

Exploring phenotypic and genetic variability in hemp (*Cannabis sativa*)

Nina Trubanová^{1†}, *Grace Pender*¹, *Paul F. McCabe*^{1,2}, *Rainer Melzer*^{1,2}, *Susanne Schilling*^{1,2*†}

¹School of Biology and Environmental Science, University College Dublin, Ireland

²UCD Earth Institute, University College Dublin, Ireland

*Corresponding author

Email: susanne.schilling@ucd.ie

† These authors contributed equally to this work.

Abstract

Hemp (*Cannabis sativa* L.) is a high-value multi-purpose crop with applications ranging from the medicinal use of its secondary metabolites to carbon-sequestering building materials. To fully capitalise on the potential of hemp as a crop for more sustainable agriculture, detailed phenotypic and genotypic characterizations are required in order to inform targeted breeding programmes.

Here, we present a detailed morphological and genomic analysis of 10 hemp cultivars. We found high variability in agronomically important traits such as flowering time, plant height, and biomass in most of the hemp cultivars tested. Additionally, genotyping by sequencing demonstrates that genetic differences are pervasive, both between hemp cultivars as well as between individuals of a single cultivar. The significant genetic and phenotypic variability we observe in hemp contrasts with other crops, where cultivars are often phenotypically and genetically relatively uniform. We argue that the variability of hemp is an asset for breeding and increases the potential for further improvement of the crop but is also a challenge for today's highly automated agriculture that relies on phenotypic uniformity.

Author summary

Hemp (*Cannabis sativa* L.) stands as one of the earliest domesticated crops. This remarkable plant is a sustainable crop with high carbon sequestration capacity which can be cultivated for soil remediation. Furthermore, hemp oil and fibre are used for many applications ranging from cooking to manufacturing bioplastics, textiles, or building materials of superb characteristics, and its secondary metabolites are sought after because of their medicinal properties. However, in contrast to many modern crops, hemp exhibits extensive variability in key agricultural traits, such as plant height and flowering time. This variability presents a challenge for both farmers and processors. To unravel the fundamentals of hemp diversity we conducted a comprehensive study of phenotypic and genetic characterisation of ten diverse hemp cultivars. We present findings confirming substantial variability not only among individuals of different cultivars but also within the same cultivar. Additionally, we explore heterozygosity in the context of other hemp studies and other crops. Understanding this variability in the context of a single hemp cultivar and across multiple cultivars is paramount for breeding novel, more uniform hemp varieties which will allow us to unlock the full potential of hemp as a crop of the future.

Introduction

Hemp is possibly one of the oldest crops utilised by humans and was likely initially domesticated in the early Neolithic period in East Asia, in what is modern-day China (1). Its fibre was, and still is, used to make textiles, ropes, fishnets, and paper. Meanwhile, the oil extracted from its seeds has served various purposes, including cooking and industrial applications. Additionally, the plant held significance for medicinal and ritual practices (2). Today, the range of utilisation and importance of hemp as a crop constantly increases; indeed within the last two decades, hemp has gained traction as a biomass and fibre crop, as well as for its medicinal uses (3–7). Because of hemp’s rapid biomass accumulation with low inputs in both conventional and organic farming (8), it has a high carbon sequestration capacity (3) and can be grown as an environmentally beneficial crop. Hemp is also suitable for phytoremediation (9,10), as a defence against erosion (11), and the waste biomass can be processed into biochar which can be used for soil amendment (12). From hemp biomass, it is also possible to produce bioethanol or biogas, and biodiesel from oil extracted from seeds (3,10,13,14). Hemp fibre is used for manufacturing textiles, paper, bioplastics, and biocomposites, as well as construction materials, such as hemp wool used for insulation, hemp clay bricks, or hempcrete (3,5,7–11,13,15,16). Waste products, such as hurd, are used as animal bedding (6,7,11,15,16). Nevertheless, the highest value hemp products are cannabinoids, terpenes, and flavonoids extracted from the hemp biomass, primarily from the female inflorescence, and also from other parts of the plant. Cannabinoids and other secondary metabolites are predominantly used for medicinal and recreational purposes (3–5,15,16), but can also be used as bio-pesticides (6), or for other applications. Hemp oil extracted from

seeds is sought-after not only in the cosmetic industry (7,11,13), but because of the ratio of ω -6 to ω -3 fatty acids and protein content (17,18) also as an addition to human and animal diets (4,7,11,16) either in a form of oil or food supplements (3,5). Additionally, besides hemp protein, hemp sprouts and microgreens rich in macronutrients, micronutrients, polyphenols, and antioxidants (4,19) can become a valuable source of nutrients in the human diet.

Meanwhile, novel applications are still emerging, such as functional textiles (10,15) and surgical devices thanks to the antibacterial properties of hemp (15), batteries from carbon nanosheets made of hemp pulp (5), non-petroleum hemp-based ink for tattooing, and hemp filaments for 3D-printing, or oil absorbent materials (5).

Though the plethora of purposes hemp can be utilised for and the importance of this crop is indisputable, hemp research lags behind other major crops. Because of the legal status of hemp (20) in the previous decades little research has been done to maximise the potential of use of this versatile crop. For the implementation of findings in practical applications, including breeding, more relevant hemp variability data must be collected and analysed. Previous studies have begun to explore population genomics and trait variability in order to get a better understanding of the genetics of hemp and to ultimately unlock the potential of this genetically diverse and variable crop (21,22).

While in nature genetic diversity is common among plant species and the basis of evolutionary change (23), in agriculture crops generally exhibit genetic uniformity in their main agronomic traits and this uniformity is favoured by farmers, processors, and consumers (24). Genetic diversity is the sum of the genetic characteristics of a species. In crops, genetic diversity can confer the ability to survive under diverse environmental conditions and in changing climates, as well as provide a genetic

reserve of resistance to various biotic and abiotic stressors. It also harbours a collection of genetic variability that is a prerequisite for breeding novel cultivars with improved morphological and agronomical traits (25). Whereas genetic diversity encompasses the distinction of the whole genotypes within the species, variability refers to one or a few different traits of the organism. Genetic variability is observable as distinct phenotypes resulting from contrasting alleles of a gene in individuals of the species or population. Utilising this variability allows breeders to target specific breeding targets.

A strong association of homozygous allelic variations with key agronomic traits has been found in many important crops, such as rice (26), cotton (27), and tomato (28). Among these crops, tomato (29) and certain rice cultivars, particularly its introgression lines (30), have relatively low levels of heterozygosity, while others, like cotton (31), are highly heterozygous. Additionally, other crops like cassava (32,33), mango (34) and grapevine (35,36) were found to be highly heterozygous. Another crop that was found to have high heterozygosity is hemp (*Cannabis sativa*) (21,22).

Here, we assess the phenotypic and genotypic variability of individual plants from ten different hemp cultivars and landraces. Besides intervarietal variability, we find a considerable amount of intravarietal variability, concerning phenotypic as well as genetic traits of the analysed plants. Our results confirm the potential of hemp genetic variability for breeding but caution that at the present individuals of a specific hemp cultivar can differ greatly in one or multiple traits.

Results

Hemp has been reported to display large phenotypic variability (21,37,38). To explore this variability, ten different hemp cultivars and landraces that varied in flowering time, sex determination, and primary agricultural purpose were cultivated in the greenhouse (Table 1): five monoecious French fibre cultivars that included early flowering ‘Fedora 17’, medium-late flowering ‘Santhica 27’ and ‘Felina 32’, ‘Futura’, and very late flowering ‘Futura 75’ (8,39,40), two accessions of the dioecious hungarian fibre cultivar ‘Kompolti’ (39) with different average flowering time, dioecious oilseed cultivar ‘Finola’, formerly ‘FIN-314’, a cross of two northern Russian landraces (41) that grows well in the arable regions of subarctic Canada and northern Europe (42), a dioecious landrace ‘Georgien’ from the Caucasus region in Georgia, and a dioecious landrace ‘Korea’ from Korea that is genetically more distinct to European and West Asian cultivars (8) (Fig 1a, S1 Fig).

Table 1. Key traits of ten hemp cultivars grown under greenhouse conditions.

	Hemp Cultivar	Flowering	Primary purpose	Plant height [cm]	Stem area [cm ²]	Fresh biomass weight [g]	Flowering [days]
Monoecious Industrial Cultivars	Fedora 17	early	fibre	196.7 ± 24.0	1.4 ± 0.2	256.7 ± 73.1	73.1 ± 18.2
	Felina 32	medium-late	fibre	165.0 ± 42.0	1.8 ± 0.4	245.6 ± 103.0	99.9 ± 19.2
	Futura	medium-late	fibre	170.0 ± 55.8	2.0 ± 0.5	268.6 ± 74.5	96.8 ± 20.7
	Futura 75	late	fibre	166.0 ± 33.6	1.9 ± 0.7	250.0 ± 105.6	116.0 ± 16.2
	Santhica 27	medium-late	fibre	158.7 ± 52.9	1.6 ± 0.8	220.0 ± 109.0	102.0 ± 15.0
Dioecious Land races	Finola	photoperiod insensitive	oilseed	122.8 ± 9.2	0.8 ± 0.1	158.3 ± 24.0	21.3 ± 4.1
	Kompolti (CAN 56)	late	fibre	133.3 ± 25.2	1.9 ± 0.9	213.3 ± 96.1	130.0 ± 0.0
	Kompolti (CAN 70)	medium	fibre	165.6 ± 65.1	1.6 ± 0.6	174.3 ± 123.0	87.1 ± 25.4
	Korea	late	-	174.6 ± 27.4	2.2 ± 0.5	282.5 ± 62.8	127.7 ± 8.1
	Georgien	early	-	180.0 ± 33.5	1.2 ± 0.5	263.3 ± 118.3	73.0 ± 23.7

Fig 1. Inter- and intravarietal differences in hemp plant height. The tallest (top) and smallest (bottom) plant per cultivar are depicted (a), scale bars 20 cm. Hemp plant height is represented as box plots measured on days 7, 13, 18, 21, 25, 32, 39, 45, 51, 58, and 74 after sowing. Each panel represents one hemp cultivar (b).

Hemp cultivars show high variability in biomass and fibre-related traits

Hemp plant height is a good indicator of fibre and biomass yield. Plants grew rapidly over the course of 130 days (Fig 1b). On average, plant height was 163.3 cm over all cultivars. The hemp plants showed high variability in plant height with the tallest individual being four times taller than the shortest individual on day 130 after sowing (Fig 1a). Plant height differences were high, with an average difference between the shortest and the tallest plant within cultivar of 97.2 cm (Fig 2a, S1 Table). On average, the standard deviation of plant height was 36.9 cm, with standard deviations up to 65.1 cm for 'Kompolti (CAN 70)', reflecting the variability found in this trait. In some cultivars, such as 'Santhica 27' or 'Futura', the tallest plant was more than three times taller than the shortest plant observed. Other cultivars, such as 'Finola', were more consistent in their height with the smallest individual being 110 cm tall and the tallest being 132 cm tall, and a mean of 122.8 cm (Table 1, S1 Table). Because of large variability within cultivars, significant differences in height between cultivars were only observed between 'Fedora17' and 'Finola' (Tukey's HSD, $p = 0.039$, Fig 2a).

Fig 2. Hemp plant architectural traits and flowering times vary within and between cultivars. Hemp plants were phenotyped for plant height at day 130 after sowing (a), fresh biomass yield (b), and length of the longest side branch (c). Flowering time was observed as the number of days between sowing and the onset of flowering per hemp cultivar (d). Significance levels are indicated with letters ($p < 0.05$).

Hemp biomass is an important factor for multiple applications, including secondary metabolites extraction (43,44), phytoremediation, and bioethanol production (45), as well as for calculating the carbon sequestration capacity (46). In the hemp plants cultivated, aboveground fresh biomass yield varied almost ten-fold between individuals of the same cultivar, with a mean of 230 g over all cultivars (Fig 2b, Table 1, S1 Table). Large intravarietal and intervariatal differences were observed. No significant differences were observed between cultivars.

Nodes, zones of emerging axillary buds that give rise to side branches, can disrupt the production of hemp fibres, affecting their length negatively (47). The number of nodes per metre determines the quality and length of hemp primary bast fibres and is therefore an important parameter when growing hemp for textile fibre production (47). Between all plants, the number of nodes ranged from four to ten per metre (S2a Fig). Within some cultivars, such as 'Kompolti (CAN 70)' or 'Futura', the highest number of nodes per metre was twice as high as the lowest number of nodes per metre (S2a Fig, S1 Table) showing the extensive intravarietal variability of this parameter.

Stem area is a good indicator of fibre yield if sowing density is optimised (48–50). The stem area 5 cm above ground ranged from 0.38 to 2.99 cm², with a mean of 1.94 cm² (S2b Fig, Table 1, S1 Table). The mean stem area of the landrace 'Korea' of 2.24 cm² was more than twice as large as the stem area of 'Finola' of 0.81 cm² (Table 1, S1 Table).

Plant architecture within the *Cannabis* genus is highly variable. To approximate plant architecture, the length of the longest side branch was measured. We observed that some individuals were slender while others were considerably wider (S1 Fig), with the longest side branches ranging from 10 to 80 cm (Fig 2c, S1 Table). The highest

intravarietal differences in the length of the longest side branch were measured in ‘Felina 32’ with an average difference of 55 cm between individuals (Fig 2c, S1 Table).

Hemp cultivars show high variability in flowering-related traits

All studied cultivars, with the exception of ‘Finola’ which is a photoperiod-insensitive cultivar, are short-day plants. Besides day length, hemp flowering time is affected by environmental conditions. Hemp cultivars grown in warmer climates with a higher number of cumulative sunlight hours tend to flower earlier and over a longer period than when grown in colder climates (51). Moreover, male individuals flower and senescence earlier than female individuals (52,53). ‘Finola’ male plants took 19 days or less to flower. In female ‘Finola’ plants, flowering time varied between 24 and 33 days after germination. For all other hemp cultivars, flowering time varied between 50 and 130 days (Fig 2d, Table 1, S1 Table). Late-flowering cultivars, namely ‘Futura 75’ and ‘Kompolti CAN 56’, and ‘Korea’ landrace were the latest flowering hemp cultivars. Some individual plants of the monoecious cultivars ‘Futura 75’ and ‘Santhica 27’ did not flower at all prior to the harvest on day 130, underlining a vast intravarietal variability within flowering time.

Inflorescence density is a good predictor of future seed and potential cannabinoids, terpenes, flavonoids, and sterols yield (54,55). The variability of flower density was higher in monoecious cultivars, such as ‘Felina 32’, ‘Futura’, ‘Futura 75’, and ‘Santhica 27’ (S3 Fig). The highest flower density was observed in female individuals of the cultivar ‘Finola’, which is grown for flowers and seeds.

Hemp is a naturally dioecious plant with male and female flowers located on separate plants (56). Monoecious cultivars with male and female flowers on the same plant that

originated from breeding are also available and favoured for some applications (57,58). Large sexual dimorphism affects phenotypic traits and plant morphology. In three out of five monoecious cultivars, 'Fedora 17', 'Felina 32', and 'Futura', only female flowers and no male flowers were observed in single individuals (S4 Fig).

Individual hemp plants are genetically variable

We detected substantial inter- and intravarietal phenotypic variation for every trait analysed. Therefore, we aimed to understand whether this phenotypic variation is mirrored at the genetic level. We employed genotyping-by-sequencing of 22 selected plants (S2 Table) to investigate single nucleotide variants (SNVs) within and between cultivars.

The number of SNVs per individual varied between 407,799 and 485,430 SNVs (Fig 3a, S2 Table). The number of SNVs per chromosome was significantly different between individuals (Chi-square, p -value $< 2.2e-16$ for all individuals). Visualisation of the density of SNVs showed that SNVs are overall unevenly distributed on all chromosomes of individual hemp plants (Figs 3b and 3c, S5 Fig). In general, density of SNVs was higher at putative subtelomeric regions and low in putative centromeric regions (59).

Fig 3. Short nucleotide variants (SNVs) of ten hemp cultivars. The number of filtered SNVs generated by genotyping-by-sequencing per hemp cultivar and of hemp cultivars previously published (21) were visualised as boxplots (a). The mean density of SNVs of all cultivars was visualised as barplots with SD per chromosome (b). The distribution of SNVs for each individual plant was visualised as a circos plot

(CircosVCF, 79) (c). The density of SNVs (b) and their distribution (c) are represented per 100,000 bp using the *C. sativa* reference genome (93).

The genetic variability of the studied plants was further examined by calculating their individual heterozygosity based on the ratio of heterozygous to homozygous SNVs (Fig 4a, S3 Table). The heterozygosity on average was 0.269, ranging from 0.189 in one of the ‘Santhica 27’ individuals to 0.338 in one of the ‘Kompolti CAN 70’ individuals. This is similar to previously published data (Fig 4b, S3 Table) (21,22). ‘Santhica 27’ displayed the lowest average heterozygosity (0.216). Conversely, the ‘Futura’ hemp cultivar exhibited the highest average heterozygosity (0.313). To contextualise this finding, we compared the observed heterozygosity in hemp to SNV data of thousands of accessions of 10 other crop species (S6 Fig). Crops with an average heterozygosity higher than collated hemp data (mean: 0.269, median: 0.267) were *Brassica napus* (mean: 0.543, median: 0.490), *Cucurbita pepo* (mean: 0.458, median: 0.439), and *Cucumis sativus* (mean: 0.376, median: 0.286). A number of cultivars from other species had a considerably lower heterozygosity than hemp cultivars, including watermelon (*Citrullus lantanus*), soybean (*Glycine max*), cotton (*Gossypium hirsutum*), and common bean (*Phaseolus vulgaris*). Also, some improved or inbred lines from maize (*Zea mays*) and wheat (*Triticum aestivum*) had very low levels of heterozygosity.

Fig 4. Hemp genomes have a high degree of heterozygosity. Variability in heterozygosity of ten hemp cultivars was calculated as a ratio of heterozygous to homozygous alternative alleles for ten hemp cultivars (a) and other publicly available hemp data (21,22) (b). Each VCF file from Sawler *et al.* (2015) (21) data and Lynch *et*

al. (2016) (22) data was intersected with a VCF file generated with FreeBayes (79,80) from our collection merged in the output to a single file using VCF-VCFintersect (83) to use limit analysis to sites present in GBS data. The resulting VCF files contained only sites identical to those found in our data.

Individuals from the same hemp cultivar differ significantly genetically

Observed phenotypic differences and differences in the number of SNVs and heterozygosity levels between hemp plants of the same cultivar were in many cases as high as between cultivars. Therefore, we analysed whether hemp plants can be associated with a specific cultivar based on the SNV genotyping. A principal component analysis (PCA) analysing SNV profiles of individual hemp plants based on the variance-standardised relationship matrix showed a strong trend of clustering for individuals belonging to the same variety (Fig 5a, S7 Fig). Individuals from the two accessions of 'Kompolti' (CAN 56 and CAN 70) created one cluster, similarly individuals of 'Futura' and 'Futura 75' clustered together. A large cluster was created by plants belonging to monoecious French hemp cultivars 'Fedora 17', 'Felina 32', and 'Santhica 27', and dioecious cultivar 'Finola' (Fig 5a). Dendrograms visualising results of different types of clustering, specifically based on pairwise dissimilarity (Fig 5b), pairwise identity-by-state (S8 Fig), and hierarchical hclust clustering (S8 Fig), confirmed results of the PCA and close genetic relationship of hemp plants belonging to the same cultivar.

Fig 5. Individual hemp plants from the same cultivar show a high degree of

similarity. A principal component analysis (PC1 and PC2) shows a clustering of individuals of the same hemp cultivar (a). For PC1 and PC3, and PC2 and PC3 see S7 Fig.

A hierarchical cluster dendrogram based on pairwise dissimilarity demonstrates clustering of individuals from the same cultivar (b). Relative amounts of alternative SNV positions overlap between individuals show an increased similarity between individuals from the same cultivar (c). The percentage of shared SNVs is expressed as a ratio of shared SNVs to all SNVs of both individuals combined in percentage; i.e. if individuals share 100% of positions the value is 100, if they share 50% of positions the value is 33.3.

The large amount of differences in the number of SNVs and heterozygosity of individuals prompted us to investigate whether individuals differed in their SNV positions. The chromosomal positions of SNVs (hetero- and homozygous) in each sequenced individual were compared to SNV positions of all other individuals. The number of shared SNVs was then expressed as a ratio to all SNVs of the two individuals as percentage (Fig 5c). Individuals of the same cultivar showed an overlap, but not a full congruence of SNV positions (Fig 5c). For example, between two 'Felina 32' individuals, only 32.09% of SNV positions were identical. However, the percentage of SNV positions identity between individuals of the same cultivar was overall higher than between individuals of different cultivars. European cultivars had a higher percentage of shared SNVs (Fig 5c) than the landraces 'Korea' or 'Georgien' which is in line with the results of the PCA and clustering (Fig 5a).

Discussion

In this study, we conducted a comprehensive phenotypic analysis involving more than 80 individuals representing ten distinct hemp cultivars. To unravel the genetic variability of these cultivars we genotyped 22 of these individuals using genotyping-by-sequencing.

Our data show that individuals within a given cultivar exhibited a higher degree of genetic similarity to one another when compared to individuals from different cultivars. The relationships constructed on the basis of our sequencing data align closely with documented breeding histories, for instance, the close relationships of French fibre cultivars.

Hemp phenotypic traits are diverse, both between as well as within cultivars

Hemp is a fast-growing crop which accumulates biomass rapidly (Figs 1 and 2a). Therefore, hemp has a huge potential as a carbon-sequestering crop, with a multitude of industrial applications (3). Its short growth cycle makes it particularly interesting as a break crop since it has soil-remediating properties (9,53).

While hemp cultivars are firmly established and used in both agriculture and industry, we found that specific cultivars often do not significantly differ in agronomically important traits, including plant height and biomass yield, plant architecture, and flowering time (Fig 2). In fact, for most traits we investigated, the variability among individuals within the same cultivar was so pronounced that significant differences between cultivars could not be detected (Fig 2, S2 Fig). Consequently, while certain

general trends can be observed among hemp cultivars, the extensive intravarietal variability in all the observed phenotypic traits often makes it impossible to classify an individual plant based on our data as belonging to a specific cultivar. It is worth noting that other traits not addressed in this study, such as cannabinoid content or fibre quality, may or may not exhibit similar levels of variability (13,32).

Hemp genomes display a high degree of heterozygosity

We hypothesise that the substantial phenotypic differences among individuals are, to a significant extent, genetically determined. Our SNV genotyping reveals a relatively high degree of individual heterozygosity. This heterozygosity implies that offspring will exhibit variability in genetically determined traits controlled by genes that are heterozygous in the parental plants. While a certain degree of heterozygosity offers advantages for plant populations (24,61), homogeneity is desirable in agricultural and industrial contexts. In the case of drug-type *C. sativa* (marijuana), the industry standard involves working with clonal propagation of mother plants that exhibit the desired phenotype, rather than relying on seed propagation. This approach is generally impractical for field-grown, high-volume crops like fibre hemp cultivars of *C. sativa*. Reducing the heterozygosity of hemp cultivars through breeding efforts introduces a potential avenue to increase the profitability of the entire value chain by improving harvesting and processing steps.

Conversely, selfed populations characterised by low heterozygosity can suffer from inbreeding depression (62), manifested in reduced yields compared to cross-pollinated progeny (63). Hence, a certain level of heterozygosity may be desirable. However, achieving this balance requires the development of low heterozygosity parental lines

and the optimisation of the subsequent seed production in the context of hemp cultivation.

C. sativa is primarily a dioecious plant, although monoecious populations exist as well (64). In monoecious cultivars, lower hemp heterogeneity resulting from autogamy (65) allows the harvest of both seeds and stems for fibre (54). Self-pollination has proven effective as a method for determining the presence and levels of cannabinoids in autogamous populations. It also serves as a breeding technique to create starting material with stable absence or presence of these compounds (66). Interestingly, in our study, both dioecious and monoecious cultivars exhibited comparable levels of heterozygosity (Fig 4a).

We compared the SNV positions among sequenced individuals. Our findings revealed that similarities were most pronounced among individuals of the same cultivar, followed by individuals of closely related cultivars. The lowest number of SNV positions overlapping were detected between individuals of dissimilar cultivars (Fig 5c). The relationships derived from our sequencing data (Fig 5b) corroborate known breeding histories. For instance, French fibre cultivars such as ‘Felina 32’, ‘Santhica’, and ‘Futura’ (39) are more closely related to each other than to landraces ‘Korea’ and ‘Georgien’. The higher density of SNVs in distal parts of the chromosomes and lower density in centromeric regions (Fig 3c) is in consensus with a high recombination rate at chromosome ends and reduced rates in the putative centromeric regions (67).

High heterozygosity levels are similar to other highly heterozygous crops such as cucumber and *Brassicaceae* (S6 Fig). The heterozygosity in hemp is however higher than that of many crops, including maize and soybean, where even landraces do not possess heterozygosity levels as high as observed in hemp. Numerous studies have previously described high rates of heterozygosity, especially in *Brassicaceae* and

Cucurbitaceae species, and linked them to diverse phenotypes between cultivars of the respective crop species (68–71). In *Brassicaceae*, the high level of heterozygosity in commercial cultivars stems from the generation of F1 hybrids (71). High heterozygosity as a sign of heterosis effects is one strategy for crop improvement in these crops, while in other cases, high heterozygosity might be counterproductive (72).

Genetic and phenotypic variability presents both opportunities and challenges in hemp breeding

The extensive range of phenotypic trait values and the underlying genetic heterogeneity makes farming more challenging. To illustrate, variations in flowering times among individuals of the same cultivar (Fig 2d) inevitably impact the timing of harvest and processing for inflorescences and seeds. Therefore, achieving uniformity in traits related to hemp inflorescences, such as flowering time, inflorescence density, and consistent sex expression in monoecious cultivars, should be a central focus of breeding efforts. Furthermore, for research and genetic analyses, sequences from inbred individuals can be utilised to improve the current genomic assemblies (38) that are prone to misassembly due to heterozygosity (73).

We observed significant variability not only among individuals from different cultivars but also among individuals of the same cultivar. Consequently, when conducting comparative studies involving different hemp cultivars, as well as cultivars of other highly heterozygous crops, relying on a single plant (21,22,74) or just a few individuals (16,66) as representative samples is insufficient.

To expand the *C. sativa* gene pool, a comprehensive genomics-based assessment of the *C. sativa* germplasm is essential to define the available diversity (73). This

403 approach would facilitate the identification of superior alleles contributing to hybrid
 404 vigour and the development of elite cultivars, much like what has been achieved in
 405 rice (4). Our research highlights the substantial diversity within hemp genomes,
 406 offering valuable resources for both breeding and research. This genetic wealth can
 407 be leveraged to develop novel cultivars that exhibit high-value characteristics and
 408 uniformity in traits crucial to various applications. Hemp's notable phenotypic plasticity
 409 positions it favourably for successful cultivation across diverse climates and
 410 environmental conditions.

Materials and methods

Plant growth

Ten cultivars of hemp (*C. sativa*) were grown in a greenhouse between May and September 2019. Seeds of ‘Futura’ (CAN 69), ‘Georgien’ (CAN 22), ‘Korea’(CAN 23), and both accessions of ‘Kompolti’ (CAN 56 and CAN 70) cultivars were obtained from IPK Gatersleben, and seeds of ‘Fedora 17’, ‘Finola’, ‘Futura 75’, ‘Felina 32’, and ‘Santhica 27’ cultivars were obtained commercially. Plants were grown in 2L pots with a custom soil mix of 1:1:1 John Innes No. 2:Vermiculite:Perlite and fertilised once a week with water-soluble fertiliser (NPK 24-8-16, Water Soluble All Purpose Plant Food, Scotts Miracle-Gro Products, Inc.). The position of plants was randomised and plants were rotated multiple times during the trial to avoid bias associated with their position in the greenhouse.

Phenotypic measurements and records

Plant height was measured at several time intervals during the plants’ growth on 7 days, 13 days, 18 days, 21 days, 25 days, 32 days, 39 days, 45 days, 51 days, 58 days, 74 days, and at harvest. Plant height was measured from the base of the plant to the top of the main stem. Flowering was defined as the presence of a female flower with visible stigmata or a male flower bulb with a length of at least 3 mm. The day when flowers were first observed was recorded as flowering time for each plant expressed as a number of days after sowing. After day 67, flowering was observed on day 74, day 102 and day 130, hence observed flowering times are upper estimates.

Each plant was sexed and recorded as dioecious if being either male or female, or monoecious if both, male and female flowers, were observed on the same plant. Data related to other phenotypic parameters, such as flower density, number of nodes per one metre, longest side branch length, stem area, and above-ground biomass were collected at the time of the harvest. Flower density was subjectively categorised using the 'low', 'medium-low', 'medium', 'medium-high', and 'high' density scale. The number of nodes per one metre was taken from the bottom of the stem. Stem surface was calculated using ImageJ from aerial photos of plant stems cut ~5 cm above the soil line.

Tissue harvesting and DNA extraction

After about eight weeks, tissue samples were harvested from the young leaves. Samples were snap-frozen in liquid nitrogen and stored at -80°C before DNA extraction. DNA was extracted from each sample using a DNeasy Plant Mini Kit (Qiagen, Germany). DNA was quality-controlled using gel electrophoresis and photospectrometry.

Sequencing and data acquisition

Sequencing data were obtained using genotyping by sequencing (GBS) method essentially as previously described (74), but DNA was digested with MseI and the sequencing platform used was Illumina HiSeq. Paired-end 150 bp sequencing was performed (CD Genomics, USA). The raw sequence has been deposited in the NIH Sequence Read Archive (SRA), under BioProject PRJNA1023932. Publicly available genotyping by sequencing (GBS) data (21) and whole-genome shotgun sequencing

(WGS) data (22) for hemp were obtained from the NCBI. SNV data for other crops were retrieved from the Plant Imputation database (75).

Mapping and variant calling

A collection of trimmed sequences was aligned to cs10 *C. sativa* reference genome assembly GCF_900626175.2 using the Burrows-Wheeler Alignment (BWA-MEM) tool (76–78). Variants were called with FreeBayes (79,80) on each file individually for calculating heterozygosity and SNV densities, and on all files together for PCA, dendrograms, and SNVs positions comparison. All analyses were performed using the platforms UseGalaxy.org and UseGalaxy.eu (81).

Heterozygosity analysis

For calculating heterozygosity, variants were filtered for quality > 30 and read depth > 5 or 9 for GBS and WGS respectively using bcftools annotate (82) and VCFfilter (83). Out of all types of called variants, only SNVs were kept using bcftools annotate (82). Homozygous reference sites were removed using the bcftools filter (82). Each VCF file from Sawler *et al.* (2015) (21) data and Lynch *et al.* (2016) (22) data was intersected with a VCF file generated with FreeBayes (79,80) from our collection merged in the output to a single file using VCF-VCFintersect (83) to use limit analysis to sites present in GBS data. The resulting VCF files contained only sites identical to those found in our data. The heterozygosity of each plant was calculated as a ratio between the number of genotyped heterozygous SNVs and the number of non-reference homozygous SNVs using HetHomAlleles (83). The same tool was used to calculate individual heterozygosity from SNV data obtained from the Plant Imputation database (75).

483

484 **PCA and clustering**

485 Firstly, variants were filtered for quality > 30 and read depth > 2 using bcftools annotate
 486 (82) and VCFfilter (83). Further filtering was performed in PLINK 1.9 (84–86) after
 487 which data included only SNVs with minor allele frequency $\geq$ 2.5% and 80%
 488 genotyping rate. Principal components analysis (PCA) based on the variance-
 489 standardised relationship matrix was performed in PLINK 1.9 (84–86) and generated
 490 eigenvectors and eigenvalues were visualised in RStudio v2021.9.0.351 (87) using
 491 the tidyverse package (88).

492 Dendrograms visualising clustering results from identity-by-state matrix and
 493 dissimilarity matrices were created in RStudio v2021.9.0.351 (87). Dendrogram
 494 visualising hierarchical clustering performed in RStudio was visualised using FigTree
 495 v1.4.4 (89).

496

497 **Density and position of single nucleotide variants**

498 The density of SNVs per 100,000 bp, filtered the same way as above, was visualised
 499 with CircosVCF (90) with each circle representing an individual plant. The mean
 500 density of SNVs per 100,000 bp of the length of each chromosome of the *C. sativa*
 501 reference genome (93) was visualised using the RStudio v2021.9.0.351 (87) and
 502 tidyverse package (88). Chi-square statistics for the number of SNVs per chromosome
 503 between samples were calculated in RStudio v2021.9.0.351 (87) using the tidyverse
 504 package (88). Overlaps in positions of SNVs between individual samples were
 505 calculated using the VCF Toolz (91).

Acknowledgements

NT is funded by the UCD School of Biology and Environmental Science. SS was supported by a postdoctoral scholarship of the Irish Research Council and GreenLight Medicines (Ireland) under the Enterprise Partnership Scheme (grant no. EPSPD/2019/220, Enterprise Partner GreenLight Medicines) during part of the project. We are grateful to Jack Forsythe who helped with DNA extraction. We thank Prof. Mali Salmon-Divon from Ariel University who was instrumental for troubleshooting CircosVCF. Open access funding was provided by IReL.

Conflict of interest

This research was partially funded by GreenLight Medicines (Ireland); however, the company was not involved in the study design and analysis.

Author contributions

RM, PFM and SS conceptualised the project. SS performed the experiments. NT and GP performed the data analysis. NT, SS and RM drafted the manuscript. All authors edited and finalised the manuscript.

References

1. Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, et al. Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Sci Adv*. 2021 Jul;7(29):eabg2286.
2. Li HL. An archaeological and historical account of cannabis in China. *Econ Bot*. 1973 Oct;28(4):437–48.
3. Schilling S, Dowling CA, Shi J, Ryan L, Hunt D, O'Reilly E, et al. The Cream of the Crop: Biology, Breeding and Applications of *Cannabis sativa* [Internet]. Preprints; 2020 Oct [cited 2022 May 7]. Available from: <https://www.authorea.com/users/360584/articles/483690-the-cream-of-the-crop-biology-breeding-and-applications-of-cannabis-sativa?commit=6532fb15968ae2aa7e3410d50241bbd53a2ff5cf>
4. Cerino P, Buonerba C, Cannazza G, D'Auria J, Ottoni E, Fulgione A, et al. A Review of Hemp as Food and Nutritional Supplement. *Cannabis Cannabinoid Res*. 2021 Feb 1;6(1):19–27.
5. Schluttenhofer C, Yuan L. Challenges towards Revitalizing Hemp: A Multifaceted Crop. *Trends Plant Sci*. 2017 Nov;22(11):917–29.
6. Amaducci S, Scordia D, Liu FH, Zhang Q, Guo H, Testa G, et al. Key cultivation techniques for hemp in Europe and China. *Ind Crops Prod*. 2015 Jun;68:2–16.
7. Small E. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization. *Bot Rev*. 2015 Sep;81(3):189–294.
8. Ranalli P. *Advances in Hemp Research*. CRC Press; 1999.
9. Ranalli P, Venturi G. Hemp as a raw material for industrial applications. *Euphytica*. 2004 Jan;140(1–2):1–6.
10. Salentijn EMJ, Zhang Q, Amaducci S, Yang M, Trindade LM. New developments in fiber hemp (*Cannabis sativa* L.) breeding. *Ind Crops Prod*. 2015 Jun;68:32–41.
11. Karus M, Vogt D. European hemp industry: Cultivation, processing and product lines. :6.
12. Atoloye IA, Adesina IS, Sharma H, Subedi K, Liang CL (Kathleen), Shahbazi A, et al. Hemp biochar impacts on selected biological soil health indicators across different soil types and moisture cycles. Peña-Fernández A, editor. *PLOS ONE*. 2022 Feb 28;17(2):e0264620.
13. Amaducci S, Scordia D, Liu FH, Zhang Q, Guo H, Testa G, et al. Key cultivation techniques for hemp in Europe and China. *Ind Crops Prod*. 2015 Jun;68:2–16.
14. Prade T, Svensson SE, Mattsson JE. Energy balances for biogas and solid biofuel production from industrial hemp. *Biomass Bioenergy*. 2012;40:36–52.
15. Andre CM, Hausman JF, Guerriero G. *Cannabis sativa*: The Plant of the Thousand and One Molecules. *Front Plant Sci* [Internet]. 2016 Feb 4 [cited 2022 May 7];7. Available from: <http://journal.frontiersin.org/Article/10.3389/fpls.2016.00019/abstract>
16. Carus M, Sarmiento L. *The European Hemp Industry: Cultivation, processing and applications for fibres, shivs, seeds and flowers*. European Industrial Hemp Association; 2016.
17. Callaway JC. Hempseed as a nutritional resource: An overview. *Euphytica*. 2004 Jan;140(1–2):65–72.

18. Galasso I, Russo R, Mapelli S, Ponzoni E, Brambilla IM, Battelli G, et al. Variability in Seed Traits in a Collection of *Cannabis sativa* L. Genotypes. *Front Plant Sci* [Internet]. 2016 May 20 [cited 2023 Oct 5];7. Available from: <http://journal.frontiersin.org/Article/10.3389/fpls.2016.00688/abstract>
19. Frassinetti S, Moccia E, Caltavuturo L, Gabriele M, Longo V, Bellani L, et al. Nutraceutical potential of hemp (*Cannabis sativa* L.) seeds and sprouts. *Food Chem*. 2018 Oct;262:56–66.
20. Herer J. *The Emperor Wears No Clothes*. 11th Edition. Ah Ha Publishing, Quick American Archives; 2007.
21. Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, et al. The Genetic Structure of Marijuana and Hemp. Tinker NA, editor. *PLOS ONE*. 2015 Aug 26;10(8):e0133292.
22. Lynch RC, Vergara D, Tittes S, White K, Schwartz CJ, Gibbs MJ, et al. Genomic and Chemical Diversity in *Cannabis*. *Crit Rev Plant Sci*. 2016 Nov 1;35(5–6):349–63.
23. Darwin CR. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. 1st ed. London: John Murray; 1859.
24. Managing Global Genetic Resources—Agricultural Crop issues and Policies. *Environ Int*. 1995 Jan;21(1):104.
25. Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, Glenn KC. Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci*. 2021 Mar;61(2):839–52.
26. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*. 2010 Nov;42(11):961–7.
27. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*. 2013 Jan;45(1):43–50.
28. Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, et al. Genomic analyses provide insights into the history of tomato breeding. *Nat Genet*. 2014 Nov;46(11):1220–6.
29. The 100 Tomato Genome Sequencing Consortium, Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, et al. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J*. 2014 Oct;80(1):136–48.
30. Ali J, Aslam UM, Tariq R, Murugaiyan V, Schnable PS, Li D, et al. Exploiting the Genomic Diversity of Rice (*Oryza sativa* L.): SNP-Typing in 11 Early-Backcross Introgression-Breeding Populations. *Front Plant Sci*. 2018 Jun 22;9:849.
31. Nuraida D, Abdurrajak Y, Amin M, Hastutik U. Genetic Variation Analysis of Superior Cotton Varieties of *Gossypium hirsutum* through Microsatellite Markers. *Int J Plant Biol*. 2017 Apr 12;8(1):6996.
32. Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol*. 2016 May;34(5):562–70.
33. Hu W, Ji C, Liang Z, Ye J, Ou W, Ding Z, et al. Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. *Genome Biol*. 2021 Dec;22(1):316.
34. Wang P, Luo Y, Huang J, Gao S, Zhu G, Dang Z, et al. The genome evolution

and domestication of tropical fruit mango. *Genome Biol.* 2020 Dec;21(1):60.

35. Salmaso M, Faes G, Segala C, Stefanini M, Salakhutdinov I, Zyprian E, et al. Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. :11.

36. Laucou V, Launay A, Bacilieri R, Lacombe T, Adam-Blondon AF, Bérard A, et al. Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. Chiang TY, editor. *PLOS ONE*. 2018 Feb 8;13(2):e0192540.

37. Petit J, Salentijn EMJ, Paulo MJ, Thouminot C, van Dinter BJ, Magagnini G, et al. Genetic Variability of Morphological, Flowering, and Biomass Quality Traits in Hemp (*Cannabis sativa* L.). *Front Plant Sci.* 2020 Feb 20;11:102.

38. Vergara D, Feathers C, Huscher EL, Holmes B, Haas JA, Kane NC. Widely assumed phenotypic associations in *Cannabis sativa* lack a shared genetic basis. *PeerJ*. 2021 Apr 20;9:e10672.

39. de Meijer EPM. Fibre hemp cultivars: A survey of origin, ancestry, availability and brief agronomic characteristics. *J Int Hemp Assoc.* 1995;(2(2)).

40. Butkutė B, Liaudanskienė I, Jankauskienė Z, Gruzdevienė E, Cesevičienė J, Amalevičiūtė K. Features of Carbon Stock in the Biomass of Industrial Hemp and Stinging Nettle. In: Sayigh A, editor. *Renewable Energy in the Service of Mankind Vol I: Selected Topics from the World Renewable Energy Congress WREC 2014* [Internet]. Cham: Springer International Publishing; 2015. p. 17–29. Available from: https://doi.org/10.1007/978-3-319-17777-9_2

41. Callaway JC, Laakkonen TT. Cultivation of cannabis oil seed varieties in Finland. *J Int Hemp Assoc.* 1996;(3):32–4.

42. Callaway JC. Hemp as Food at High Latitudes. *J Ind Hemp.* 2002 Jan;7(1):105–17.

43. Chen C, Pan Z. Cannabidiol and terpenes from hemp – ingredients for future foods and processing technologies. *J Future Foods.* 2021 Dec;1(2):113–27.

44. Seaman C. Cultivation Stress Techniques and the Production of Secondary Metabolites in *Cannabis sativa*. In: *Recent Advances in the Science of Cannabis*. 1st Edition. CRC Press; 2021. p. 30.

45. Rheay HT, Omondi EC, Brewer CE. Potential of hemp (*Cannabis sativa* L.) for paired phytoremediation and bioenergy production. *GCB Bioenergy.* 2021 Apr;13(4):525–36.

46. Pervaiz M, Sain MM. Carbon storage potential in natural fiber composites. *Resour Conserv Recycl.* 2003 Nov;39(4):325–40.

47. Amaducci S, Errani M, Venturi G. Plant Population Effects on Fibre Hemp Morphology and Production. *J Ind Hemp.* 2002 Jun;7(2):33–60.

48. van der Werf HMG, Wijnhuizen M, de Schutter JAA. Plant density and self-thinning affect yield and quality of fibre hemp (*Cannabis sativa* L.). *Field Crops Res.* 1995 Mar;40(3):153–64.

49. Crowley JG. The performance of *Cannabis sativa* (hemp) as a fibre source for medium density fibre board (MDF). Carlow: Crops Research Centre; 2003.

50. Burczyk H, Grabowska L, Strybe M, Konczewicz W. Effect of Sowing Density and Date of Harvest on Yields of Industrial Hemp. *J Nat Fibers.* 2009 Jun 12;6(2):204–18.

51. Dhondt F, Muthu SS. Optimal Conditions for Hemp Fibre Production. In: Dhondt F, Muthu SS, editors. *Hemp and Sustainability* [Internet]. Singapore: Springer Singapore; 2021. p. 37–50. Available from: https://doi.org/10.1007/978-981-16-3334-8_3

52. Bócsa I, Karus M. The Cultivation of Hemp: Botany, Varieties, Cultivation and Harvesting. Sebastopol: HempTech; 1998.
53. Salentijn EMJ, Petit J, Trindade LM. The Complex Interactions Between Flowering Behavior and Fiber Quality in Hemp. *Front Plant Sci*. 2019 May 16;10:614.
54. Faux AM, Draye X, Lambert R, d'Andrimont R, Raulier P, Bertin P. The relationship of stem and seed yields to flowering phenology and sex expression in monoecious hemp (*Cannabis sativa* L.). *Eur J Agron*. 2013 May;47:11–22.
55. Jin D, Dai K, Xie Z, Chen J. Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes. *Sci Rep*. 2020 Dec;10(1):3309.
56. Moliterni VMC, Cattivelli L, Ranalli P, Mandolino G. The sexual differentiation of *Cannabis sativa* L.: A morphological and molecular study. *Euphytica*. 2004 Jan;140(1–2):95–106.
57. Baldini M, Ferfua C, Piani B, Sepulcri A, Dorigo G, Zuliani F, et al. The Performance and Potentiality of Monoecious Hemp (*Cannabis sativa* L.) Cultivars as a Multipurpose Crop. *Agronomy*. 2018 Aug 24;8(9):162.
58. Ferfua C, Zuliani F, Danuso F, Piani B, Cattivello C, Dorigo G, et al. Performance and Stability of Different Monoecious Hemp Cultivars in a Multi-Environments Trial in North-Eastern Italy. *Agronomy*. 2021 Jul 16;11(7):1424.
59. Grassa CJ, Wenger JP, Dabney C, Poplawski SG, Motley ST, Michael TP, et al. A complete *Cannabis* chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content [Internet]. *Genomics*; 2018 Oct [cited 2023 Sep 26]. Available from: <http://biorxiv.org/lookup/doi/10.1101/458083>
60. Petit J, Salentijn EMJ, Paulo MJ, Denneboom C, Trindade LM. Genetic Architecture of Flowering Time and Sex Determination in Hemp (*Cannabis sativa* L.): A Genome-Wide Association Study. *Front Plant Sci*. 2020 Nov 4;11:569958.
61. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. *Genet Res Int*. 2015 Mar 19;2015:1–14.
62. Carlson CH, Stack GM, Jiang Y, Taşkıran B, Cala AR, Toth JA, et al. Morphometric relationships and their contribution to biomass and cannabinoid yield in hybrids of hemp (*Cannabis sativa*). Costa F, editor. *J Exp Bot*. 2021 Dec 4;72(22):7694–709.
63. Kurtz LE, Mahoney JD, Brand MH, Lubell-Brand JD. Comparing Genotypic and Phenotypic Variation of Selfed and Outcrossed Progeny of Hemp. *HortScience*. 2020 Aug;55(8):1206–9.
64. de Meijer EPM, Bagatta M, Carboni A, Crucitti P, Moliterni VMC, Ranalli P, et al. The Inheritance of Chemical Phenotype in *Cannabis sativa* L. *Genetics*. 2003 Jan 1;163(1):335–46.
65. Horkay E. Establishing the share of self- and cross-fertilization by means of population genetics in a monoecious hemp stand [Monoecious, dioecious, inbreeding depression, hybrid effect, seed and dry stem yielding ability]. *Novenytermeles*. 1986;(35):177–82.
66. Mishchenko SV. Phenotypic manifestation of the traits of presence and content of cannabinoids in the process of self-pollination in monoecious hemp and selection. *Plant Breed Seed Prod*. 2020 Dec 30;0(118):58–73.
67. Lavery KU, Stout JM, Sullivan MJ, Shah H, Gill N, Holbrook L, et al. A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res*. 2019 Jan;29(1):146–56.

68. Chikh-Rouhou H, Mezghani N, Mnasri S, Mezghani N, Garcés-Claver A. Assessing the Genetic Diversity and Population Structure of a Tunisian Melon (*Cucumis melo* L.) Collection Using Phenotypic Traits and SSR Molecular Markers. *Agronomy*. 2021 May 31;11(6):1121.
69. Pan Y, Wang Y, McGregor C, Liu S, Luan F, Gao M, et al. Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective. *Theor Appl Genet*. 2020 Jan;133(1):1–21.
70. Pandey S, Ansari WA, Pandey M, Singh B. Genetic diversity of cucumber estimated by morpho-physiological and EST-SSR markers. *Physiol Mol Biol Plants*. 2018 Jan;24(1):135–46.
71. Izzah NK, Lee J, Perumal S, Park JY, Ahn K, Fu D, et al. Microsatellite-based analysis of genetic diversity in 91 commercial *Brassica oleracea* L. cultivars belonging to six varietal groups. *Genet Resour Crop Evol*. 2013 Oct;60(7):1967–86.
72. Fridman E. Consequences of hybridization and heterozygosity on plant vigor and phenotypic stability. *Plant Sci*. 2015 Mar;232:35–40.
73. Hurgobin B, Tamiru-Oli M, Welling MT, Doblin MS, Bacic A, Whelan J, et al. Recent advances in *Cannabis sativa* genomics research. *New Phytol*. 2021 Apr;230(1):73–89.
74. Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, et al. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. Liu Z, editor. *PLoS ONE*. 2013 Jan 23;8(1):e54603.
75. Gao Y, Yang Z, Yang W, Yang Y, Gong J, Yang QY, et al. Plant-ImputeDB: an integrated multiple plant reference panel database for genotype imputation. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D1480–8.
76. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009 Jul 15;25(14):1754–60.
77. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010 Mar 1;26(5):589–95.
78. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio [Internet]*. 2013 May 26 [cited 2022 Feb 21]; Available from: <http://arxiv.org/abs/1303.3997>
79. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *ArXiv12073907 Q-Bio [Internet]*. 2012 Jul 20 [cited 2022 Feb 21]; Available from: <http://arxiv.org/abs/1207.3907>
80. Tange O. GNU Parallel: The Command-Line Power Tool. *Login USENIX Mag*. 2011;36(1):6.
81. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018 Jul 2;46(W1):W537–44.
82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078–9.
83. Garrison E. *vcflib* [Internet]. GitHub repository; 2015. Available from: <https://github.com/ekg/vcflib>
84. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007 Sep;81(3):559–75.
85. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-

771 generation PLINK: rising to the challenge of larger and richer datasets.
772 GigaScience. 2015 Dec;4(1):7.
773 86.PLINK: Whole genome data analysis toolset [Internet]. [cited 2022 Sep 24].
774 Available from: <https://zzz.bwh.harvard.edu/plink/>
775 87.RStudio Team. RStudio: Integrated Development for R [Internet]. RStudio,
776 PBC, Boston, MA; 2020. Available from: <http://www.rstudio.com/>
777 88.Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al.
778 Welcome to the Tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.
779 89.FigTree [Internet]. [cited 2022 Sep 24]. Available from:
780 <http://tree.bio.ed.ac.uk/software/figtree/>
781 90.Drori E, Levy D, Smirin-Yosef P, Rahimi O, Salmon-Divon M. CircosVCF:
782 circos visualization of whole-genome sequence variations stored in VCF files.
783 Hancock J, editor. Bioinformatics. 2017 May 1;33(9):1392–3.
784 91.Davis S. vcftoolz: a Python package for comparing and evaluating Variant Call
785 Format files. J Open Source Softw. 2019 Mar 26;4(35):1144.
786 92.Minevich G, Park DS, Blankenberg D, Poole RJ, Hobert O. CloudMap: A
787 Cloud-Based Pipeline for Analysis of Mutant Genome Sequences. Genetics.
788 2012 Dec 1;192(4):1249–69.
789 93.Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Timothy
790 Motley S, et al. A new *Cannabis* genome assembly associates elevated
791 cannabidiol (CBD) with hemp introgressed into marijuana. New Phytol. 2021
792 May;230(4):1665–79.

Supporting information captions

S1 Fig. Imaging of hemp plants used for this study. All hemp plants used for this study were imaged on day 130 after sowing. Individual plants are grouped by cultivar.

S2 Fig. Hemp plant architectural traits vary within and between cultivars. The ten hemp cultivars showed differences in morphology and agriculturally important traits. Variability was observed for the number of nodes measured or calculated per 1-metre length (a) and the area of the stem cut ~5 cm above the soil line at the end of the trial (b). The number of nodes indicates the number of side branches. Significance levels are indicated with letters ($p < 0.05$).

S3 Fig. Flower density varies within and between cultivars. Flower density of female and monoecious hemp plants was recorded at the end of the cultivation period. The size of points increases with the number of individuals with a given flower density.

S4 Fig. Sex expression of monoecious hemp is not uniform. Flower sex expression of monoecious hemp cultivars at the end of the cultivation period is plotted as a stacked bar graph, indicating individuals with male and female flowers (dotted), female only flowers (crosshatched) and non-flowering individuals (blank).

S5 Fig. SNV density plotted for individual cultivars. The distribution of SNVs plotted on a circos plot (CircosVCF, 79) using the *C. sativa* reference genome (93): 'Fedora 17' (a), 'Felina 32' (b), 'Finola' (c), 'Futura' (d), 'Futura 75' (e), 'Georgien' (f), 'Kompolti CAN 56' (g), 'Kompolti CAN 70' (h), 'Korea' (i), 'Santhica 27' (j). Density of

filtered SNVs is represented per 100,000 bp.

S6 Fig. Heterozygosity found in *C. sativa* and other crops. The heterozygosity of *C. sativa sativa* (this study, 21, 22) in comparison to an accession of a diverse panel of different crops including cultivars and landraces (Plant Imputation database, (75)). The comparability of hemp data is limited because the amount of individuals of other crops was nearly thousandfold higher.

S7 Fig. Principal component analysis based on SNVs for factors PC2 and PC3. Principal components from the PCA show clustering of individuals that belong to the same hemp cultivar. PC1 and PC3 (a), PC2 and PC3 (b). For PC1 vs PC2 see Figure 5a.

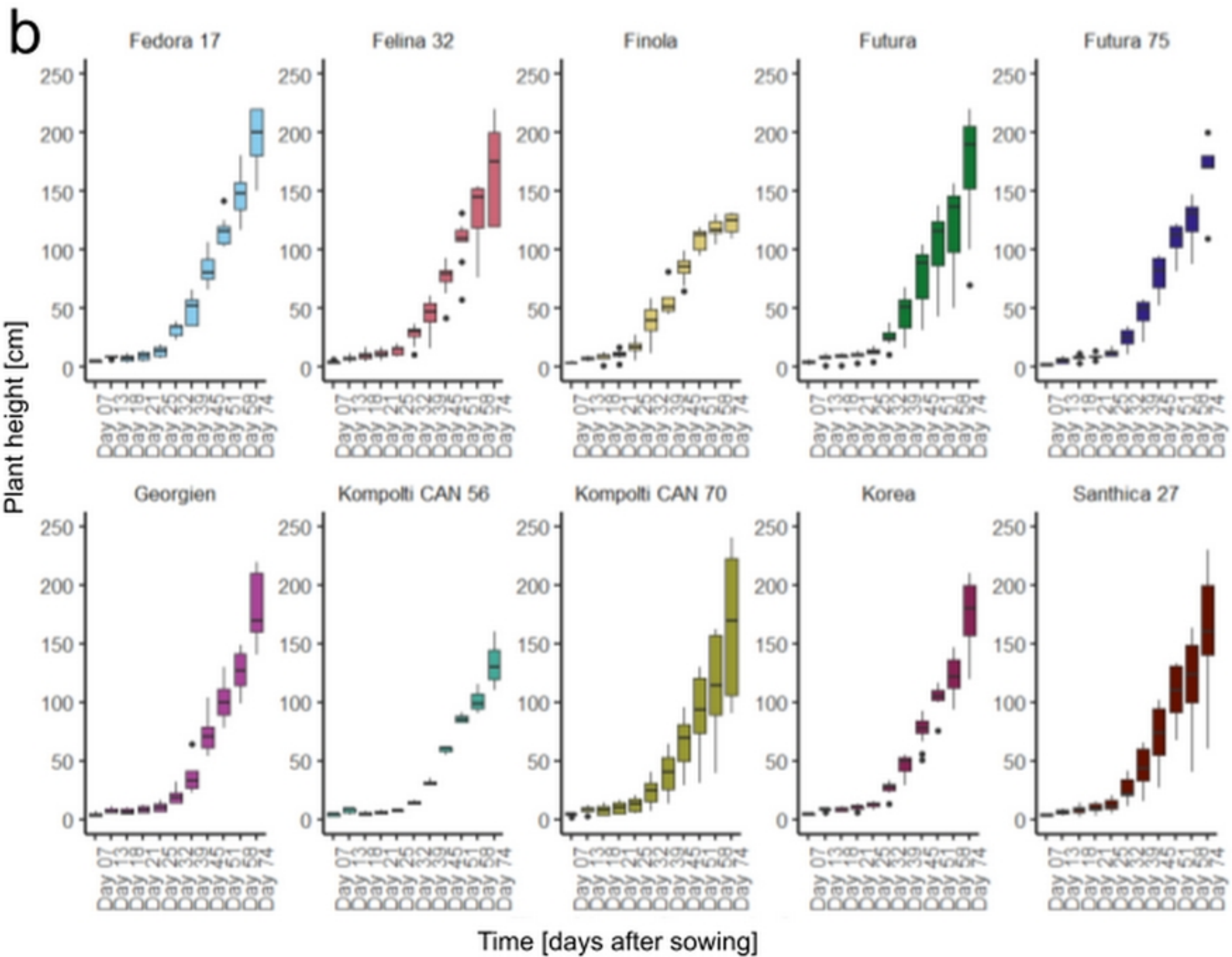
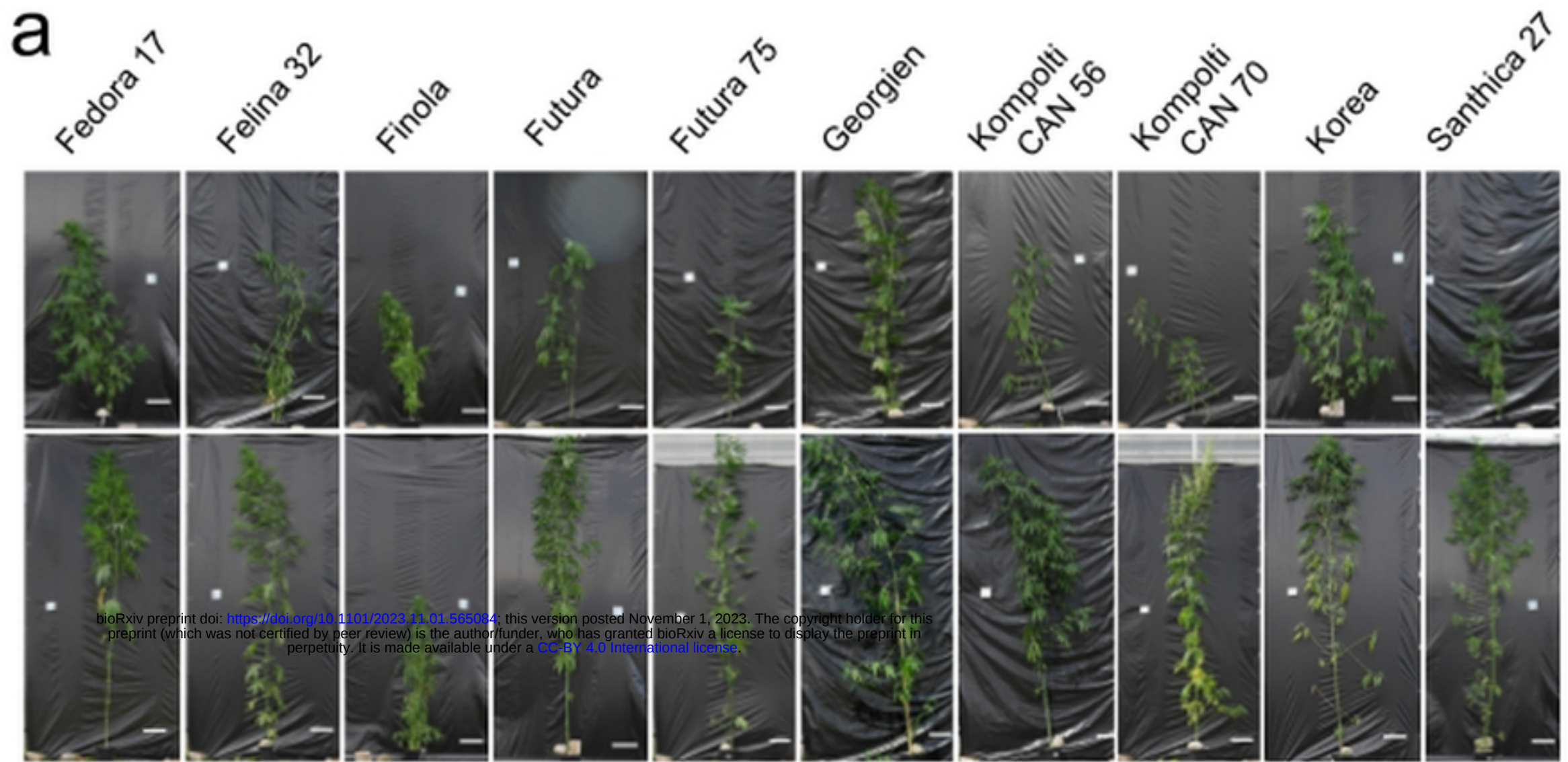
S8 Fig. Hierarchical clustering of individual hemp plants based on SNVs. Hierarchical cluster dendrogram based on pairwise identity-by-state (IBS) values (a) and a dendrogram based on hierarchical hclust clustering values (b) from SNV data of all sequenced individuals show a clustering of individuals that belong to the same hemp cultivar.

S1 Table. Detailed phenotypic and developmental traits of ten hemp cultivars grown under greenhouse conditions.

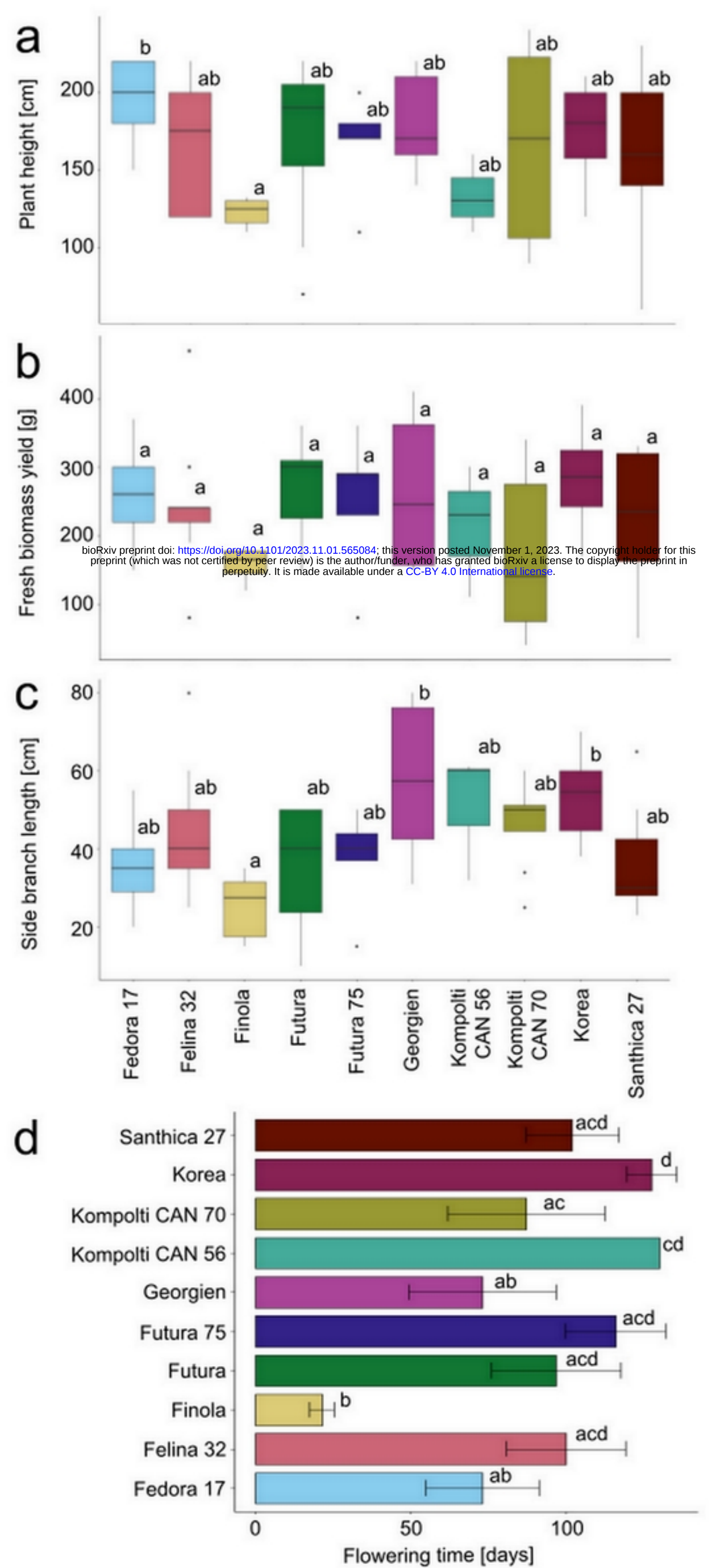
S2 Table. The number of SNVs per individual hemp plants.

S3 Table. Heterozygosity of hemp cultivars calculated as a ratio of heterozygous

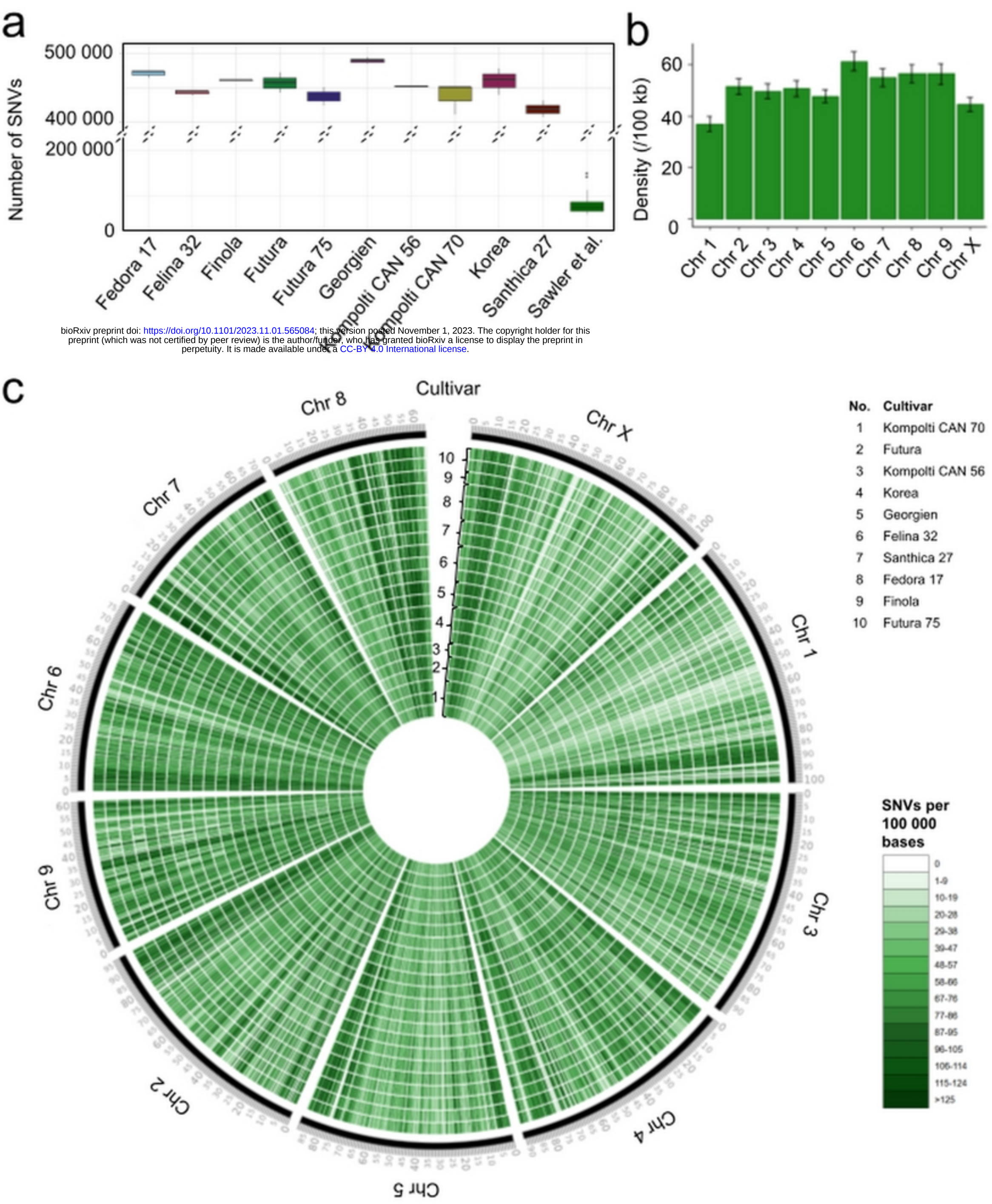
843 **to homozygous alternative alleles for hemp cultivars published here, in Sawler**
844 ***et al.* (2015) (21) and Lynch *et al.* (2016) (22).**



Figure



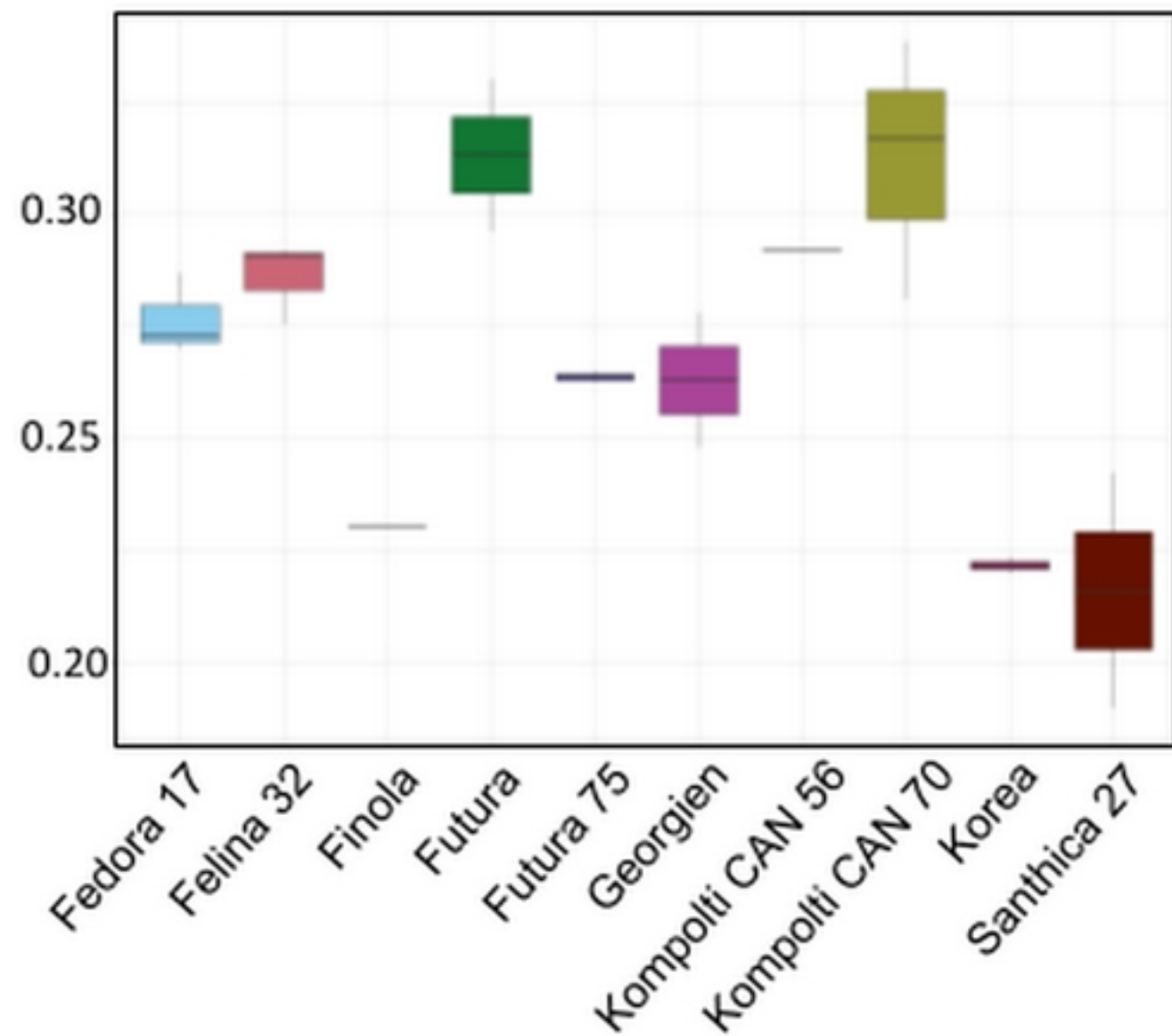
Figure



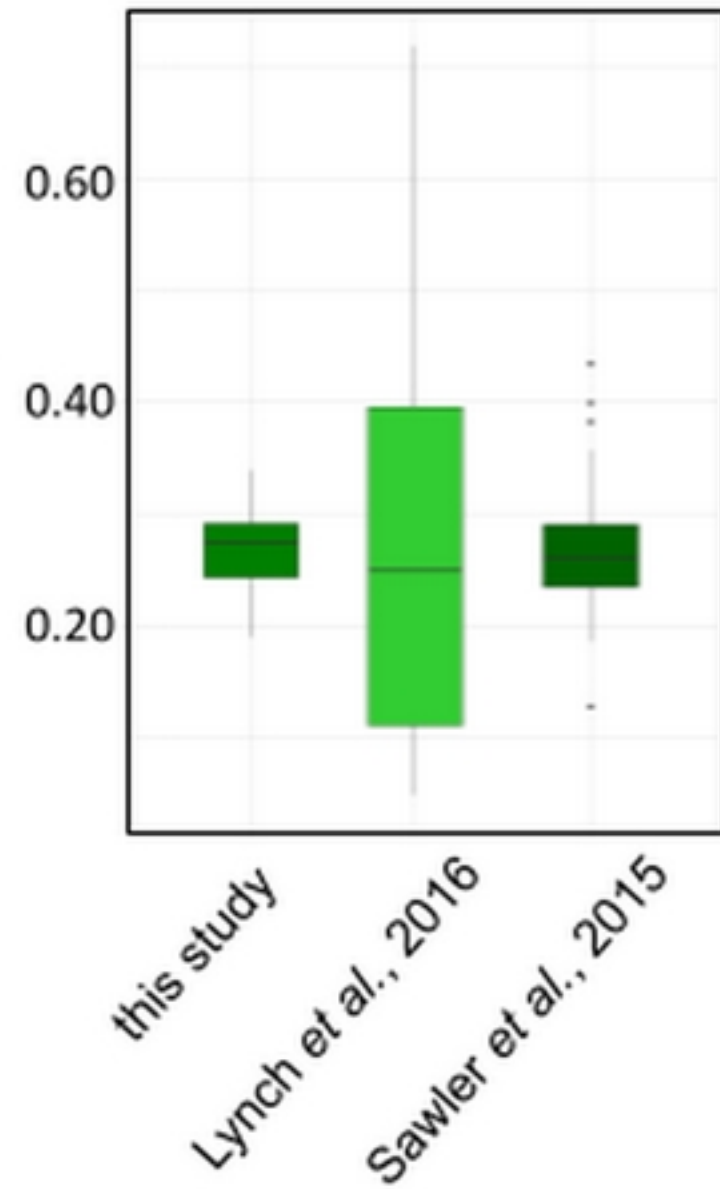
Figure

a

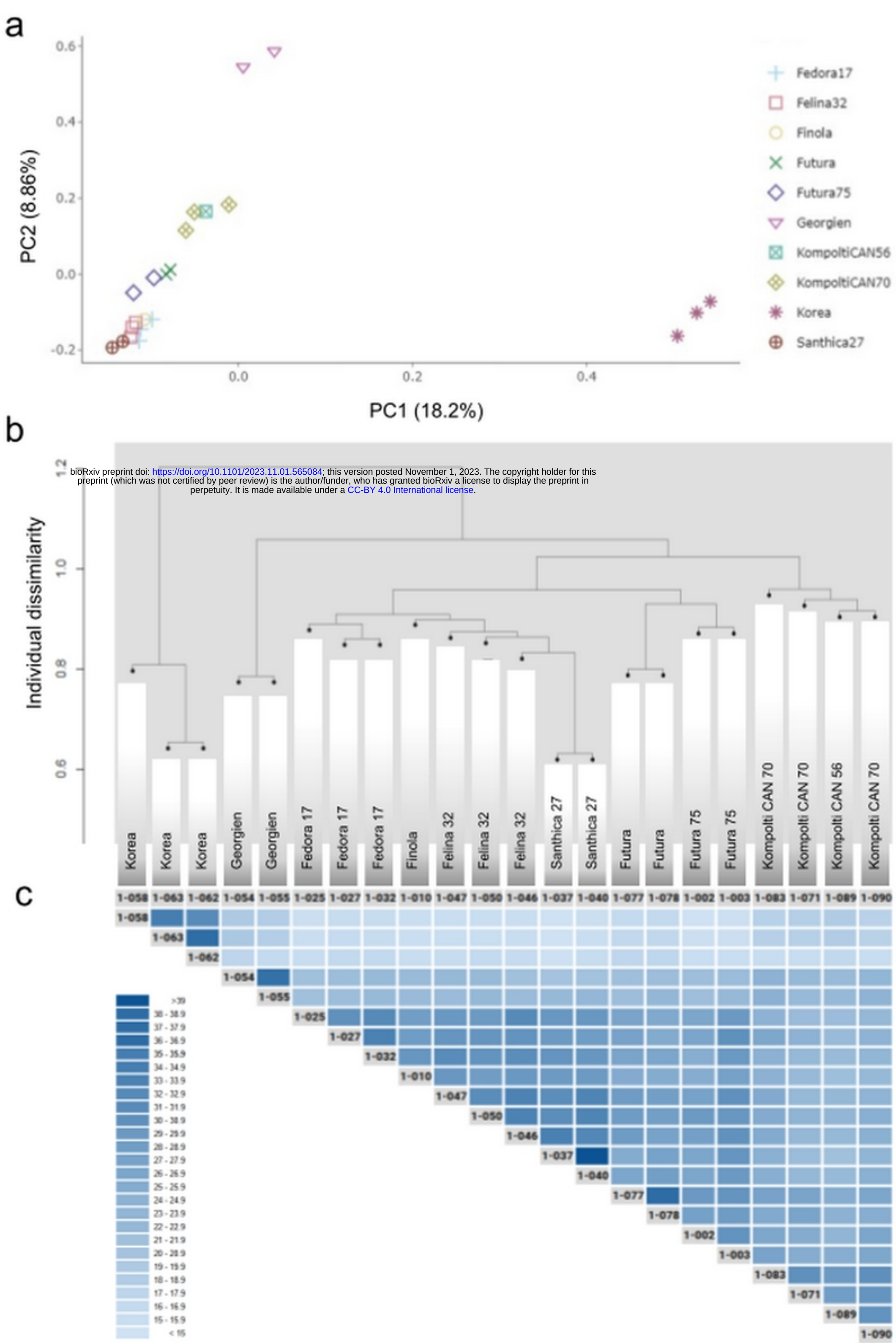
Heterozygosity

**b**

Heterozygosity



Figure



Figure