Haplotype-based phylogenetic analysis uncovers the tetraploid progenitor of sweet potato

Mengxiao Yan
Shanghai Chenshan Botanical Garden
Ming Li
Sichuan Academy of Agricultural Sciences
M-Hossein Moeinzadeh
Max Planck Institute for Molecular Genetics
Dora G. Quispe-Huamanquispe
Ghent University
Weijuan Fan
Shanghai Chenshan Botanical Garden
Haozhen Nie
Shanghai Chenshan Botanical Garden
Zhangying Wang
Guangdong Academy of Agricultural Sciences
Bettina Heider
International Potato Center (CIP)
Robert Jarret
USDA-ARS/PGRU
Jan Kreuze
International Potato Center (CIP)
Godelieve Gheysen
Department of Biotechnology, Ghent University
Hongxia Wang
Shanghai Chenshan Botanical Garden
Ralph Bock
Max Planck Institute of Molecular Plant Physiology https://orcid.org/0000-0001-7502-6940
Martin Vingron
Max Planck Institute for Molecular Genetics
Jun Yang (jyang03@cemps.ac.cn)
Shanghai Chenshan Botanical Garden https://orcid.org/0000-0002-0371-8814

Genetics Article

Keywords: sweet potato, tetraploid progenitor, haplotype-based phylogenetic analysis

Posted Date: August 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-750500/v1

Abstract

The hexaploid sweet potato is one of the most important root crops worldwide. However, its genetic origins, especially that of its tetraploid progenitor, are unclear. In this study, we conceived a pipeline consisting of a genome-wide variation-based phylogeny and a novel haplotype-based phylogenetic analysis (HPA) to determine that the tetraploid accession CIP695141 of *Ipomoea batatas* 4x from Peru is the tetraploid progenitor of sweet potato. We detected biased gene exchanges between subgenomes. The B_1 to B_2 subgenome conversions were almost 3-fold higher than the B_2 to B_1 subgenome conversions. Our analyses revealed that the genes involved in storage root formation, sugar transport, stress resistance, and maintenance of genome stability have been selected during the speciation and domestication of sweet potato. This study sheds lights on the evolution of sweet potato and paves a way for the improvement of sweet potato.

Introduction

Sweet potato, *Ipomoea batatas* (L.) Lam. (2n = 6x = 90), was first domesticated in tropical America at least 5000 years ago [1], introduced into Europe and Africa in the early 16th century, and later into the rest of the world [2]. Today, sweet potato has become an important staple food crop worldwide with an annual production of ~ 113 million tons, and is an important source of dietary calories, proteins, vitamins and minerals. For example, orange sweet potato plays a crucial role in combating vitamin A deficiency in Africa [3–5]. Unlike other important polyploid crops, such as hexaploid bread wheat (*Triticum aestivum*) and tetraploid potato (*Solanum tuberosum*), the origin of cultivated sweet potato has been the subject of considerable debate. Furthermore, the exact role of polyploidization in the origin and evolution of sweet potato has not been determined. Knowledge of its genetic origin is vital for supporting studies of biology, domestication, genetics, genetic engineering and breeding using wild relatives.

Sweet potato belongs to the genus *Ipomoea* series Batatas (Convolvulaceae). This group includes *I. batatas* 6x (sweet potato), *I. batatas* 4x, and 13 diploid species that are commonly considered as the wild relatives of the cultivated sweet potato [6]. Two main polyploidization scenarios have been proposed to account for the hexaploid genome of sweet potato. The autopolyploid hypothesis suggests that sweet potato has an autopolyploid origin with *I. trifida* likely being the only wild ancestor. This hypothesis has gained some support from molecular marker analyses (RAPD and GBS) [7,8], beta-amylase gene sequences [9], sequencing of single-copy nuclear DNA regions [10], and cytogenetic analyses [11–13]. By contrast, allopolyploidy hypotheses are relatively less consistent. Austin [1] suggested that the cultivated sweet potato was derived from a hybridization event between *I. trifida* and *I. triloba* based on morphological data. Gao et al. [14], based on *Waxy* (*Wx*) intron sequence variation, suggested that sweet potato arose via hybridization between *I. tenuissima* and *I. littoralis*. Subsequently, Gao, et al. [15] postulated that the hexaploid sweet potato may be neither a true autopolyploid nor a genuine allopolyploid based on the analysis of 811 conserved single-copy genes. Recently, Yang et al. [5] analyzed the phased genome of sweet potato and proposed that sweet potato might have resulted from a cross between a diploid and a

tetraploid progenitor. The triploid hybrid progeny of this event subsequently underwent a genome duplication resulting in the hexaploid species. This model of the origin of sweet potato is supported by cytogenetic analyses [13,16]. The diploid progenitor of the hybrid was most likely *I. trifida*, whereas the tetraploid progenitor has remained unknown [5].

As the first reported natural transgenic food crop, the genomes of sweet potato and some of its wild relatives contain horizontally transferred *Ib*T-DNA1 and/or *Ib*T-DNA2 sequences from *Agrobacterium* spp. [17,18]. *Ib*T-DNAs could serve as natural genetic markers to track the wild ancestors of cultivated sweet potato [17], and the *Ib*T-DNA positive species in the series Batatas (*I. trifida*, *I. cordatotriloba*, *I. tenuissima*, and *I. batatas* 4x) are potential wild progenitors of sweet potato [17]. Consequently, *I. batatas* 4x and other wild relatives in the *Ipomoea* series Batatas are key species to be examined in order to trace the genetic origin(s) of cultivated sweet potato [17]. As the majority of previous studies have not included *I. batatas* 4x, the relationship between sweet potato and its tetraploid wild relative is unclear, and the potential tetraploid progenitor of sweet potato is currently unknown.

Because of the highly heterozygous and complex hexaploid genome[5,19], a serious limitation in most previous studies on the genetic origin of sweet potato has been the use of consensus genomic sequences and a limited number of nuclear markers. In addition, chromosome rearrangements and homoeologous exchanges that shuffle and/or replace homoeologs among the subgenomes of polyploids [20–22] further complicate genetic studies that aim to resolve the origin of polyploid species. Currently, the best strategy for determining the origin of allopolyploids relies on the use of subgenome-level genome assemblies or the the homologous genes or variants of each subgenome to perform the phylogenetic analyses. This strategy has been successfully applied to rapeseed (*Brassica napus*), bread wheat (*Triticum aestivum*) and *Echinochloa* spp., polyploid bamboo (*Bambusa* spp.) and strawberry (*Fragaria × ananassa*) [23–28]. However, unlike these allopolyploids, the subgenomes of sweet potato are highly similar to one another due to the close genetic relationship between the diploid and the tetraploid progenitor species. Also, a subgenome-level or fully-phased reference genome of sweet potato is not yet available. Therefore, the above-mentioned strategy is not applicable to sweet potato, and a novel method that takes full advantage of genome-wide homologous variation between hexaploid sweet potato and tetraploid wild relatives is required to more fully examine the origin of sweet potato.

To this end, we employed consensus genome-wide variation analysis and developed a haplotype-based phylogenetic analysis pipeline to identify the closest tetraploid accession related to sweet potato. We identified biased gene conversion events between sweet potato subgenomes based on homologous haplotypes. Moreover, we provide new insight in the role that selection played in the domestication process of cultivated sweet potato and identified useful candidate genes for future breeding and genetic engineering efforts, and evolutionary studies. In addition, the identification of the presumptive tetraploid progenitor will accelerate work towards the generation of artificial hexaploids in the genus *Ipomoea*. Taken together, the results of the present study shed light on the evolution of sweet potato and pave the way for the genetic improvement of sweet potato.

Results

Phylogeny of sweet potato and its wild relatives

To investigate the phylogenetic relationship of sweet potato and its wild relatives, we analyzed all putative genetic donors of sweet potato as well as 23 sweet potato cultivars and landraces (Fig. 1). Considering the reported horizontal transfer of *Ib*T-DNAs from *Agrobacterium* spp. into the genomes of sweet potato and some of its wild relatives, *Ib*T-DNA1 and *Ib*T-DNA2 can serve as ideal natural genetic markers to trace the progenitors of the hexaploid cultivated sweet potato [17,18]. Therefore, we sampled the diploid and tetraploid relatives based on *Ib*T-DNA screening (Supplementary Table 1 and Supplementary Figs. 1-4). As diploid relatives, we included three accessions of *I. trifida*, the species that most likely to be the diploid progenitor of sweet potato, and the two wild relatives *I. triloba* and *I.* sp. PI553012. We also sampled ten representative *Ib*T-DNA positive tetraploid wild relatives (ten accessions of *I. batatas* 4x) from all geographical locations according to Quispe-Huamanquispe, et al. [17] and collection records of the International Potato Center (CIP) and the USDA (Supplementary Fig. 1).

Phylogenetic analyses (Fig.1; Supplementary Fig. 5) based on whole genome variations (50,062,627 SNPs) revealed that the diploid wild relatives form the basal clade in the phylogeny of sweet potato and its wild relatives. The ten accessions of *I. batatas* 4x are not monophyletic. Among them, the basal *I. batatas* 4x lineage (accessions CIP403270, CIP695141, CIP695150 and PI518474) resides at the base of a large clade composed of sweet potato cultivars and the monophyletic Ecuador *I. batatas* 4x lineage (accessions PI561246, PI561247, PI561248, PI561255, PI561258 and PI561261). The phylogenetic reconstruction showed that all sweet potato cultivars and the Ecuador *I. batatas* 4x lineage form two independent monophyletic lineages, thus suggesting a sister relationship between these two lineages.

The basal *I. batatas* 4x lineage was nested between the diploid progenitor (*I. trifida*) and the large clade composed of sweet potato cultivars and the Ecuador *I. batatas* 4x lineage. Thus, the basal *I. batatas* 4x lineage likely was the progenitor of both sweet potato cultivars and the Ecuador *I. batatas* 4x lineage (Fig.1; Supplementary Fig. 5). Specifically, the accession CIP695141 had the smallest branch length relative to sweet potato, suggesting that CIP695141 represents the tetraploid accession that is most closely related to sweet potato. It is particularly noteworthy that accession CIP695141 is the only tetraploid accession examined in this study that harbors both the *Ib*T-DNA1 and the *Ib*T-DNA2 insertions (Supplementary Table 1).

The closest tetraploid accession revealed by haplotype-based phylogenetic analysis (HPA)

To identify the closest tetraploid accession and the possible chromosomal donor to sweet potato, we developed a HPA pipeline (Supplementary Fig. 6). First, we independently phased the genome sets of five representative hexaploid sweet potato cultivar and ten tetraploid *I. batatas* 4x accession. As for the representative sweet potato cultivars, we chose one representative cultivar from each lineage in the sweet potato phylogeny (Fig.1; Supplementary Fig. 5), i.e., Huameyano, NASPOT5/58, NK259L, Y601, and Yuzi7. Each cultivar was used to extract the syntenic haplotype block with each *I. batatas* 4x accession. We obtained 439,555-760,769 haplotype blocks in the five sweet potato cultivars (Supplementary Table 2; Supplementary Fig. 7) and 380,895-542,596 haplotype blocks in the ten *I. batatas* 4x accessions (Supplementary Table 3; Supplementary Fig. 7). Second, we extracted the syntenic haplotype blocks shared between each sweet potato cultivar and each *I. batatas* 4x accession by comparing their genomic positions. In doing so, we identified 606,246-1,050,700 syntenic haplotype blocks (Supplementary Table 4; Supplementary Fig. 8). Third, we removed (i) redundant syntenic haplotype blocks that had overlapping regions with other blocks, and (ii) those blocks that consist of very short sequences (less than 3 bp). Ultimately, 406,488-642,341 syntenic haplotype blocks were extracted, which accounted for 33.8-42.7% of the sweet potato genome (Supplementary Table 5; Supplementary Fig. 9).

The previously identified syntenic haplotype blocks between each sweet potato cultivar and each *I. batatas* 4x accession were used to perform phylogenetic reconstructions independently. The phylogenetic trees were inferred by two methods: Unweighted Pair-Group Method with Arithmetic Mean (UPGMA) and maximum likelihood (ML). We calculated the monophyletic ratio and the Nsp-Nwr distance to measure the relationship between the investigated tetraploid accession and the hexaploid sweet potato (Supplementary Fig. 6d). The monophyletic ratio is defined as the proportion of trees in which sweet potato haplotypes forming a monophyletic clade (Supplementary Fig. 6d). The Nsp-Nwr distance is defined as the tree branch length between the most recent common ancestor (MCRA) node of sweet potato haplotypes (i.e., Nsp) and the MCRA node of the tetraploid accession (i.e., Nwr) (Supplementary Fig. 6d). Smaller indices indicate a closer relationship between the investigated tetraploid accession and the hexaploid sweet potato. To increase accuracy, we only included trees that had the same monophyletic judgement by both tree-building methods, and these trees were used to calculate the monophyletic ratio and Nsp-Nwr distance. Among all syntenic haplotype blocks, the 6:4 data set (composed of six haplotypes of sweet potato and four haplotypes of *I. batatas* 4x) produced the most robust results based on the two indices (Fig. 2; Supplementary Fig. 10-24).

The monophyletic ratio and the Nsp-Nwr distance both indicated that the accession CIP695141 was the closest tetraploid relative of the hexaploid sweet potato, regardless of which sweet potato cultivar was used as the hexaploid reference (Fig. 2). In addition, the accessions CIP403270, CIP695141, CIP695150, and PI518474, which belong to the basal *I. batatas* 4x lineage, showed closer relationships with sweet potato than any accession in the Ecuador *I. batatas* 4x lineage (Fig. 2). This result confirms the close

relationship between the basal *I. batatas* 4x lineage and sweet potato as revealed by the SNP-based phylogenetic analysis, and demonstrates that HPA is suitable to distinguish the closest tetraploid accessions of sweet potato.

Gene conversion between sweet potato subgenomes

Gene conversion in polyploids refers to sequence exchanges between homologous genes from different subgenomes, in which one progenitor allele overwites another [29-31]. The sweet potato genome is comprised of two B_1 and four B_2 subgenomes ($B_1B_1B_2B_2B_2B_2$). Subgenomes B_1B_1 were donated by the diploid progenitor and subgenomes $B_2B_2B_2B_2$ by the tetraploid progenitor [13] (Fig. 3a). If no conversion events occurred, each syntenic haplotype block should have two copies of the B_1 subgenome from sweet potato, four copies of the B_2 subgenome from sweet potato, and four copies of the B_2 subgenome from *I. batatas* 4x (Fig. 3a, c). If a gene were converted between B_1 and B_2 subgenomes, the copy numbers of subgenomes and tree topology should deviate from the standard 2:8 ratio between B_1 and B_2 in the hexaploid sweet potato and *I. batatas* 4x (Fig. 3c-e). To detect possible gene conversion events, we first filtered those syntenic haplotype blocks and use only blocks in gene regions with six haplotypes of sweet potato and four haplotypes of *I. batatas* 4x within. We then used 786-1,634 homogeneous haplotype blocks in gene regions of sweet potato and the closest *I. batatas* 4x accession (CIP695141) to identify gene conversion events between subgenomes (Supplementary Table 6). Using different sweet potato cultivars as references, 49.1-53.4% of gene regions in sweet potato showed evidence of conversion between subgenomes (Fig. 3b; Supplementary Table 6). We found that B_1 to B_2 subgenome gene conversions (41.1%-43.8%) were much more common than B_2 to B_1 conversions (8.0-9.7%) (Fig. 3c; Supplementary Table 6). This was to be expected, as gene conversion is known to be a copy number-dependent process [32].

Genomic signatures of selective sweeps in sweet potato

A genetic diversity comparison between sweet potato and its tetraploid wild relatives was performed by estimating the genome-wide nucleotide diversity (π) of 23 sweet potato cultivars and landraces, and ten *I. batatas* 4x accessions. The subsequent evaluation showed that sweet potato and *I. batatas* 4x have very similar genome-wide nucleotide diversities (Supplementary Fig. 26; $\pi_{\text{sweet potato}}$ =0.0227, $\pi_{\textit{I. batatas}\ 4x}$ =0.0231).

To detect potential signatures of selection during sweet potato domestication, we employed three metrics, π ratio (π wild relative/π sweet potato), Fst and XP-CLR, to identify potential selective sweeps associated with natural selection and domestication. Using a 100 kb sliding window with 10 kb steps, a total of 466

potential selective sweeps in the top 1% of π ratio, Fst and XP-CLR scores were detected. These regions contained 1559, 1438 and 8814 genes, respectively (Supplementary Fig. 27). Many of these genes are associated with root initiation and development, cell wall organization, phytohormone biogenesis and response, sugar transport, starch and sucrose metabolism, and plant defense (Supplementary Table 7). We highlighted the 20 genes supported by at least two metrics in Manhattan plots (Fig.4; Supplementary Table 8). Among them, NAC domain-containing protein 100 (NAC100), Agamous-like MADS-box protein AGL14, homeobox protein knotted-1-like (KNOX1), ethylene-responsive transcription factor RAP2-13-like (RAP2-13), and fimbrin-like protein 2 (FIM2) have been reported to be involved in storage root initiation and/or development [33-38]. Root hair defective 3 (RHD3), topless-related protein 3-like (TPR3), FAR1-related sequence 5-like (FRS5), and wall-associated kinase (WAK) are functionally related to root development in *Arabidopsis* [39-41]. The nine genes may play important roles in storage root development in sweet potato. Pectin is important for cell wall properties and storage root development [42,43]. Two genes involved in pectin biogenesis and acetylation were identified, arabinosyltransferase 1 (*ARAD1*) and pectin acetylesterase 8-like (*PAE8*) [44,45]. In addition, the sugar transporter SWEET1 known to mediate both low-affinity uptake and efflux of sugar across the plasma membrane [46], three well-known plant defense genes (i.e., *Xa21* encoding receptor kinase-like protein, *DDS* encoding dammarenediol II synthase-like, and *N* encoding TMV resistance protein N-like) were identified [47-49]. Sporamin B, a major storage protein in sweet potato storage roots, which plays additional roles in defense and development [50] was also detected. In polyploids, maintenance of genomic stability poses particular challenges due to the complex meiotic behavior of the chromosome sets and recombination [51,52]. Two genes required for maintenance of genomic stability were identified, spindle and kinetochore-associated protein 1 (*SKA1*), which is essential for proper chromosome segregation [53]; and RECQ helicase L2 (*RECQL2*), which prevents recombination events and channels repair processes into non-recombinogenic pathways [54].

Discussion

Understanding the genetic origin of crops is vital for breeding and genetic engineering efforts, and is particularly important to all genetic improvement strategies involving wild relatives. The origin of sweet potato is still the subject of fierce debate. Competing hypotheses have been put forward proposing that sweet potato is an autopolyploid, an allopolyploid, or neither a pure autopolyploid nor a true allopolyploid [8–10, 12,14–16, 19,55,56]. The genetic origin of sweet potato has remained unresolved because of the high complexity of the genome, due to its hexaploid nature and high degree of heterozygosity [5,19]. In addition, the two progenitors of sweet potato are genetically closely related, thus adding to the difficulties with distinguishing the subgenomes of sweet potato. The phased genome sequence of sweet potato distinguished the two sets of chromosomes contributed by the diploid progenitor from the four chromosome sets coming from the tetraploid progenitor [5], and confirmed the $B_1B_1B_2B_2B_2B_2$ genome architecture that has been revealed by earlier cytogenetic studies [13,16]. The diploid progenitor of sweet potato is most likely *I. trifida*, but the tetraploid progenitor has remained unknown. Although the extant tetraploid relatives exhibit a closer relationship with sweet potato than *I. trifida* [9,55–57], they have not

received sufficient attention in previous studies that aimed to resolve the genetic origin of sweet potato. As the tetraploid progenitor contributed two thirds of the chromosomes to the sweet potato genome, it potentially plays more important roles in physiology and genetic behavior of sweet potato than the diploid ancestor.

In order to identify the tetraploid progenitor of sweet potato, we investigated representative *Ib*T-DNA positive tetraploid wild relatives from all relevant geographic locations (based on the previous study by Quispe-Huamanquispe, et al. [17]) as well as collection records of the USDA and CIP (Supplementary Fig. 1). The phylogeny based on genome-wide variation revealed that the tetraploid relatives (*I. batatas* 4x) have a closer relationship with sweet potato than the diploid progenitor (*I. trifida*) and showed that the closest tetraploid accession to sweet potato among those included in this study is CIP695141. It is important to note that all polyploid accessions were treated as diploids during the variant calling step to meet the data format required by downstream analysis. This procedure artificially decreased the nucleotide diversity of polyploids and unavoidably, resulted in uncertainty in the conclusions that could be drawn. This represents a common problem in studies on the origin of polyploid species that rely on consensus variation. To solve this problem, we developed a HPA pipeline that takes full advantage of homologous variation while maintaining the true nucleotide diversity of the polyploid species. Using this new pipeline, we successfully identified the closest tetraploid accession to sweet potato and authenticated the result of the consensus genome-wide variation analysis. Encouraged by the success with sweet potato, we are currently employing the HPA pipeline to resolve the genetic origin in other complex polyploid species.

The majority of sweet potato cultivars and landraces have both the *Ib*T-DNA1 and the *Ib*T-DNA2 insertions in their genomes (Supplementary Table 1). The *Ib*T-DNA2 of sweet potato could be inherited from the diploid progenitor or the tetraploid progenitor or both, while the *Ib*T-DNA1 is most likely inherited from the tetraploid progenitor, because *I. trifida* does not have the *Ib*T-DNA1 insertion as revealed by PCR screening of 29 *I. trifida* accessions [17]. The accession CIP695141 is the only tetraploid accession that has both *Ib*T-DNA1 and *Ib*T-DNA2, in line with its close relationship with sweet potato. Based on our analysis of all available tetraploid accessions collected (to date) in South America, we conclude that the accession CIP695141 from Peru is the tetraploid progenitor of hexaploid sweet potato, or at least shares the most recent tetraploid progenitor with hexaploid sweet potato. However, there is still the remote possibility that the true tetraploid progenitor of sweet potato has not been discovered yet. This possibility should be addressed in future research by conducting a wider survey and full examination of tetraploid species from Peru. Sweet potato breeders have been working for decades towards generating artificial hexaploids from diploid and tetraploid wild relatives of sweet potato [58]. The discovery of CIP695141 as the closest tetraploid accession to sweet potato not only contributes to our understanding of the genetics of sweet potato, but also provides a critical natural resource for future breeding programs.

Another important application of HPA lies in the use of homologous haplotypes to detect gene conversion between subgenomes. In sweet potato, almost half of the gene regions show evidence of conversion between subgenomes. Taking advantages of phased haplotypes, the identified gene conversion events

between subgenomes also shed light on the evolution and domestication of hexaploid sweet potato. B_1 to B_2 conversion events are approximately 3-times more frequent than B_2 to B_1 conversions (Fig. 3). Rampant gene conversion and conversion biases increase genome complexity in sweet potato and may suggest an important role for gene conversion in genome evolution and domestication of sweet potato. Subgenome-biased conversion has been reported in several allopolyploid crop plants including cotton, canola, peanut and strawberry [28,30,59,60]. However, the molecular mechanisms underlying the conversion bias are largely unknown. In the case of sweet potato, the dosage effect (of the tetraploid B_2 versus the diploid B_1 genome) may explain the more prevalent conversion of B_1 alleles to B_2 alleles. Because gene conversion is known to be a copy number-dependent process [32].

The domestication syndrome of vegetatively propagated field crops includes the mode of reproduction, the yield of the edible parts of the plant, the time and ease of harvesting, defense adaptations and plant architecture [61]. At least some of these phenotypic traits are likely associated with domestication in sweet potato. Compared with its wild relatives, the most remarkable feature of sweet potato is its large edible storage roots. The evolutionary history of the sweet potato storage root has not been elucidated yet. A storage root is found in some accessions of the diploid progenitor, *I. trifida* [62] (Supplementary Fig. 28; Supplementary Table 1). A substantial enlargement of the fibrous root in several *I. batatas* 4x accessions has also been reported [63], although the *I. batatas* 4x accessions included in the present study do not produce tuberous roots (Supplementary Fig. 28–30; Supplementary Table 1). The original root traits in primitive sweet potatoes are unknown, because the speciation of sweet potato occurred during pre-human times [10]. It is noteworthy in this regard that the wild hexaploid sweet potato (accession Y601) produces a slightly enlarged storage root (Supplementary Fig. 28; Supplementary Table 1). All available data suggest that both the progenitor species and the wild hexaploid sweet potato have the potential to form storage roots, but the typical edible storage root of modern cultivars are the result of domestication and were selected by early hunter-gatherers, farmers and breeders. Therefore, formation of a starchy storage root has likely been under strong selection pressure during the domestication of sweet potato by humans [64], which is also supported by genomic signatures. The majority of selective sweeps are functionally related to root development. Five genes (*KNOX1*, *AGL14*, *RAP2-13*, *FIM2* and *NAC100*) that are required for storage root initiation and development in sweet potato and/or cassava [33–38], were identified in this study. Four additional candidate genes (*RHD3*, *TPR3*, *FRS5* and *WAK*) are known to be involved in root development in *Arabidopsis* [39–41]. These nine genes represent attractive targets of functional validation experiments and the future genetic improvement of tuber crops.

Starch accumulation is another agronomically important trait that was under selection during domestication of sweet potato [64]. Sugars are important substrates for starch biosynthesis. Hence, sugar transport is critical for both source-sink relations and starch accumulation in the storage root. In this study, we have identified the sugar transporter SWEET1 as a target of selection during sweet potato domestication. SWEET1 likely acts as a bidirectional glucose transporter in sweet potato. In addition, biotic resistance is also a trait strongly selected during both human and natural selection [64]. The whole

sweet potato plant is edible, and both leaves and storage roots attract herbivorous insects, and pathogenic viruses, fungi and bacteria [65–68]. Therefore, plant defense is important to both plant survival and storage root production in sweet potato. Four well-known plant resistance genes (*Xa21*, *DDS*, *Sporamin B*, and *N-like*) were identified as genes carrying signatures of selective sweeps.

As information on domestication-related genes and their genomic and subgenomic distribution continues to accumulate, new opportunities become available to improve the crop by increasing yields and development of tailor-made varieties. Clearly, a combination of applied and theoretical approaches (involving computational and systems biology-based models) will be required to meet the challenges involved [69]. The new knowledge on sweet potato genomics and domestication revealed in this study will contribute to this goal and aid future breeding and genetic engineering approaches in this important staple crop.

Methods

Plant materials. Five diploid wild relatives of sweet potato (including three accessions of *I. trifida*, one accession of *I. triloba* and one accession of *I.* sp), ten *Ib*T-DNA positive tetraploid wild relatives of sweet potato (*I. batatas* 4x) and 23 sweet potato cultivars/landraces were utilized. Among these, sequencing data from cultivars Taizhong6, Xushu18, Y601, Yuzi263 and Yuzi7 were newly generated in this study. All other data was downloaded from NCBI, including cultivars Tanzania, Beauregard and 16 cultivars in the Mwanga diversity panel (MDP) [19]. Detailed information on the plant materials is given in Supplementary Table 1 and Supplementary Fig. 1-4.

***Ib*T-DNA detection**. The PCR detection of *Ib*T-DNA1 and *Ib*T-DNA2 genes was performed as previously described in Quispe-Huamanquispe, et al. [17]. To estimate the coverage depth of *Ib*T-DNA insertion, the WGS paired-end reads were aligned to *Ib*T-DNA1 and *Ib*T-DNA2 reference sequence (GenBank: KM052616 and KM052617) using bwa-mem (version 0.7.17) [70] and the coverage depth were calculated with samtools (version 1.10) [71]. The depth ratio was defined as the ratio between *Ib*T-DNA depth (Dt) and genome depth (Dg), to estimate the relative depth of *Ib*T-DNA region compared with the normal genome region. For simplification, the average depth of F-box gene was used as the genome depth.

$$\text{Depth ratio}=\frac{Dt}{Dg}$$

Variant calling. The WGS paired-end reads were aligned to the reference sweet potato genome (https://ipomoea-genome.org/download_genome.html) using bwa-mem (version 0.7.17) [70] and sorted by samtools (version 1.10) [71] with the default parameters. Picard (version 2.23.4) [72] was used to label PCR duplicates based on the mapping coordinates. Genetic variants including SNPs and INDELs were detected as diploid using the Genome Analysis Toolkit (GATK, version 4.1.8.1) [73]. SNPs were filtered using bcftools (version 1.11, http://samtools.github.io/bcftools) with the following parameters: DP>3 && QUAL>=30 && F_MISSING<=0.8.

Phylogenetic analysis. VCF-kit (version 0.2.8, https://github.com/AndersenLab/VCF-kit) was used to generate a fasta file by concatenating all SNPs from the VCF file. A phylogenetic tree of sweet potato cultivars/landraces and wild relatives was reconstructed using IQ-TREE (version 1.6.12)[74] with 1,000 ultrafast bootstrap replicates. The nucleotide substitution model (GTR+F+I+G4) was selected by IQ-TREE. The phylogenetic tree was rooted with the diploid wild relatives as outgroup and all accessions were plotted onto world map using the R package phytools (version 0.7-70) [75].

Haplotype-based phylogenetic analysis. We developed the HPA pipeline to investigate the relationship of each tetraploid accession to cultivated sweet potato (Supplementary Fig. 6).

Haplotyping. The WGS paired-end reads from the sweet potato cultivars and *I. batatas* 4x were mapped to the sweet potato reference genome using bwa-mem (version 0.7.17-r1188). Freebayes (version v1.3.1-17-gaa2ace8) [76] was used to call variants (setting -p 6 for sweet potato and –p 4 for *I. batatas* 4x). Ranbow (version 2.0) [77] was used for genome haplotyping.

Phylogenetic analysis. The syntenic haplotype blocks between each sweet potato cultivar and each tetraploid accession were extracted and filtered using HPA pipeline. Sequences within each syntenic haplotype block were aligned by MAFFT[78]. The UPGMA tree and ML tree for each syntenic haplotype block were reconstructed independently using MEGA-CC (version 10.1.8) [79,80] and IQ-TREE respectively. The monophyletic ratio and Nsp-Nwr distance were calculated using HPA pipeline. To increase the accuracy, only those trees which had the same monophyletic judgement by two tree-building methods (trees generated based on the same syntenic block by two methods are both monophyletic or both not monophyletic) were used to calculate monophyletic ratio and Nsp-Nwr distance. The detailed identification procedures are descripted in the Supplementary Note.

Gene conversion. The syntenic haplotype blocks that had six haplotypes of sweet potato and four haplotypes of *I. batatas* 4x, within gene regions, were extracted to detect gene conversion between subgenomes. When ignoring the reverse gene conversion, if there is no gene conversion in a specific syntenic haplotype block, the block is expected to have two B_1 subgenome haplotypes and four B_2 subgenome haplotypes from sweet potato, and four B_2 subgenome haplotypes from *I. batatas* 4x. The phylogenetic tree should form two clades corresponding to haplotypes of each subgenome. If a gene was converted between subgenomes, the number of haplotypes and the tree topology is expected to vary. Gene conversions were identified based on tree topology (Supplementary Fig. 25). The detailed identification procedures are descripted in the Supplementary Note. The detailed procedures for determining gene conversion are provided in the Supplementary Note.

Population genetic diversity and selective sweep detection. Nucleotide diversity (θπ) and Population differentiation (F_{ST}) were determined for the tetraploid wild relative population (ten accessions of *I. batatas* 4x) and the sweet potato population (23 cultivars/landraces) with VCFtools (version 0.1.17) [81] using a 100 kb sliding window with a 10 kb step size. A composite likelihood approach (XP-CLR) was applied to scan for genome-wide selective sweeps [82] and a python version was employed (https://github.com/hardingnj/xpclr), with a 100 kb sliding window and a 10 kb step size. We considered the tetraploid wild relative (*I. batatas* 4x) population as a reference population and the sweet potato population a query population to identify the potential evolution/breeding sweeps. To detect selective sweeps, we calculated the π ratio ($\pi_{\text{wild relatives}}/\pi_{\text{sweet potato}}$) within the same sliding windows. The regions that scored in the top 1% of the π ratio, F_{ST} or XP-CLR values were defined as candidate domestication sweeps.

Declarations

Data availability

The raw DNA sequencing data are deposited in BIGD under accession number PRJCA004953.

Code availability

The HPA pipeline and relevant instructions are available at the Github website (https://github.com/YanMengxiao/HPA). Other analysis command lines are given in the Supplementary Data file.

Acknowledgements

This work was funded by the Ministry of Science and Technology of China (2019YFD1000703 to J.Y., 2019YFD1000704-2 to M.X.Y., and 2019YFD1000701-2 to W.J.F.), the Innovation Promotion Foundation of Sichuan Academy of Agricultural Sciences (2019LJRC017, 2019QYXK006 to M.L.), the National Natural Science Foundation of China (31771854 to H.X.W.), the Shanghai Municipal Afforestation & City Appearance and Environmental Sanitation Administration (G182402, G192413, G192414 and G202402 to J.Y.), the State Key Laboratory of Subtropical Silviculture (KF2019 to J.Y.), Youth Innovation Promotion Association CAS, and the Bureau of Science and Technology for Development CAS (KFJ-BRP-017-42 to J.Y.).

Author contributions

M.X.Y. developed HPA pipeline, conducted data analysis and wrote the manuscript. M.L. provided plant materials, part of sequencing data. M.H.M. updated the Ranbow software. D.G.Q. performed the PCR screening of most plants. W.J.F. and H.Z.N. performed the PCR screening of DNA samples. Z.Y.W. helped to access the plants and DNA samples. B.H. tracked back the collecting information of CIP accessions. R.J. provided the Ecuador accessions. J.F.K. and G.G. discussed and contributed to the early phase of the project. H.X.W. and R.B. revised the manuscript. R.B. suggested and contributed to the gene conversion analysis. M.V. discussed and contributed to the haplotyping analysis. J.Y. designed the study and contributed to the original concept of the project.

Competing interests

The authors declare no competing financial interests.

References

1. Austin, D.F. The taxonomy, evolution and genetic diversity of sweet potatoes and related wild species. in *Exploration, Maintenance, and Utilization of Sweetpotato Genetic Resources* 27–60 (CIP, 1988).
2. Roullier, C., Benoit, L., McKey, D.B. & Lebot, V. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proceedings of the National Academy of Sciences* **110**, 2205–2210 (2013).
3. Food and Agriculture Organization. FAOSTAT Statistics Database. http://www.fao.org/faostat/. (2019).
4. Kurabachew, H. The role of orange fleshed sweet potato *(Ipomea batatas*) for combating vitamin A deficiency in Ethiopia: A review. *International Journal of Food Sciences and Nutrition Engineer* **5**, 141–146 (2015).

5. Yang, J. *et al.* Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nature Plants* **3**, 696–703 (2017).
6. Huaman, Z. *Systematic Botany and Morphology of the Sweetpotato Plant*, (International Potato Center, Lima, Peru, 1992).
7. Mollinari, M. *et al.* Unraveling the hexaploid sweetpotato inheritance using ultra-dense multilocus mapping. *bioRxiv*, 689638 (2019).
8. Ukoskit, K. & Thompson, P.G. Autopolyploidy versus allopolyploidy and low-density randomly amplified polymorphic DNA linkage maps of sweetpotato. *Journal of the American Society for Horticultural Science* **122**, 822–828 (1997).
9. Rajapakse, S. *et al.* Phylogenetic relationships of the sweetpotato in *Ipomoea* series Batatas (Convolvulaceae) based on nuclear β-amylase gene sequences. *Molecular Phylogenetics and Evolution* **30**, 623–632 (2004).
10. Muñoz-Rodríguez, P. *et al.* Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to Polynesia. *Current Biology* **28**, 1246–1256 (2018).
11. Nishiyama, I. Evolution and domestication of the sweet potato. *Botanical Magazine* **84**, 377–387 (1971).
12. Srisuwan, S., Sihachakr, D. & Siljak-Yakovlev, S. The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Plant Science* **171**, 424–433 (2006).
13. Shiotani, I. & Kawase, T. Synthetic hexaploids derived from wild species related to sweet potato. *Japanese Journal of Breeding* **37**, 367–376 (1987).
14. Gao, M. *et al. Wx* intron variations support an allohexaploid origin of the sweetpotato [*Ipomoea batatas* (L.) Lam]. *Euphytica* **177**, 111–133 (2011).
15. Gao, M., Soriano, S.F., Cao, Q., Yang, X. & Lu, G. Hexaploid sweetpotato (*Ipomoea batatas* (L.) Lam.) may not be a true type to either auto-or allopolyploid. *PloS one* **15**, e0229624 (2020).
16. Magoon, M., Krishnan, R. & Bai, K.V. Cytological evidence on the origin of sweet potato. *Theoretical and Applied Genetics* **40**, 360–366 (1970).
17. Quispe-Huamanquispe, D.G. *et al.* The horizontal gene transfer of *Agrobacterium* T-DNAs into the series Batatas (Genus *Ipomoea*) genome is not confined to hexaploid sweetpotato. *Scientific Reports* **9**, 1–13 (2019).
18. Kyndt, T. *et al.* The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: an example of a naturally transgenic food crop. *Proceedings of the National Academy of Sciences* **112**, 5844–5849 (2015).
19. Wu, S. *et al.* Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nature Communications* **9**, 1–12 (2018).
20. Bertioli, D.J. *et al.* The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature Genetics* **51**, 877–884 (2019).

21. Gaeta, R.T. & Chris Pires, J. Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytologist* **186**, 18–28 (2010).
22. Wang, M. *et al.* Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nature Genetics* **51**, 224–229 (2019).
23. Lu, K. *et al.* Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nature Communications* **10**, 1154 (2019).
24. An, H. *et al.* Transcriptome and organellar sequencing highlights the complex origin and diversification of allotetraploid *Brassica napus*. *Nature Communications* **10**, 2878 (2019).
25. Zhou, Y. *et al. Triticum* population sequencing provides insights into wheat adaptation. *Nature Genetics* **52**, 1412–1422 (2020).
26. Ye, C.Y. *et al.* The genomes of the allohexaploid *Echinochloa crus-galli* and its progenitors provide insights into polyploidization-driven adaptation. *Molecular Plant* **13**, 1298–1310 (2020).
27. Guo, Z.H. *et al.* Genome sequences provide insights into the reticulate origin and unique traits of woody bamboos. *Molecular Plant* **12**, 1353–1365 (2019).
28. Edger, P.P. *et al.* Origin and evolution of the octoploid strawberry genome. *Nature Genetics* **51**, 541–547 (2019).
29. Wang, X.Y. & Paterson, A.H. Gene conversion in angiosperm genomes with an emphasis on genes duplicated by polyploidization. *Genes* **2**, 1–20 (2011).
30. Chen, X. *et al.* Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proceedings of the National Academy of Sciences* **113**, 6785–6790 (2016).
31. Cenci, A., Combes, M.-C. & Lashermes, P. Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. *Plant Molecular Biology* **78**, 135–145 (2012).
32. Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal* **46**, 85–94 (2006).
33. Sojikul, P. *et al.* Genome-wide analysis reveals phytohormone action during cassava storage root initiation. *Plant Molecular Biology* **88**, 531–543 (2015).
34. Tanaka, M., Kato, N., Nakayama, H., Nakatani, M. & Takahata, Y. Expression of class I *knotted1*-like homeobox genes in the storage roots of sweetpotato (*Ipomoea batatas*). *Journal of Plant Physiology* **165**, 1726–1735 (2008).
35. Firon, N. *et al.* Transcriptional profiling of sweetpotato (*Ipomoea batatas*) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *Annals of Botany* **14**, 1–25 (2013).
36. Ku, A.T., Huang, Y.S., Wang, Y.S., Ma, D.F. & Yeh, K.W. *IbMADS1* (*Ipomoea batatas MADS-box 1* gene) is involved in tuberous root initiation in sweet potato (*Ipomoea batatas*). *Annals of Botany* **102**, 57–67 (2008).

37. Garay-Arroyo, A. *et al.* The MADS transcription factor XAL2/AGL14 modulates auxin transport during Arabidopsis root development by regulating PIN expression. *The EMBO Journal* **32**, 2884–2895 (2013).
38. Tanabe, N., Ito, A., Tamoi, M. & Shigeoka, S. Identification of promoter for adventitious root-specific gene expression from sweet potato. *Plant Root* **12**, 31–44 (2018).
39. Schiefelbein, J.W. & Somerville, C. Genetic control of root hair development in *Arabidopsis thaliana*. *The Plant Cell* **2**, 235–243 (1990).
40. Pi, L. *et al.* Organizer-derived WOX5 signal maintains root columella stem cells through chromatin-mediated repression of *CDF4* expression. *Developmental Cell* **33**, 576–588 (2015).
41. Ma, L. & Li, G. FAR1-related sequence (FRS) and FRS-related factor (FRF) family proteins in *Arabidopsis* growth and development. *Frontiers in Plant Science* **9**, 692 (2018).
42. Guillemin, F. *et al.* Distribution of pectic epitopes in cell walls of the sugar beet root. *Planta* **222**, 355–371 (2005).
43. Dong, W. *et al.* Changes in cell wall components and polysaccharide-degrading enzymes in relation to differences in texture during sweetpotato storage root growth. *Journal of Plant Physiology* **254**, 153282 (2020).
44. Harholt, J. *et al.* ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in *Arabidopsis*. *Plant Physiology* **140**, 49–58 (2006).
45. de Souza, A., Hull, P.A., Gille, S. & Pauly, M. Identification and functional characterization of the distinct plant pectin esterases PAE8 and PAE9 and their deletion mutants. *Planta* **240**, 1123–1138 (2014).
46. Chen, L.Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532 (2010).
47. Park, C.-J. & Ronald, P.C. Cleavage and nuclear localization of the rice XA21 immune receptor. *Nature Communications* **3**, 1–6 (2012).
48. Li, J. *et al.* Fungal elicitors enhance ginsenosides biosynthesis, expression of functional genes as well as signal molecules accumulation in adventitious roots of *Panax ginseng* CA Mey. *Journal of Biotechnology* **239**, 106–114 (2016).
49. Marathe, R., Anandalakshmi, R., Liu, Y. & Dinesh-Kumar, S. The tobacco mosaic virus resistance gene, N. *Molecular Plant Pathology* **3**, 167–172 (2002).
50. Senthilkumar, R. & Yeh, K.W. Multiple biological functions of sporamin related to stress tolerance in sweet potato (*Ipomoea batatas* Lam). *Biotechnology Advances* **30**, 1309–1317 (2012).
51. Comai, L. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**, 836–846 (2005).
52. Hollister, J.D. Polyploidy: adaptation to the genomic environment. *New Phytologist* **205**, 1034–1039 (2015).

53. Hanisch, A., Silljé, H.H. & Nigg, E.A. Timely anaphase onset requires a novel spindle and kinetochore complex comprising Ska1 and Ska2. *The EMBO Journal* **25**, 5504–5515 (2006).
54. Kobbe, D., Blanck, S., Demand, K., Focke, M. & Puchta, H. AtRECQ2, a RecQ helicase homologue from *Arabidopsis thaliana*, is able to disrupt various recombinogenic DNA structures *in vitro*. *The Plant Journal* **55**, 397–405 (2008).
55. Jarret, R. & Austin, D. Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. *Genetic Resources and Crop Evolution* **41**, 165–173 (1994).
56. Jarret, R., Gawel, N. & Whittemore, A. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. *Journal of the American Society for Horticultural Science* **117**, 633–637 (1992).
57. Feng, J.Y. *et al.* Analysis of evolution and genetic diversity of sweetpotato and its related different polyploidy wild species *I. trifida* using RAD-sEq. *BMC Plant Biology* **18**, 1–12 (2018).
58. Nishiyama, I., Miyazaki, T. & Sakamoto, S. Evolutionary autoploidy in the sweet potato (*Ipomoea batatas* (L.) Lam.) and its progenitors. *Euphytica* **24**, 197–208 (1975).
59. Chalhoub, B. *et al.* Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953 (2014).
60. Paterson, A.H. *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427 (2012).
61. Denham, T. *et al.* The domestication syndrome in vegetatively propagated field crops. *Annals of Botany* **125**, 581–597 (2020).
62. Li, M. *et al.* The wild sweetpotato (*Ipomoea trifida*) genome provides insights into storage root development. *BMC Plant Biology* **19**, 119 (2019).
63. Díaz, J., De La Puente, F. & Austin, D.F. Enlargement of fibrous roots in *Ipomoea* section Batatas (Convolvulaceae). *Economic Botany* **46**, 322–329 (1992).
64. Liu, Q. Improvement for agronomically important traits by gene engineering in sweetpotato. *Breeding Science*, 16126 (2017).
65. Jansson, R.K. & Raman, K.V. *Sweet Potato Pest Management: a Global Perspective*, (International Potato Center, 1991).
66. Ogawa, K. & Komada, H. Biological Control of Fusarium Wilt of Sweet Potato by Non-pathogenic *Fusarium oxysorum*. *Japanese Journal of Phytopathology* **50**, 1–9 (1984).
67. Jang, I.C. *et al.* Differential expression of 10 sweetpotato peroxidase genes in response to bacterial pathogen, *Pectobacterium chrysanthemi*. *Plant Physiology and Biochemistry* **42**, 451–455 (2004).
68. Gutiérrez, D.L., Fuentes, S. & Salazar, L. Sweetpotato virus disease (SPVD): distribution, incidence, and effect on sweetpotato yield in Peru. *Plant Disease* **87**, 297–302 (2003).
69. Vaughan, D., Balazs, E. & Heslop-Harrison, J. From crop domestication to super-domestication. *Annals of Botany* **100**, 893–901 (2007).

70. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
71. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
72. Picard toolkit. in *Broad Institute, GitHub repository* (Broad Institute, 2019).
73. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178 (2017).
74. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2014).
75. Revell, L.J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).
76. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv*, 1207.3907 (2012).
77. Moeinzadeh, M.-H. *et al.* Ranbow: A fast and accurate method for polyploid haplotype reconstruction. *PLoS Computational Biology* **16**, e1007843 (2020).
78. Katoh, K. & Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
79. Kumar, S., Stecher, G., Peterson, D. & Tamura, K. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* **28**, 2685–2686 (2012).
80. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–1549 (2018).
81. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
82. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Research* **20**, 393–402 (2010).

Figures

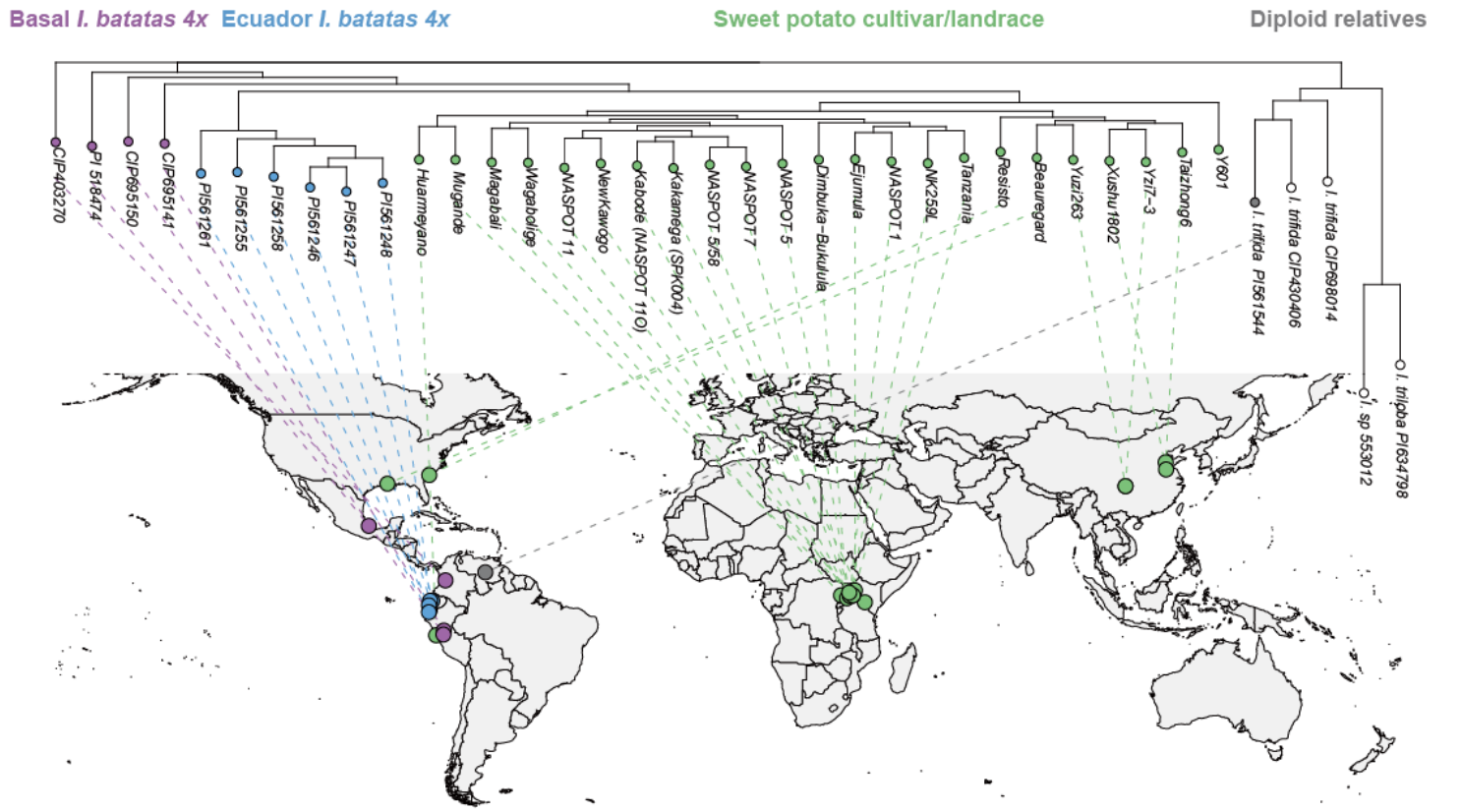

Figure 1

Phylogeny and geographic distribution of the sampled sweet potato cultivars and wild relatives. Phylogenetic relationships between sweet potato cultivars and landraces, and their diploid and tetraploid relatives were inferred using the maximum likelihood method. The colored dashed lines link the phylogenetic position on the tree with the geographic location on the map for each accession. The purple, blue and green dots denote the basal I. batatas 4x, the Ecuador I. batatas 4x, and the sweet potato cultivar/landrace lineages, respectively. The grey and white dots mark the diploid relatives. The accessions with unknown geographic location are not linked to map.

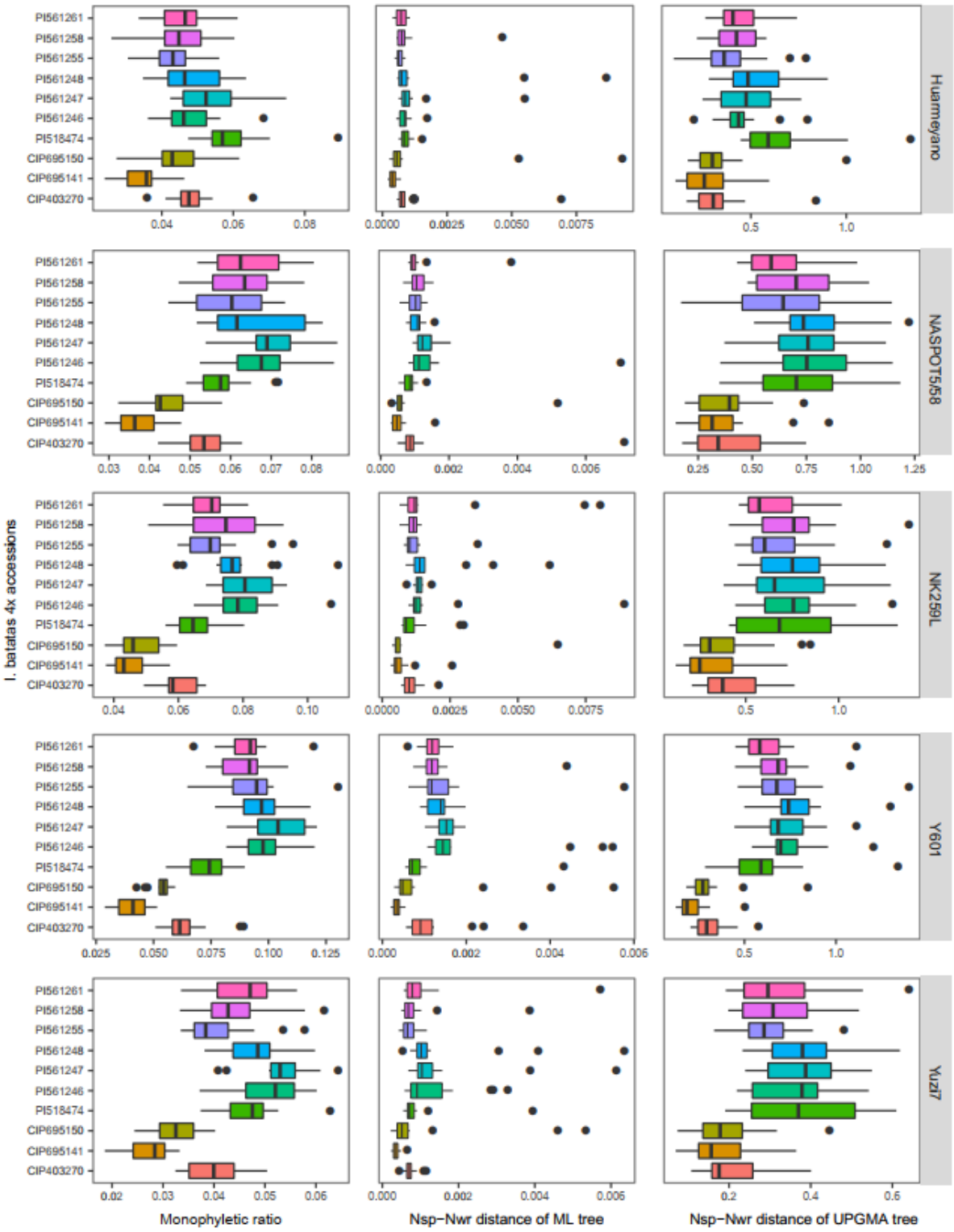

Figure 2

Relationships between sweet potato cultivars and I. batatas 4x accessions as revealed by HPA. Boxplots of the monophyletic ratios and the Nsp-Nwr distances of 15 chromosomes among ten I. batatas 4x accessions. The results achieved by each hexaploid reference (sweet potato cultivar) are presented in one row.

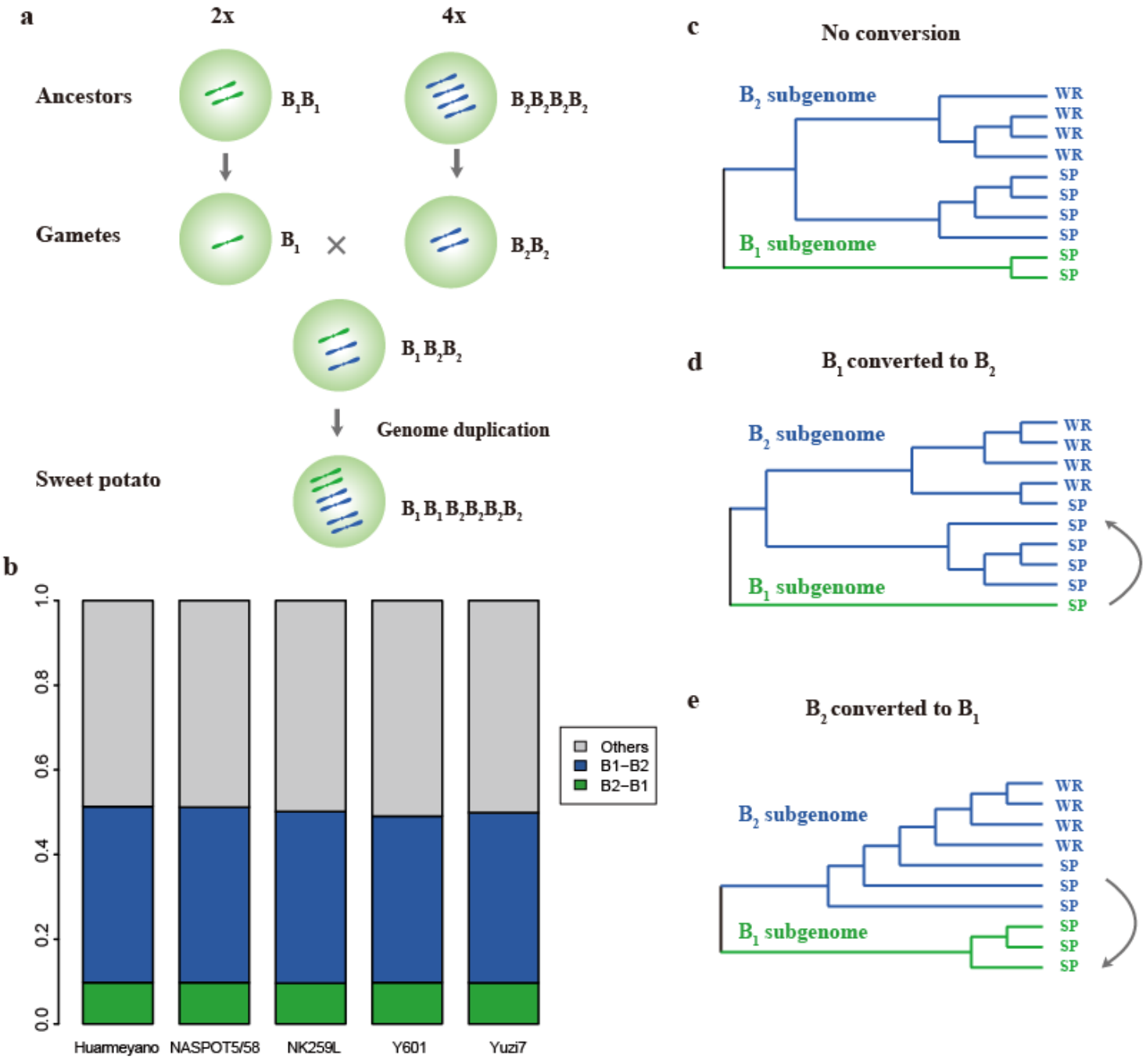

Figure 3

Gene conversions between sweet potato subgenomes. a, Model of the genetic origin of sweet potato. B1 subgenome chromosomes are represented in green and B2 subgenome chromosomes in blue. b, Gene conversion ratios in hexaploid sweet potato cultivars/landraces using accession CIP695141 as reference. B1 – B2, gene conversion events from the B1 to the B2 subgenome. B2 - B1, conversion events from B2 to B1 subgenome. Others, other scenarios, including no conversion and scenarios that could not be resolved. c-e, Examples of tree topologies under the scenarios of no conversion (c), B1 to B2 gene conversion (d), and B2 to B1 gene conversion (e). The B1 subgenome is shown in green and the B2 subgenome in blue. SP, sweet potato. WR, I. batatas 4x.

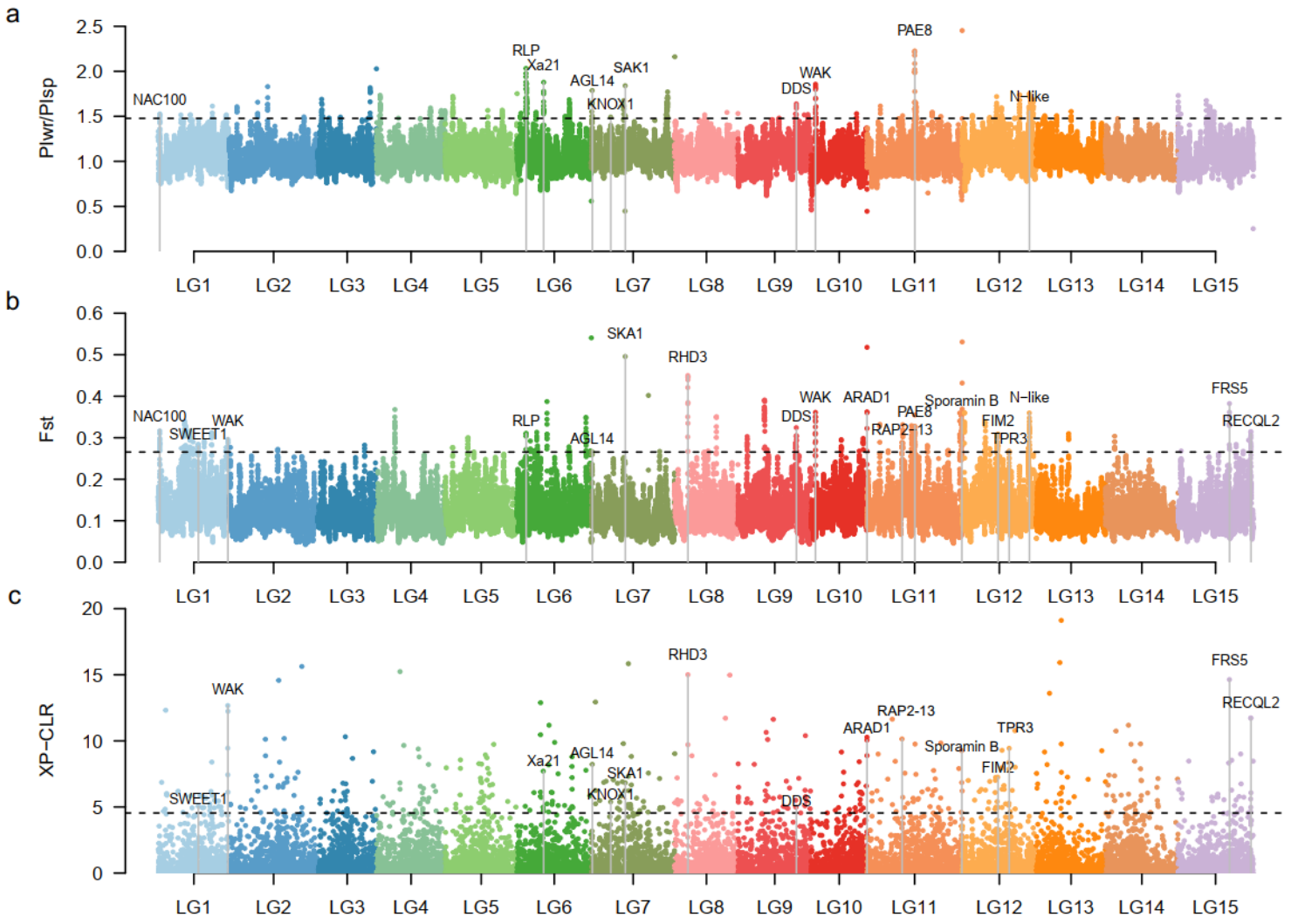

Figure 4

Profiling of the selective sweeps during sweet potato speciation and domestication. Selective sweep regions identified by (a) the greatest reduction in nucleotide diversity (π ratio, π wild relative/π sweet potato), (b) population differentiation (Fst) between wild relatives and sweet potato cultivars, and (c) cross-population composite likelihood ratio (XP-CLR) between wild relatives and sweet potato cultivars. Dashed lines indicate the regions that scored in the top 1% of the π ratio, Fst and XP-CLR values. Grey vertical bars show the positions of candidate genes related to domestication traits. wr, wild relative (refers to I. batatas 4x). sp, sweet potato.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HPAanalyzingnote.html
- Supplementarynotes2021722.docx
- codesummary.html

- tables20210527.xlsx
- Supplementaryfigures.docx