2$


a
0
5000
10000
15000
20000
bZIP HY5
LHCB
LD161390
1.0
0.8
0.6
0.4
0.2
0.0
F_{st}
6000
8000
10000
12000
14000
Position (bp)
b
0
10000
20000
30000
40000
50000
LHCII CP47
LD140230
1.0
0.8
0.6
0.4
0.2
0.0
F_{st}
5000
10000
15000
20000
25000
30000
35000
40000
45000
Position (bp)
CV vs. neomexicanus
CV vs. lupuloides