

# Lawrence Berkeley National Laboratory

LBL Publications

## Title

Diversification of mandarin citrus by hybrid speciation and apomixis

## Permalink

<https://escholarship.org/uc/item/3gx0n207>

## Journal

Nature Communications, 12(1)

## ISSN

2041-1723

## Authors

Wu, Guohong Albert

Sugimoto, Chikatoshi

Kinjo, Hideyasu

et al.

## Publication Date








2021

## DOI

10.1038/s41467-021-24653-0

Peer reviewed

# Diversification of mandarin citrus by hybrid speciation and apomixis

Guohong Albert Wu <sup>1,8</sup>, Chikatoshi Sugimoto<sup>2,8</sup>, Hideyasu Kinjo<sup>3</sup>, Chika Azama <sup>2</sup>, Fumimasa Mitsube<sup>3</sup>, Manuel Talon<sup>4</sup>, Frederick G. Gmitter Jr. <sup>5</sup> & Daniel S. Rokhsar <sup>1,2,6,7</sup>

The origin and dispersal of cultivated and wild mandarin and related citrus are poorly understood. Here, comparative genome analysis of 69 new east Asian genomes and other mainland Asian citrus reveals a previously unrecognized wild sexual species native to the Ryukyu Islands: *C. ryukyuensis* sp. nov. The taxonomic complexity of east Asian mandarins then collapses to a satisfying simplicity, accounting for tachibana, shiikuwasha, and other traditional Ryukyuan mandarin types as homoploid hybrid species formed by combining *C. ryukyuensis* with various mainland mandarins. These hybrid species reproduce clonally by apomictic seed, a trait shared with oranges, grapefruits, lemons and many cultivated mandarins. We trace the origin of apomixis alleles in citrus to mangshanyaju wild mandarins, which played a central role in citrus domestication via adaptive wild introgression. Our results provide a coherent biogeographic framework for understanding the diversity and domestication of mandarin-type citrus through speciation, admixture, and rapid diffusion of apomictic reproduction.

<sup>1</sup>DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>2</sup>Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan. <sup>3</sup>Okinawa Prefectural Agricultural Research Center, Nago Branch, Nago, Okinawa, Japan. <sup>4</sup>Centro de Genómica, Instituto Valenciano de Investigaciones Agrarias, Valencia, Spain. <sup>5</sup>Citrus Research and Education Center, Institute of Food and Agricultural Sciences, University of Florida, Lake Alfred, FL, USA. <sup>6</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. <sup>7</sup>Chan-Zuckerberg BioHub, San Francisco, CA, USA. <sup>8</sup>These authors contributed equally: Guohong Albert Wu, Chikatoshi Sugimoto. ✉email: [gwu@lbl.gov](mailto:gwu@lbl.gov); [fgmitter@ufl.edu](mailto:fgmitter@ufl.edu); [dsroksar@gmail.com](mailto:dsroksar@gmail.com)

Mandarin-type citrus comprise a heterogeneous group of east Asian citrus with small and typically easily peelable fruit<sup>1,2</sup>. Because of their consumer-friendly attributes, mandarins have seen the greatest percentage increase in global citrus production<sup>3</sup>. The phenotypic and genomic diversity of mandarin types in the Nanling region of southern China has driven speculations that mandarins first arose and were domesticated in this region<sup>4</sup>. Yet the nature of the domestication process, and the relationships among these mainland Asian types remain poorly understood.

These mysteries are compounded by the extensive mandarin diversity of the Ryukyu islands and mainland Japan<sup>5</sup>, including: tachibana [*C. tachibana* (Makino) Yu Tanaka], a culturally significant ornamental citrus grown throughout mainland Japan; shiikuwasha [*C. depressa* Hayata], grown in the Ryukyus and renowned for its health promoting qualities<sup>6</sup>; and other traditional and wild Ryukyuan citrus of unknown ancestry (Supplementary Note 1). Since tachibana and shiikuwasha have been found in wild forests, and are described in the earliest poetry of the Japanese and Ryukyuan kingdoms<sup>7–10</sup>, they have been presumed to be indigenous or natural species<sup>1,11</sup>, although some researchers have suggested that they are interspecific hybrids of various kinds<sup>1,12–14</sup>. The complexity of relationships among indigenous and cultivated mandarins across east Asia remains unclear, and is a barrier to understanding the origin and domestication of mandarins.

Most domesticated mandarins can reproduce true to type from seed (apomixis) by generating maternal clones from somatic tissue through the process of nucellar embryony<sup>15</sup>, which allows desirable genotypes to be replicated at scale. Although apomixis has been shown to be inherited in a dominant Mendelian fashion<sup>15–18</sup>, its natural origin and dispersal across diverse citrus are obscure. Mandarins are also widely hybridized with other citrus species to produce a diversity of familiar cultivated varieties including oranges, grapefruit, and lemons<sup>19,20</sup>, which also reproduce apomictically.

In order to resolve the relationships among wild and cultivated mandarins and explore the nature, evolution, and biogeography of east Asian citrus, here we present the genome sequences of 69 traditional, wild, and atypical citrus of the Ryukyus and southern mainland Japan (Supplementary Data 1 and 2; Supplementary Note 2), and analyze these genomes together with previously sequenced wild and domesticated Chinese mandarins, including those from Mangshan in the Nanling mountain range, and other citrus<sup>4,20,21</sup> (Fig. 1). We find that the complexity of mandarin relationships is considerably simplified by the discovery of three ancestral lineages which, together with pummelo, gave rise to all extant mandarin diversity by hybridization and introgression. One of these groups is a previously unknown wild species currently found in the Ryukyu islands; the other two are previously unrecognized sister subspecies of mainland Asian mandarin. Our analysis leads to a comprehensive revision of the origin and diversification of east Asian citrus, including the elucidation of the origins of apomixis in mandarin and its spread to related citrus including oranges, grapefruits and lemons.

## Results

**Citrus ryukyuensis is a new species of mandarin citrus.** Most strikingly, we identified a new wild citrus species native to the Ryukyu islands that we designate *C. ryukyuensis* sp. nov. (Supplementary Fig. 1, Supplementary Note 3). This new species is represented in our collection by eight wild Okinawan accessions that form a cluster of genomes separated from all previously sequenced species of *Citrus* (Fig. 1a). These accessions include ‘tanibuta’ types (“big seeds” in Okinawan dialect; Supplementary Note 1) that were described by Tanaka<sup>5</sup> as a Ryukyuan variant of

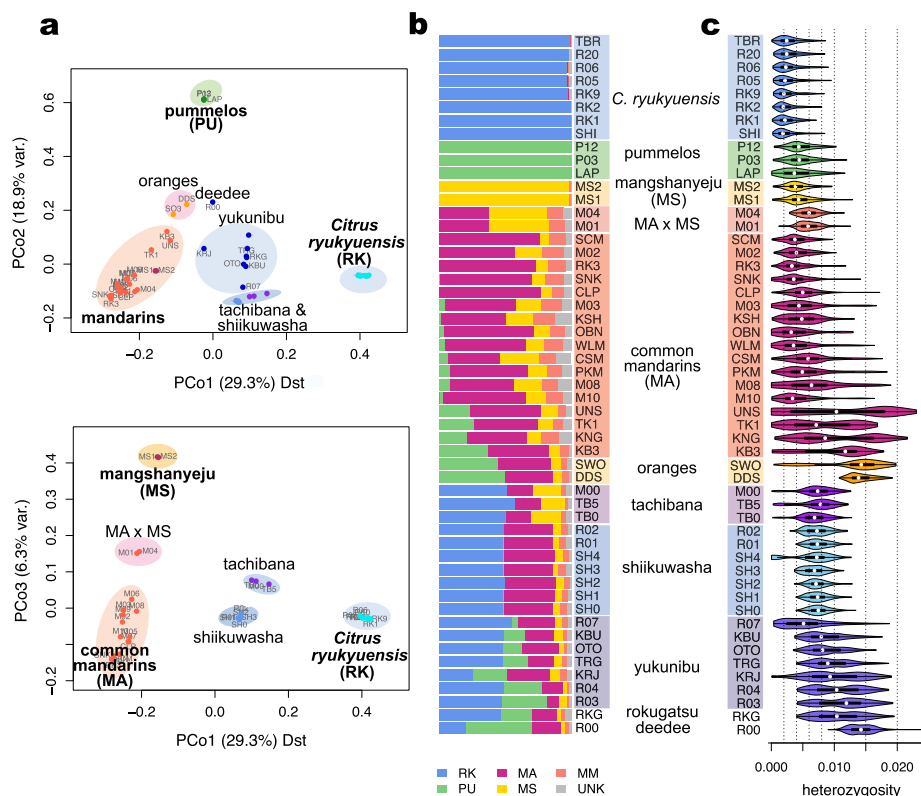
tachibana. We find that *C. ryukyuensis* is related to but genetically distinct from tachibana and shiikuwasha. Among their differences, *C. ryukyuensis* is a sexual species that produces monoembryonic seeds, while tachibana<sup>22</sup> and shiikuwasha<sup>11</sup> produce polyembryonic (apomictic) seeds.

The identification of *C. ryukyuensis* as a pure species (i.e., a distinct sexually reproducing population without admixture) is supported by three findings (Fig. 1). First, this population has low genome-wide heterozygosity (0.2–0.3%) that is comparable to or less than the variation seen within other recognized citrus species<sup>20</sup>, and smaller than the typical interspecific variation in citrus<sup>21</sup> (Fig. 1c). Second, *C. ryukyuensis* nuclear and cpDNA types form distinct clades to the exclusion of other mainland Asian citrus species (Supplementary Fig. 2). *C. ryukyuensis* is strongly differentiated from both *C. reticulata* (i.e., mainland Asian mandarin;  $F_{ST} = 0.67$ ) and *C. maxima* (pummelo;  $F_{ST} = 0.82$ ). Based on sequence comparisons, we estimate that *C. ryukyuensis* diverged from mainland Asian mandarins around ~2.2–2.8 Mya (Fig. 2, Supplementary Notes 3, 11). This divergence time is comparable to the split between other recognized citrus species, e.g., *Poncirus trifoliata* and *P. polyandra*<sup>23</sup>. Finally, the allele frequency spectrum in the *C. ryukyuensis* population decays roughly as expected for a panmictic sexual population (Supplementary Fig. 3), an observation that is consistent with monoembryony (Supplementary Fig. 1b) and zygotic (sexual) reproduction.

**Common mandarin and mangshanyejū are two subspecies of mainland Asian mandarin.** We uncovered further surprises when we analyzed the mainland Chinese wild mandarins sequenced by Wang et al.<sup>4</sup> in our expanded context (Supplementary Note 9). We find that wild mainland Asian mandarins comprise two sister populations with substantial genetic differentiation (Fig. 1, Supplementary Figs. 2 and 4), in contrast to Wang et al.’s description of a single wild population of Chinese mandarins from which domesticated varieties were selected. For taxonomic simplicity, we consider the two sister populations as sub-species of *C. reticulata* (Blanco). One sub-species, which we call ‘common mandarin,’ is the predominant contributor to domesticated mandarin. Many domesticated types, however, also contain admixture from the newly recognized second mandarin sub-species and from pummelo (Fig. 1b).

We identify the second mainland mandarin subspecies with types that are colloquially referred to as “mangshanyejū”, i.e., wild mandarin (“ju”) from the Mangshan region of the Nanling mountain range. Although Wang et al.<sup>4</sup> regard wild mandarins as an undifferentiated group, we find that mangshanyejū (MS) and common mandarin (MA) populations are sufficiently differentiated from each other ( $F_{ST} \sim 0.49$ ) that they should be considered as at least distinct sub-species (Supplementary Note 9). We estimate that these two populations diverged 1.4–1.7 million years ago (Fig. 2 and Supplementary Note 11). We find that the collection of Wang et al.<sup>4</sup> includes two pure mangshanyejū (MS1 and MS2) and two distinct F1 hybrids of mangshanyejū with nearly pure common mandarins (M01 and M04) (see Figs. 1 and 4c, Supplementary Fig. 4). Other citrus also have mangshanyejū ancestry (Fig. 1). For example, we find that yuzu, cultivated for its pleasing flavor and aroma valued in gastronomy and aromatherapy, is an F1 hybrid of mangshanyejū with an Ichang papeda seed parent (Supplementary Note 8). We caution that “mangshanyejū” should not be confused with ‘mangshanyegan’ (wild citrus (‘gan’) from Mangshan, *C. mangshanensis*), which is a distantly related citrus species from the same region<sup>4,20,21</sup>.

The estimated divergence times between *C. ryukyuensis* and *C. reticulata* (2.2–2.8 Mya), and between mangshanyejū and



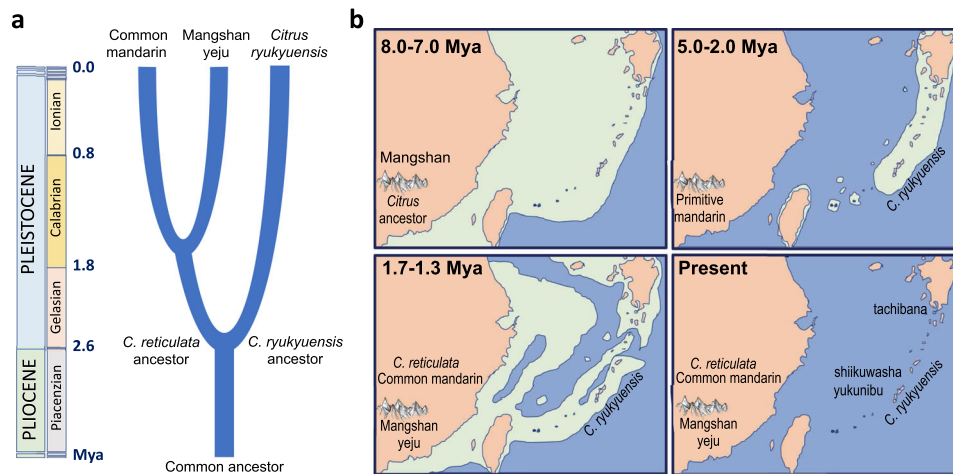
**Fig. 1 Population structure, genetic admixture, and heterozygosity of east Asian citrus.** **a** Multidimensional scaling (MDS) plot of 51 citrus accessions. Projection onto the first two principal coordinates (upper panel) shows *C. ryukyuensis* as a distinct population from tachibana, shiikuwasha, and other Ryukyuan hybrids (yukunibu and deedees). The third principal coordinate (lower panel) separates the two Mangshan wild mandarins (MS) from other mandarins. It also separates tachibana from shiikuwasha. For easier visualization, accessions with significant pummelo ancestry (pummelos, oranges, some mandarins, yukunibus) are not shown in the lower panel. See Supplementary Data 1 and 3 for accession code and names. **b** Four-way admixture plot of 53 citrus accessions based on local ancestry inference. PU=pummelo (*C. maxima*), RK=*C. ryukyuensis*, MS=mangshanyeu, MA=common mandarin, MM=generic *C. reticulata* without subspecies assignment (MS vs MA), UNK=unknown. Note that tachibana has more MS alleles than shiikuwasha and other Ryukyuan hybrids. Some wild mandarins (M01, M04) are hybrids with nearly equal contribution from the two subspecies of MS and MA. Common mandarins display varying degree of MS admixture. **c** Heterozygosity distribution violin plot for the same accessions as in **b**, for non-overlapping windows of 500,000 callable sites. *C. ryukyuensis* shows the lowest heterozygosity compared to tachibana, shiikuwasha and other hybrid types as well as accessions from *C. reticulata* and *C. maxima*. Median and quartiles are denoted by the white dot and black bar limits respectively, and whiskers are 1.5× inter-quartile range. Source data are provided as a Source Data file.

common mandarins (1.4–1.7 Mya) are consistent with the paleogeology of the region (Fig. 2, Supplementary Fig. 5, Supplementary Note 11). During the early diversification of citrus throughout southeast Asia in the Late Miocene (11.61–5.33 Mya)<sup>20</sup>, the boundary of mainland Asia included what is now the Ryukyu arc and the main islands of Japan<sup>24,25</sup>. Sea level rise and tectonic activity isolated the Ryukyu archipelago in the Pliocene (5.33–2.58 Mya) and Pleistocene, with intervals of connectivity to the south near present day Taiwan and north to the Japanese islands. This variable connectivity and associated climatic variation led to the emergence of new species in this region in many plant<sup>26–30</sup> and animal<sup>31–34</sup> taxa, coinciding with our estimates for the divergence of the distinct *C. ryukyuensis* from mainland Asian mandarin. The emergence of *C. ryukyuensis* by allopatric speciation was accompanied by a population bottleneck, suggested by its reduced heterozygosity relative to mainland mandarins (Supplementary Note 11).

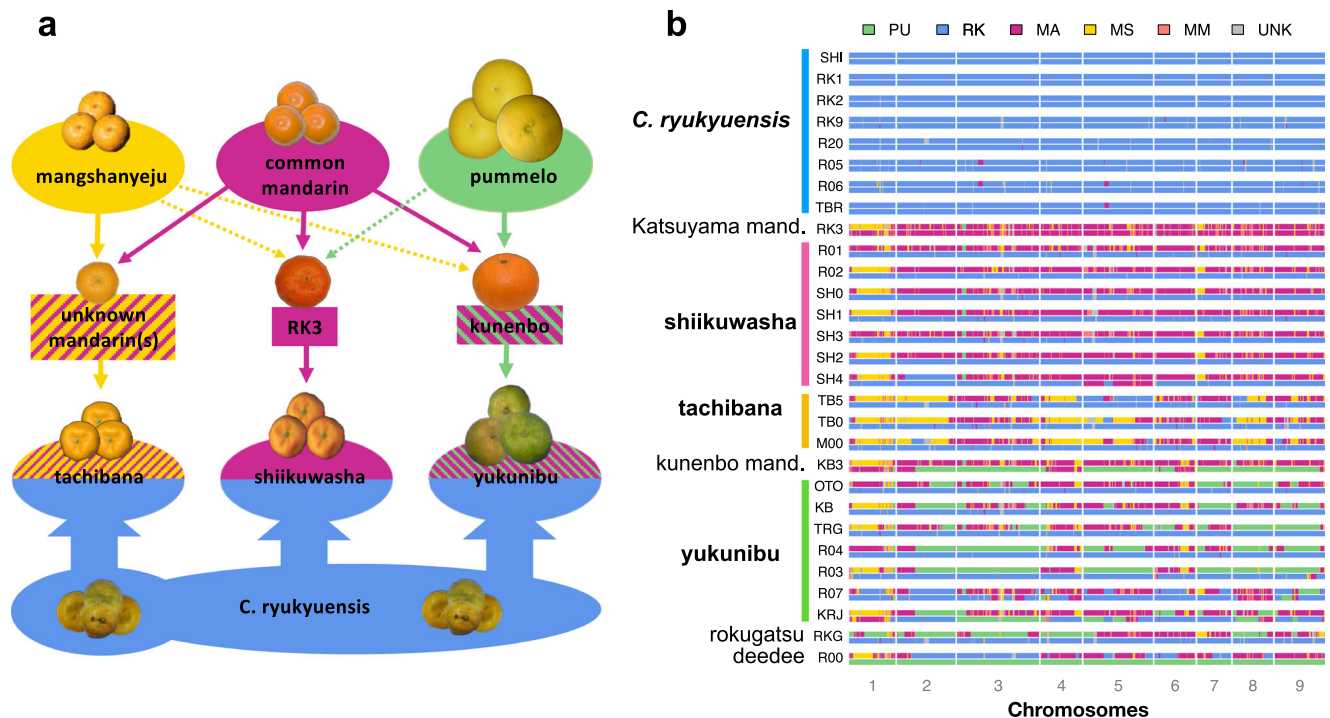
**New species illuminates origins of shiikuwasha, tachibana, and other Ryukyu types.** *C. ryukyuensis* provides the key to unlocking the origin and diversity of Ryukyuan and mainland Japanese citrus (Fig. 3). The shiikuwasha in our collection form a large half-sibling family with one mainland mandarin parent but distinct

*C. ryukyuensis* parents (Fig. 3, Supplementary Note 4). Unexpectedly, we found a clonal relative of this mainland mandarin parent of all shiikuwasha in a private collection in Nago City, Okinawa (RK3 in our designation, Fig. 3a). RK3 is referred to colloquially as an “ishikunibu” type, but is distinct from the shiikuwasha variety of the same name (Supplementary Note 4). Sequence comparison shows that RK3 is a close relative of the pure<sup>20</sup> Chinese mandarin Sun Chu Sha Kat (SCM) (coefficient of relatedness 0.41), but contains a single 2.4 Mbp introgressed pummelo segment.

The observation that shiikuwasha form a large half-sib family explains the previously puzzling finding that shiikuwasha chloroplast DNAs (cpDNAs) are of two distinct types<sup>35,36</sup>, either matching tachibana mandarins (here recognized as *C. ryukyuensis* type), or matching certain mainland Asian mandarins (here recognized as a *C. reticulata* type). Evidently, RK3 mandarin served as both seed and pollen parent in the numerous hybridization events that generated shiikuwasha. The hybrid nature of shiikuwasha accounts for its previously noted genotypic and phenotypic diversity<sup>8,12</sup>, and is consistent with previous suggestions that shiikuwas has hybrids based on high levels of nucleotide polymorphism<sup>1,12,37</sup>. More detailed understanding was elusive since *C. ryukyuensis* had not been recognized or characterized.



**Fig. 2 Chronogram of east Asian mandarin citrus speciation and biogeography in the Ryukyu Arc and mainland Japan.** **a** Population divergence times of *C. ryukyuensis* (2.2–2.8 Mya) and two subspecies of mainland Asian mandarins (*C. reticulata*): common mandarin and mangshanyeuju (1.4–1.7 Mya). Extant common mandarins are recent admixtures with both mangshanyeuju and pummelos. **b** Geological history of the Ryukyu Arc and evolutionary origins of east Asian citrus during four representative time periods: (1) initial radiation of citrus during the late Miocene<sup>20</sup> with subsequent dispersal to regions including Mangshan of the Nanling mountain range. The exact arrival time of primitive mandarins at Mangshan cannot be determined and could be as late as the Pliocene epoch (5.3–2.6 Mya) (top left), (2) geographical isolation and genetic divergence of *C. ryukyuensis* in the Ryukyu Arc from mainland Asian mandarins during early Pleistocene (top right), (3) divergence of mangshanyeuju and common mandarins (bottom left), and (4) current distribution of east Asian citrus with *C. ryukyuensis* ancestry in the Ryukyu Arc and mainland Japan, as a result of distinct hybridization events with different migrant mainland mandarins (bottom right). (Maps are adapted from Kimura<sup>25</sup> with paleo-landmasses in light green.) Source data underlying Fig. 2a are provided as a Source Data file.



**Fig. 3 Hybrid speciation and admixture map of Ryukyuan and mainland Japanese citrus.** **a** Origin of Ryukyuan and mainland Japanese citrus types (tachibana, shiikuwasha, yukunibu) derived from four ancestral populations. Thick arrows denote ancestry involving multiple individuals from a population, whereas a thin arrow stands for single individual ancestry. Dotted and solid lines from the top row denote small and significant introgression, respectively. For example, RK3 has small amount of pummelo admixture whereas kunenbo has significant pummelo introgression. The shiikuwashas are half-sibs sharing the same mainland Asian mandarin parent (RK3) but different *C. ryukyuensis* parents. Kunenbo (KB3) is the seed parent of the yukunibu group. **b** Four-way admixture map for Ryukyuan and mainland Japanese citrus types. Population code as in Fig. 1b. Tachibana genomes are characterized by both significant admixture with MS and segments of diploid *C. ryukyuensis* genotype. SH4 is a seedless shiikuwasha. Source data underlying Fig. 3b are provided as a Source Data file.



We find that tachibana is also a collection of hybrids between *C. ryukyuensis* and mainland Asian mandarins, but distinct from shiikuwasha (Fig. 3b, Supplementary Note 5). The extensive sharing of mainland mandarin haplotypes among our tachibana genotypes is consistent with a single Chinese mandarin-like parent although we cannot rule out a small number of closely related mandarin parents (Supplementary Fig. 6). Importantly, the mainland parents of shiikuwasha and tachibana are not related, implying that these geographically separated hybrid species arose independently. In contrast to the simple interspecific F1 hybrid origin of shiikuwasha, tachibana genotypes are more complex. Each tachibana carries 4–6 multi-megabase segments of diploid *C. ryukyuensis* within an otherwise *C. ryukyuensis* × *C. reticulata* hybrid background, which implies that the direct mandarin-like parents of tachibana themselves had prior introgression of *C. ryukyuensis* (Fig. 3b). Our study shows that tachibana are not generally full siblings, as suggested by an earlier marker-based analysis of three accessions<sup>13</sup>.

Finally, much of the remaining diversity of indigenous Ryukyuan citrus (including several other named species<sup>5,38–40</sup>) can be organized into a third hybrid family that we named yukunibu, meaning “sour citrus” in Okinawan dialect. Yukunibu citrus are F1 hybrids with a kunenbo-mikan seed parent and diverse *C. ryukyuensis* pollen parents (Fig. 3, Supplementary Note 6). The yukunibu group unites three cultivated accessions (oto, kabuchii, and tarogayo, grown for juice) with two others. While kabuchii’s kunenbo-mikan ancestry was previously suggested<sup>13,39</sup>, its other familial relationships were not previously recognized. The yukunibu family presumably arose soon after kunenbo-mikan was introduced to the Ryukyus (and then mainland Japan) from Indochina sometime between the 8th and 12th centuries<sup>41–45</sup>. Our collection also contains other members of the extended yukunibu family as well as other unrelated hybrid genotypes with *C. ryukyuensis* ancestry (Supplementary Note 7, Supplementary Fig. 7).

**Apomixis.** While *C. ryukyuensis* is a sexually reproducing species, its hybrid derivatives (shiikuwasha, tachibana, and yukunibu) reproduce apomictically by nucellar embryony. This implies that the apomixis trait was transmitted to these three hybrid species by their migrant mainland mandarin parents. Apomictic reproduction of shiikuwasha<sup>46</sup> and tachibana<sup>22</sup> enabled the rapid establishment and dispersal of these new hybrid species after their formation by hybridization with the pre-existing locally adapted *C. ryukyuensis* population. Notably, the mainland mandarin parents of shiikuwasha and yukunibu (RK3 and kunenbo-mikan) both produce polyembryonic seed<sup>16</sup>.

All apomictically reproducing citrus in our collection carry a recently described MITE (miniature inverted-repeat transposable element) DNA transposon insertion in the promoter of the *CitRKD1*<sup>18</sup> gene (also known as *CitRWP*<sup>17</sup>) that dominantly confers an apomictic phenotype. We find that this MITE insertion arose in the mangshanyejū population in the early Pleistocene and spread into other mandarins, oranges, grapefruits, and lemons as well as shiikuwasha, tachibana, and yukunibu (Fig. 4; Supplementary Note 10). The four mangshanyejū-derived MITE haplotypes of *CitRKD1* (from two related groups) occur in different combinations (along with the ancestral allele without MITE insertion) in diverse apomictic citrus, highlighting the complex history of this critical genomic region (Supplementary Fig. 8; Supplementary Data 1 and 3).

Selection for apomixis explains widespread mangshanyejū admixture across cultivated and wild mandarins, especially on chromosome 1 around the *CitRKD1* gene (Figs. 3b, 4c). The adaptive wild introgression of apomixis alleles was a driver of

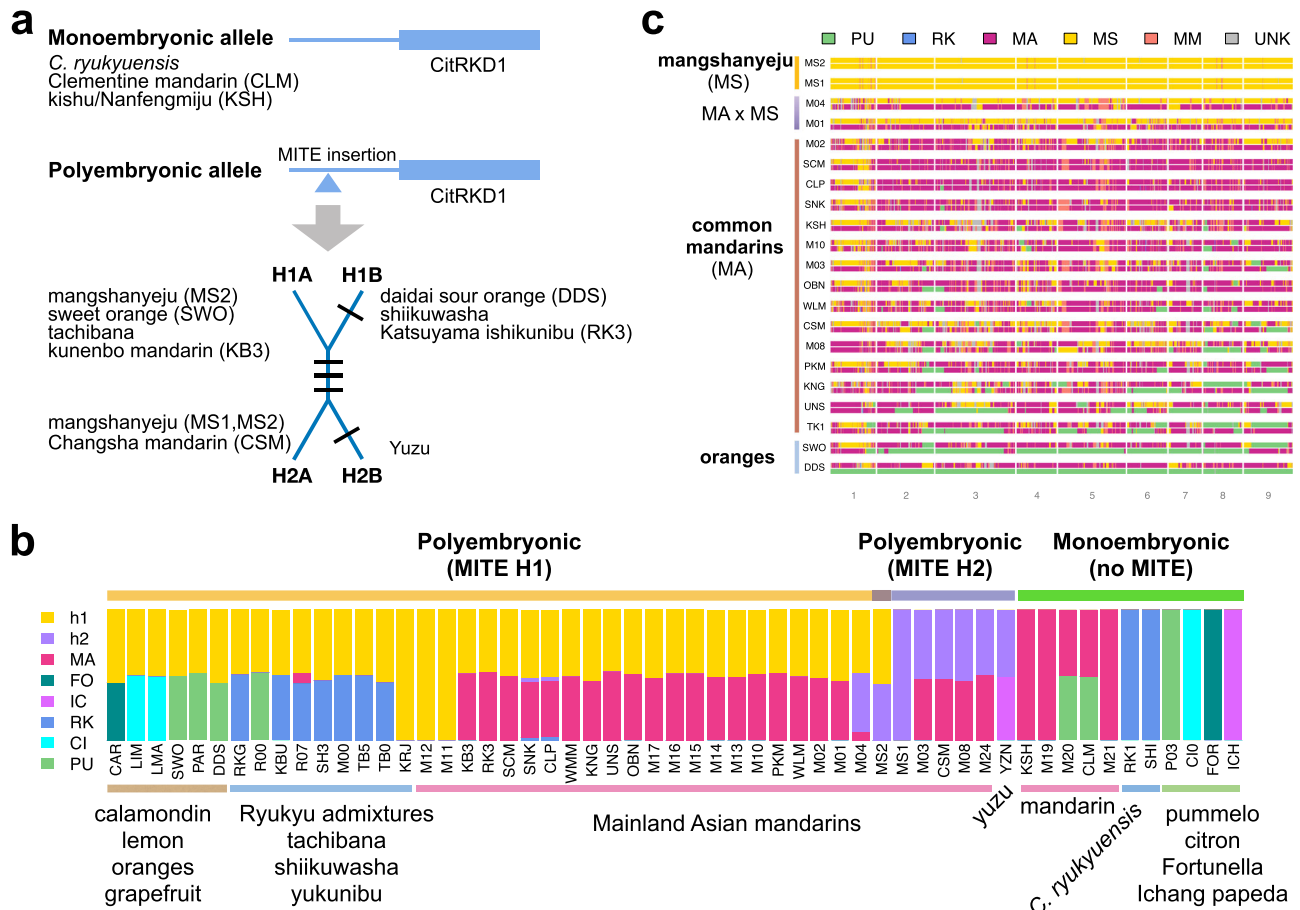
domestication because it allowed the mass production of desirable types. This is consistent with the limited diversity of chromosome 1 haplotypes around the apomixis locus in cultivated mandarin and mandarin hybrids (Fig. 4b). In contrast, the other eight chromosomes in mandarins are typically dominated by alleles from the “common” mandarin sub-population, including those believed to confer low acidity<sup>20,47,48</sup>, anthocyanin production<sup>49–51</sup>, self-compatibility<sup>43,52</sup>, and to regulate flesh and peel color<sup>53–55</sup> and volatile aromatics that contribute to flavor<sup>56</sup>.

**Formation of east-Asian mandarin types.** The three native Ryukyuan and Japanese hybrid species with mixed ancestry—shiikuwasha, tachibana, and yukunibu—arose by independent hybridizations of one or a few mainland mandarin migrants with the native wild sexual *C. ryukyuensis*. Since shiikuwasha and tachibana are referenced in native poetry and songs, and were found in wild forests, we suggest that the mainland Asian mandarin founders of these species were chance prehistoric introductions during intervals of connectivity with mainland Asia (40,000–200,000 years ago)<sup>25</sup>. Alternatively, their seeds could have arrived via the Kuroshio Current<sup>34,57</sup>, a powerful south-to-north warm current that passes by Taiwan, the Ryukyu Arc and mainland Japan. In contrast, kunenbo-mikan (the mainland Asian parent of yukunibu) was introduced to the Ryukyus by trade between the 8th and 12th centuries<sup>42</sup>. Other imported mainland Asian citrus such as sour orange (daidai) and pummelo also hybridized with *C. ryukyuensis* and shiikuwasha to produce rokgatsu and Ryukyu sour orange (deedee), respectively (Supplementary Fig. 7, Supplementary Note 7).

Although some authors have suggested a close relationship between shiikuwasha and tachibana<sup>1,36</sup>, we find them to be distinct hybrid species. Shiikuwasha and tachibana differ both in their mainland Asian mandarin parents and by the differentiation of their *C. ryukyuensis* ancestors. While the mainland Asian mandarin parent of shiikuwasha had prior pummelo admixture but limited mangshanyejū introgression, the unknown mainland Asian mandarin parent(s) of tachibana had substantial mangshanyejū admixture (48–54%, comparable to an F1 hybrid of mangshanyejū and common mandarin) and no pummelo admixture. On the *C. ryukyuensis* side, genetic differentiation among tachibana and shiikuwasha haplotypes ( $F_{st} = 0.17–0.20$ ) is consistent with separate mainland Japanese and Ryukyuan *C. ryukyuensis* populations that diverged ~220,000–350,000 years ago (Supplementary Note 11). The *C. ryukyuensis* haplotypes of shiikuwasha and yukunibu are more closely related to the extant *C. ryukyuensis* population in Okinawa.

## Discussion

Here we have shown that extant mandarin diversity arose from three ancestral populations distinguished by comparative genome analyses: the island species *C. ryukyuensis*, and common mandarin and mangshanyejū from mainland Asia. Complex patterns of admixture involving these previously unrecognized founding populations, combined with clonal propagation by apomixis, produced the extensive heterogeneity of mandarin citrus. Our approach consists of identifying natural or pure species based on distinctive patterns of genetic variation, followed by characterization of hybrid genotypes in terms of these founding types<sup>20,21,58</sup>. These new insights into wild mandarin diversity, coupled with the recognition that widespread pummelo introgression has also contributed to mandarin domestication, enables a comprehensive admixture-informed classification scheme for mandarin citrus<sup>20,59</sup> (Supplementary Note 12). Consideration of domestication phenotypes including nucellar embryony, fruit size and palatability<sup>20</sup> supports its practical use.



**Fig. 4** Ancestry of apomixis alleles and two subspecies of mainland Asian mandarins (*C. reticulata*). **a** Diversity of the apomixis alleles in mandarins and inter-specific mandarin hybrids. The ancestral allele does not have the MITE transposon insertion in the promoter of the *CitRKD1* gene regulating citrus apomixis. Derived alleles with the MITE insertion are dominant for the nucellar embryony phenotype. Four MITE haplotypes in two haplogroups (H1=H1A and H1B; H2=H2A and H2B) are observed among sequenced mandarins and hybrids with each black line denoting a segregating SNP. Listed next to each MITE allele type are representative citrus accessions containing that allele. **b** Genetic ancestry of the citrus polyembryonic locus (200 kb region flanking *CitRKD1* gene). Fifty-five accessions derived from six progenitor species are analyzed with ADMIXTURE<sup>68</sup> and the eight-population ( $K=8$ ) structure is presented with additional figures shown in Supplementary Fig. 8. (PU=pummelo, CI=citron, RK=*C. ryukyuensis*, IC=Ichang papeda, FO=*Fortunella* (kumquat), MA= common mandarin, h1 and h2 have mangshanyejū ancestry). Accessions with h1 ancestry contain MITE H1A or H1B, whereas those with h2 ancestry have MITE H2A or H2B. MS1 and MS2 are two mangshanyejū accessions. All sequenced polyembryonic accessions carry the dominant allele with the MITE insertion and have mangshanyejū ancestry at this locus, whereas monoembryonic accessions have common mandarin but not mangshanyejū ancestry. **c** Genome-wide local ancestry inference of mainland East Asian citrus with four ancestral populations including two subspecies of *C. reticulata* (MS, MA). Population code as in Fig. 1b. This figure complements Fig. 3b by considering 21 accessions without *C. ryukyuensis* ancestry. Note that the apomixis locus is located near the end of chromosome 1 (based on the Clementine reference sequence) which exhibits extensive MS admixture in common mandarins relative to other chromosomes. MS admixture is widespread in all sequenced mandarins. Two wild mandarins (M01=Daoxian wild mandarin and clonal relatives, M04=Suanpangan) show hybrid ancestry with nearly equal contribution from MS and MA. Source data underlying Fig. 4b and c are provided as a Source Data file.

We have shown that shiikuwasha, tachibana, and yukunibu are distinct homoploid (i.e., non-polyploid) hybrid species. Homoploid hybrid species are rare and are generally complex mixtures of parental species due to the absence of reproductive isolation between F1 hybrids and their parents<sup>60,61</sup>. Such introgressive hybridization promotes gene flow and the potential for fixing novel genetic combinations from the parental species. In contrast, shiikuwasha, tachibana, and yukunibu are fixed F1 hybrids that carry substantial genetic variation from their sexual parental species (*C. ryukyuensis*) on one haplotype, but only limited variation from their apomictic mainland parent(s) on the other. These hybrid genetic configurations are stably propagated by ongoing apomictic reproduction, which also reproductively isolates the hybrids from their parental species. This scenario provides a new model for homoploid hybrid speciation involving apomixis that could also apply to other plant taxa.

Our new conceptual framework for mandarin evolution and domestication illuminates the origin of other key traits besides apomixis. For example, loss of anthocyanin production is widespread among domesticated mandarins and some of their hybrid relatives<sup>50</sup>, and results in the appealing white flowers celebrated in ancient poetry compared with the ancestral purple color seen in other citrus. While elegant genetic analyses have shown that this recessive trait is due to defective alleles of the MYB transcription factor *Ruby*<sup>49,50</sup>, its evolutionary origin is unknown. Since the same deletion allele is fixed in both mangshanyejū and *C. ryukyuensis*, it was likely already fixed in the common Pleistocene ancestor of all mandarins. (An overlapping second deletion allele is now also segregating in common mandarins). It follows that loss of anthocyanin in mandarins preceded domestication (Supplementary Note 10). While adaptive introgressions from wild populations have played a notable role in crop and animal

domestication<sup>62,63</sup>, the introgression of apomixis and anthocyanin loss from wild populations occurred prior to domestication of mandarins, highlighting the importance of wild alleles at all stages of domestication.

From an agronomic point of view, apomixis contributed to the spread of appealing phenotypes, accelerating domestication and impacting subsequent breeding strategies. This behavior is especially desirable in citrus and other woody plants with longer reproductive cycles. In our scenario, the domestication of mandarin citrus is characterized by three main events: (1) the rise of apomixis in the mainland mandarin lineage during Pleistocene, (2) pummelo introgression that incorporated desirable traits into the ancestral mandarin genome<sup>21</sup>, and (3) the myriad of crosses between those ancestral hybrids and admixtures, that gave rise to the current basic types of edible citrus<sup>20,58</sup> (Supplementary Note 12). These findings provide inspiration for the breeding of new hybrids with disease-tolerance—for example, one of the top performing rootstocks tolerant to Huanglongbing (citrus greening disease) is a hybrid with shiikuwasha parentage<sup>64</sup>—as well as other desirable horticultural and nutritional characteristics<sup>6,65</sup>.

## Methods

**Sampling of diverse east Asian citrus.** Currently available genomic data for mandarin (*C. reticulata*) and related citrus are concentrated on (1) cultivated varieties<sup>17,20,21</sup> and (2) wild varieties from southeastern China, the center of diversity for mandarin citrus<sup>4</sup>. The extensive wild and cultivated diversity in the Ryukyu islands and southern mainland Japan has been largely neglected with only two tachibana accessions included in previous collections<sup>4,20</sup>.

To sample diverse east Asian citrus, we collected (1) named varieties of shiikuwasha and other traditional cultivated varieties (oto, kabuchi, tarogayo) from local Okinawan growers and the Okinawa Prefectural Agricultural Research Center, (2) unnamed local mandarin-type citrus trees from personal gardens and small farms, including the Katsuyama Shiikuwasha Co., Ltd., (3) wild citrus trees from various sources including naturally forested areas, including samples later identified as belonging to the new species *C. ryukyuensis*, (4) local island-grown trees of known or uncertain provenance from Okinawa World Theme Park, and (5) wild and cultivated tachibana from southern mainland Japan. Our collection also includes two distant relatives of citrus in the Rutaceae family (*Murraya paniculata* and *Toddalia asiatica*) that are not used in the present analysis. In total, 71 samples including 69 citrus were collected with appropriate permissions. More detailed information about these collections is provided in Supplementary Note 2 and Supplementary Data 2.

**Genome sequencing and genotyping.** Shoots or young leaves were collected from each sample in the field. Leaves were frozen with liquid nitrogen, and then crushed in a mortar. Total DNA was extracted from the frozen powder of leaves using a DNeasy® Plant Mini Kit (QIAGEN Co., Hilden, Germany). After libraries had been prepared with a KAPA HyperPlus Library Preparation Kit (F. Hoffmann-La Roche, Ltd., Basel, Switzerland) and NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina (NEW ENGLAND BioLabs, Inc., Ipswich, MA), 150 and 250 bp pair-end reads were obtained using a HiSeq 2500 Rapid v2, HiSeq 4000 and NovaSeq 6000 SP (Illumina, Inc., San Diego, CA). Each sample was sequenced at more than 30-fold redundancy.

Illumina paired-end reads from each accession (both new and from refs. 4,17,20,21) were mapped to the haploid Clementine reference sequence v1.0<sup>21</sup> using BWA-MEM (version 0.7.8-r455)<sup>66</sup>. We used the Clementine because it is the best quality available mandarin-type reference genome, and therefore suitable for exploring variation in mandarin. It has previously been established<sup>20,21</sup> that Illumina data from diverse citrus species can be readily aligned to this reference sequence, and the average mapping rate across samples sequenced in this study is 97.4%. Duplicate reads were removed using picard MarkDuplicates (version 2.13.2). Variants were called using GATK HaplotypeCaller (version 3.7-0-gcfed67)<sup>67</sup> with filtering based on read map quality, base quality, read depth and allele balance (Supplementary Note 2).

**Identification of ancestral populations.** In order to identify natural or pure species or sub-species, we sought groups of individuals whose genomes are closely related to each other as currently recognized species and conversely lack high heterozygosity characteristics of interspecific hybridization. We used three complementary approaches, sliding window analysis of pairwise genomic distance, multidimensional scaling, and admixture analysis.

The distance measure  $D$  between two diploid genomes 1 and 2 is defined by<sup>21</sup>

$$D = 1 - \frac{\pi_1 + \pi_2}{4\pi_{12}} \quad (1)$$

where  $\pi_1$  and  $\pi_2$  are the respective heterozygosities (i.e., nucleotide diversity) of the two accessions, and  $\pi_{12}$  is their sequence divergence (i.e., probability that randomly chosen alleles from the two diploids are different). The value of  $D$  ranges from 0 to 1, with  $D = 0$  for clones,  $D = 0.5$  for two unrelated individuals from a panmictic population, and  $D$  approaching 1 for two deeply divergent species.

We used a sliding window of 200 kb to calculate the distance  $D$  between two differentiated populations (e.g., different species or sub-species) taking one accession from each population, along with the heterozygosity of each genome. Genome wide values of  $D$  consistently above 0.5 without abrupt changes in heterozygosity indicate pure genomes without admixture. Deviations from this pattern at certain windows suggest admixture for those genomic regions<sup>21</sup>.

We performed multidimensional scaling analysis using the R (version 3.5.1) function `cmdscale` based on pairwise genomic distances ( $D$  defined above). We first identified clones ( $D \approx 0$ ) and chose the accession with the highest sequencing depth to represent each clonal group. Results for east Asian mandarin-type citrus are shown in Fig. 1a. Pure taxa (PU, RK, MS, MA) are found as corner clusters, and accessions lying between corners or near the middle of the diagram are found to be hybrids formed from these pure taxa.

We also identified ancestral populations and estimated genome-wide ancestry proportions using ADMIXTURE<sup>68</sup>. For mainland Asian citrus, the lowest cross-validation error was obtained with  $K = 6$  corresponding to common mandarin (MA), mangshanyejū (MS), pummelo (PU), citron, *Fortunella* (kumquat), and Ichang papeda (Supplementary Fig. 4a, Supplementary Note 9). Since citron, *Fortunella*, and Ichang papeda are absent from our mandarin core set, they do not appear in the multidimensional scaling of Fig. 1a. ADMIXTURE analysis of the mandarin core set of Fig. 1a (i.e. omitting known citron, kumquat, papeda and their hybrids) also identifies *C. ryukyuensis*, common mandarin, and mangshanyejū to be distinct populations [data not shown]. Further evidence for the distinctness of mangshanyejū and common mandarin is found in clustering by pairwise genomic distance (Supplementary Fig. 4b). We note that accessions MS1 and MS2 are pure mangshanyejū, and M01 and M04 are F1 hybrids of mangshanyejū and common mandarin.

**Phylogenetic inference.** For phylogenetic inference of Asian citrus species based on nuclear genomes, we used single nucleotide polymorphisms in the introns and UTRs to minimize selection pressure bias. Each diploid genome was reduced to a haploid sequence by randomly sampling one allele at each variant position (the species phylogeny is insensitive to this sampling procedure). We required complete coverage across all representative individuals (i.e., no missing genotype calls) and recovered a total of 209,124 single nucleotide variable sites. Using these characters we constructed a maximum likelihood tree with RAxML<sup>69</sup> under the general time-reversal model of nucleotide substitution with 1000 bootstrap replicates (“raxmlHPC -m GTRGAMMA -N 1000”). The tree was rooted with Chinese box orange (*Severinia buxifolia*, also known as *Atalantia buxifolia*). The highly supported topology of this species tree (Supplementary Fig. 2a) is in agreement with our previously published nuclear genome phylogeny based on SNPs in complementary regions of the genome (non-genic, non-repetitive, and non-centromeric)<sup>20</sup>, affirming the robustness of the tree topology.

We determined the chloroplast genotype of each accession by mapping reads to the chloroplast genome sequence of sweet orange<sup>70</sup>. This is an appropriate reference for genome wide pan-citrus comparisons. We constructed maximum likelihood phylogenetic trees using RAxML<sup>69</sup> under the general time-reversal model of nucleotide substitution with 1000 bootstrap replicates. (Supplementary Fig. 2b, Supplementary Note 2).

**C. ryukyuensis allele frequency spectrum.** The allele frequency spectrum (AFS) for the Okinawa *C. ryukyuensis* population was computed based on single nucleotide polymorphisms of the eight distinct sequenced accessions, excluding short admixed genomic segments and using only sites with no missing data (Supplementary Fig. 3). For demographic inference, we used moments<sup>71</sup> to model the folded AFS to detect the possible existence of a population bottleneck. A likelihood ratio test between a panmictic constant effective population size model (no bottleneck) and a two-epoch model shows that the two-epoch model with a population bottleneck provides a better fit. Though the strength of the bottleneck cannot be determined based on the AFS alone due to the small sample size (Supplementary Note 3), it can be estimated in the context of a 4-population divergence model (Supplementary Note 11).

**Divergence time and effective population size estimates.** To estimate the population divergence times and effective population sizes of pummelos, *C. ryukyuensis*, mangshanyejū and common mandarins (Supplementary Fig. 5a), we first derived the joint allele frequency spectrum of eight accessions (two per population) from genomic regions without admixture. We implemented demographic inference using moments<sup>71</sup>, a python package that can efficiently simulate multidimensional allele frequency spectrum and infer demographic history. Time calibration was based on a late Miocene citrus leaf fossil<sup>72</sup> and a previous estimate for mandarin-pummelo divergence<sup>20</sup> (Supplementary Note 11). For effective population size estimates, we used a generation time of 10 years. For moments simulations, multiple runs were performed with independent starting points in the high-



dimensional parameter space and checked for convergence of the likelihood and model parameter values. The estimated model parameters are listed in Supplementary Fig. 5b with uncertainties reflecting the time calibration range.

**Ancestry informative markers and local ancestry analysis.** In order to characterize the ancestry of mandarin types in detail, we first identified a genome-wide set of ancestry informative markers (AIMs) for four ancestral populations (RK=*C. ryukyuensis*, MS=mangshanyaju, MA=common mandarin, PU=pummelo) using pure, or mostly pure, individuals that were identified by sliding window analysis of pairwise genomic distance (*D*) and multidimensional scaling. AIMs were derived using three pure pummelos, three pure *C. ryukyuensis* accessions, two pure mangshanyaju, and three common mandarins as exemplars for the four ancestral populations. We note that our dataset only includes two pure mangshanyaju. We also found that common mandarins generally contained sub-chromosomal genomic segments with MS or PU ancestry. Segments of these exemplars with evidence for admixture were excluded. (Supplementary Note 3).

AIMs for each target population were defined as single nucleotide variants that are fixed in the target population exemplars relative to the other population exemplars, i.e., homozygous in the target exemplars but not found in the others. Since MS and MA are more weakly differentiated than other population pairs, we introduced a super-population MM to represent markers that are fixed in *C. reticulata* (combined MA and MS, together with mainland Asian mandarin) relative to PU and RK. In this way, we obtained a total of 397,887 ancestry informative markers: 268,383 for PU, 54,325 for RK, 35,067 for MA, 22,482 for MM, and 17,630 for MS.

We used these genome-wide AIMs to carry out local ancestry analysis in our collection of east Asian citrus (Figs. 1b, 3b, 4c). Sliding windows of 500 AIMs were employed and the ancestry for each window was assigned using a likelihood-based approach, following Wu et al.<sup>20</sup>. To call ancestry within a window, we required at least 5 AIMs for each ancestral population, otherwise the ancestry was assigned as Unknown. We note that our local ancestry method is in general agreement with but more sensitive than ADMIXTURE<sup>68</sup>, which may fail to detect short blocks of admixture at the a few percent level.

**Haplotype sharing and familial relationships.** Genetic relatedness between a pair of diploid individuals can be quantified by the proportion of their genomes that share zero, one or two haplotypes that are ‘identical by descent’ (IBD0, IBD1, and IBD2). The familial coefficient of relatedness is defined from these quantities via  $r = \frac{1}{2} \text{IBD1} + \text{IBD2}$ . Following<sup>20,21</sup>, we infer identity by descent over non-overlapping 200 kb windows using (1) the genetic distance *D* defined above and (2) the identity-by-state ratio,

$$\text{IBSR} = \text{IBS2} / (\text{IBS2} + \text{IBS0}) \quad (2)$$

where IBS2 is the number of shared heterozygous sites in a window (i.e., joint-genotype AB|AB, sharing two different alleles identical-by-state), and IBS0 is the number of homozygous differences (i.e., joint-genotype AA|BB, no allele sharing). For individuals from the same population, IBSR is independent of allele frequencies and has a mean of 2/3 for two unrelated individuals if the population is panmictic<sup>73</sup>. Importantly, if two individuals share the same haplotype across a window, then IBS0=0 and IBSR=1. We infer the IBD state for each window using the following criteria<sup>20</sup>: If IBSR < 0.95, the genomic window is assigned IBD0. If IBSR >= 0.95 and *D* < 0.05, the window is assigned IBD2. If IBSR >= 0.95 and *D* > 0.05 the window is assigned IBD1.

Genomic windows for two interspecific hybrids need to be treated differently, since in these cases the IBSR value is inflated by species-specific alleles and does not reflect shared haplotypes<sup>21</sup>. For such regions, we inferred the IBD state by comparing phased haplotypes instead of diploid genotypes. We performed inter-specific phasing using representative accessions from the two parental populations<sup>20</sup>. To allow errors from SNP calling and phasing, we consider two haplotypes identical if the mismatch rate is below  $2 \times 10^{-4}$ .

By this method we find that the mainland Asian mandarin RK3 shares one haplotype with all shiikuwasha across the entire genome. Since shiikuwasha are F1 interspecific hybrids, this implies that RK3 is the common parent of shiikuwasha (or more properly in light of its apomictic reproduction, a clone of the common parent). We found that the elite Nakamoto seedless shiikuwasha (SH4) is a somatic mutant of one of the six basic shiikuwasha genotypes (SH2) (Supplementary Note 4). Similarly, kunenbo-mikan shares one haplotype with oto, kabuchii, tarogayo, and two other accessions, implying that the kunenbo-mikan genotype is parental to these accessions (which we call the yukunibu group). The *C. ryukyuensis* haplotypes of various shiikuwasahas and yukunibus are found to be unrelated by direct comparison.

For tachibana we could not identify a mainland mandarin parent in our collection, but interspecific phasing and haplotype sharing analysis shows that there are at most two distinct *C. reticulata* haplotypes across three distinct tachibana genotypes. This is consistent with a single mainland Asian mandarin ancestor. By contrast, each tachibana carries a distinct *C. ryukyuensis* haplotype (Supplementary Fig. 6a) implying different *C. ryukyuensis* parents. The pairwise genomic IBD proportions and the coefficient of relatedness are shown in Supplementary Fig. 6b (see also Supplementary Note 5).

Other familial relationships discovered in this work are described in Supplementary Notes 7 and 8.

**Genetic differentiation between extant and inferred *C. ryukyuensis* populations.** To measure the genetic differentiation between *C. ryukyuensis* populations, we estimated Weir-Cockerham’s *F*<sub>st</sub> with vcfTools<sup>74</sup> using genomic regions without admixture. For the purpose of comparing progenitor *C. ryukyuensis* populations of shiikuwasha, tachibana, and yukunibu, we extracted *C. ryukyuensis* haplotypes from each group and formed pseudo-diploids before using vcfTools to estimate *F*<sub>st</sub>. As an alternative approach to calculate the genetic differentiation between the extant *C. ryukyuensis* population in Okinawa and the *C. ryukyuensis* ancestors of mainland Japan tachibana population, we compared diploid *C. ryukyuensis* segments of tachibana to the corresponding genomic regions in Okinawa *C. ryukyuensis* accessions (Supplementary Note 11).

**Characterizing apomixis haplotypes.** To investigate the genetic ancestry of the *CitRKD1* MITE-insertion allele and its connection to the mangshanyaju population, we examined the 200 kb region flanking the *CitRKD1* gene (Ciclev10010497m) (chromosome 1: 25,380,489–25,582,037 of the Clementine reference sequence<sup>21</sup>) across a collection of 55 citrus accessions including mandarins and interspecific mandarin hybrids and admixtures derived from six citrus species (PU=pummelo, CI=citron, IC=Ichang papeda, FO=Fortunella, RK=*C. ryukyuensis* and *C. reticulata*). We carried out genetic admixture analysis for this 200 kb window using ADMIXTURE<sup>68</sup> for *K*=3–11. For each value of *K*, we performed twenty independent runs and used the run with lowest cross-validation error. Figure 4b shows the population ancestry composition at *K*=8, with additional plots for *K*=8–10 included in Supplementary Fig. 8c (See Supplementary Note 10). For *K*=8 all six citrus species are resolved with further differentiation of mandarins (*C. reticulata*) into three sub-populations, namely, common mandarin (MA), and two mangshanyaju sub-populations h1 and h2 across this window. Progressively finer resolution is observed for *K*=9 and 10. At *K*=9, two sub-populations (m1, m2) are differentiated within common mandarins. With ten ancestral populations (*K*=10), further differentiation within pummelos (PU, p2) is revealed. Importantly, the two mangshanyaju (h1, h2) sub-population ancestry compositions remain unchanged for *K*=8–10.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. High coverage (average 46x) whole-genome shotgun-sequencing data of 69 citrus accessions generated in this study have been deposited at NCBI under BioProject PRJNA670310, with summary information for each accession in Supplementary Data 1 and 2. Previously published resequencing data used in this study are listed in Supplementary Data 3. Source data are provided with this paper.

## Code availability

Custom wrapper scripts for demographic inference using moments are available at Github [<https://github.com/citruscomp/RyukyuCitrus.git>].

Received: 4 May 2021; Accepted: 25 June 2021;

Published online: 26 July 2021

## References

- Swingle, W. T. & Reece, P. C. in *The citrus industry*, revised 2nd ed., History, World Distribution, Botany, and Varieties Vol. 1 (ed.s W. Reuther, H. J. Webber, & L. D. Batchelor) 190–430 (University of California, Berkeley, California, USA, 1967).
- Tanaka, T. Fundamental discussion of *Citrus* classification. *Stud. Citrologica* **14**, 1–6 (1977).
- USDA Foreign Agricultural Service. Citrus: world markets and trade. <https://apps.fas.usda.gov/psdonline/circulars/citrus.pdf> (2021).
- Wang, L. et al. Genome of wild mandarin and domestication history of mandarin. *Mol. Plant* **11**, 1024–1037 (2018).
- Tanaka, T. A revision of Ryukyu *Rutaceae-Aurantioideae*, Revisio Aurantiacearum X. *Sci bull Agr Home Eco Division. Univ. Ryukyus* **4**, 91–116 (1957).
- Inafuku-Teramoto, S. et al. Local citrus genetic resources and their polymethoxyflavones content in northern part of Okinawa Island. *Hortic. Res. (Japan)* **9**, 263–271 (2010).
- Tanaka, T. Citrus fruits of Japan: with notes on their history and the origin of varieties through bud variation. *J. Heredity* **13**, 243–253 (1922).

8. Ishikawa, R. *et al.* Multi-lineages of Shikuwasha (*Citrus depressa* Hayata) evaluated by using whole chloroplast genome sequences and its bio-diversity in Okinawa, Japan. *Breeding Sci.* 15151 (2016).
9. Omodaka, H. *Man'yōshū with Annotations* p 115 (in Japanese) (Dai Nihon Insatsu, Tokyo, Japan, 1958).
10. Hokama, S., Hateruma, E. *The Standard Text of Omoro Sōshi* p 13–16, p 18 (Kadokawashoten, Tokyo, Japan (in Japanese) 2002).
11. Tanaka, T. Taxonomy of the citrus fruits of the Pacific region. *Mem. Tanaka Citrus Exp. Sta.* 1, 15–36 (1927).
12. Yamamoto, M., Matsuo, Y., Kuniga, T., Matsumoto, R. & Yamada, Y. Isozyme and RAPD analyses of Shikuwashes (*Citrus depressa* Hayata). *Bulletin of the Fruit Tree Research Station (Japan)* (1998).
13. Shimizu, T. *et al.* Hybrid origins of citrus varieties inferred from DNA marker analysis of nuclear and organelle genomes. *PLoS One* 11, e0166969 (2016).
14. Garcia-Lor, A., Luro, F., Ollitrault, P. & Navarro, L. Genetic diversity and population structure analysis of mandarin germplasm by nuclear, chloroplastic and mitochondrial markers. *Tree Genet. Genomes* 11, 1–15 (2015).
15. Parlevliet, J. E. & Cameron, J. W. Evidence on the inheritance of nucellar embryony in *Citrus Proc. Am. Soc. Hortic. Sci.* 74, 252–260 (1959).
16. Iwamasa, M., Ueno, I. & Nishiura, M. Inheritance of nucellar embryony in citrus. *Bull. Hort. Res. Stn.* 7, 1–10 (1967).
17. Wang, X. *et al.* Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* 49, 765 (2017).
18. Shimada, T. *et al.* MITE insertion-dependent expression of CitRKD1 with a RWP-RK domain regulates somatic embryogenesis in citrus nucellar tissues. *BMC Plant Biol.* 18, 166 (2018).
19. Curk, F. *et al.* Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. *Ann. Bot.* 117, 565–583 (2016).
20. Wu, G. A. *et al.* Genomics of the origin and evolution of *Citrus*. *Nature* 554, 311 (2018).
21. Wu, G. A. *et al.* Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* 32, 656–662 (2014).
22. Hirai, M., Mitsue, S., Kita, K. & Kajiura, I. A survey and isozyme analysis of wild mandarin, tachibana (*Citrus tachibana* (Mak.) Tanaka) growing in Japan. *J. Jpn. Soc. Hortic. Sci.* 59, 1–7 (1990).
23. Peng, Z. *et al.* A chromosome-scale reference genome of trifoliolate orange (*Poncirus trifoliata*) provides insights into disease resistance, cold tolerance and genome evolution in *Citrus*. *Plant J.* 104, 1215–1232 (2020).
24. Kimura, M. Paleogeography of the Ryukyu Islands. *Tropics* 10, 5–24 (2000).
25. Kimura, M. The Formation of the Ryukyu Arc and Migration of Biota to the Arc [in Japanese]. (Okinawa Times, Naha, 2002).
26. Nakamura, K. *et al.* Phylogeography of *Ophiorrhiza japonica* (Rubiaceae) in continental islands, the Ryukyu Archipelago, Japan. *J. Biogeogr.* 37, 1907–1918 (2010).
27. Nakamura, K. *et al.* Phylogeny and biogeography of the *Viola iwagawae-tashiroi* species complex (*Violaceae*, section *Plagiostigma*) endemic to the Ryukyu Archipelago, Japan. *Plant Syst. Evolut.* 301, 337–351 (2015).
28. Chiang, T.-Y. & Schaal, B. A. Phylogeography of plants in Taiwan and the Ryukyu Archipelago. *Taxon* 55, 31–41 (2006).
29. Hiramatsu, M., Ii, K., Okubo, H., Huang, K. L. & Huang, C. W. Biogeography and origin of *Lilium longiflorum* and *L. formosanum* (*Liliaceae*) endemic to the Ryukyu Archipelago and Taiwan as determined by allozyme diversity. *Am. J. Bot.* 88, 1230–1239 (2001).
30. Liao, C.-C. & Chen, C.-H. Investigation of floristic similarities between Taiwan and terrestrial ecoregions in Asia using GBIF data. *Bot. Stud.* 58, 15 (2017).
31. Otsuka, H. & Takahashi, A. Pleistocene vertebrate faunas in the Ryukyu Islands: their migration and extinction. *Tropics* 10, 25–40 (2000).
32. Ota, H. Geographic patterns of endemism and speciation in amphibians and reptiles of the Ryukyu Archipelago, Japan, with special reference to their paleogeographical implications. *Res. Popul. Ecol.* 40, 189–204 (1998).
33. Yu, G.-H. *et al.* From mainland to islands: colonization history in the tree frog *Kurixalus* (*Anura: Rhacophoridae*). *Current Zoology* (2020).
34. Yang, S. F., Komaki, S., Brown, R. M. & Lin, S. M. Riding the Kuroshio current: stepping stone dispersal of the Okinawa tree lizard across the East Asian Island Arc. *J. Biogeogr.* 45, 37–50 (2018).
35. Yamamoto, M. *et al.* Diversity of chloroplast DNA in various mandarins (*Citrus spp.*) and other citrus demonstrated by CAPS analysis. *J. Jpn. Soc. Hortic. Sci.* 82, 106–113 (2013).
36. Yamamoto, M., Takakura, A., Tanabe, A., Teramoto, S. & Kita, M. Diversity of *Citrus depressa* Hayata (Shikuwasha) revealed by DNA analysis. *Genet. Resour. Crop Evolut.* 64, 805–814 (2017).
37. Nagano, Y. *et al.* Characterization of chloroplast matK sequences of *Citrus tachibana* and *Citrus depressa*, two indigenous species in Japan. *Adv. Hortic. Sci.* 28 (2014).
38. Tanaka, Y. An Iconography of Japanese Citrus Fruits. Vol. 2 (1948).
39. Yamamoto, M. Citrus genetic resources grown on the Ryukyu Islands, Japan. 南太平洋海域調査研究報告 = Occasional papers, 9–15 (2014).
40. Lam, D. T. & Ishikawa, R. Molecular discrimination of landraces of *Citrus* species in the Okinawa, Japan. *Genet. Resour. Crop Evolut.* 66, 321–333 (2019).
41. Shimizu, T. *et al.* A model for the domestication and diversification processes of modern citrus cultivars in Japan. *Acta Hortic.* 1230, 7–14 (2019).
42. Kajira, I. *A chronological table of fruit history in Japan [in Japanese]*. (Yokendo, Yokyo, 2008).
43. Zhou, X.-H. *et al.* Distribution and evolution of citrus with S 4 and/or S 5 gene alleles for self-incompatibility with special focus on the origin of satsuma mandarin (*Citrus unshiu* Marc.; S f S 4). *Genet. Resour. Crop Evolut.* 65, 1013–1033 (2018).
44. Tanaka, Y. *Developmental history of citrus varieties (1)*. *Kankitsu* 8, 308–315 (1941).
45. Tanaka, Y. *Developmental history of citrus varieties (2)*. *Kankitsu* 9, 5–16 (1942).
46. Ueno, I., Iwamasa, M. & Nishiura, M. Embryo number of various varieties of citrus and its relatives. *Bull. Hort. Res. Sta. Jpn., Ser. B* 7, 11–22 (1967).
47. Strazzer, P. *et al.* Hyperacidification of citrus fruits by a vacuolar proton-pumping P-ATPase complex. *Nat. Commun.* 10, 744 (2019).
48. Butelli, E. *et al.* Noemi controls production of flavonoid pigments and fruit acidity and illustrates the domestication routes of modern citrus varieties. *Curr. Biol.* 29, 158–164 (2019). e152.
49. Butelli, E. *et al.* Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255 (2012).
50. Butelli, E. *et al.* Changes in anthocyanin production during domestication of *Citrus*. *Plant Physiol.* 173, 2225–2242 (2017).
51. Huang, D. *et al.* Subfunctionalization of the *Ruby2-Ruby1* gene cluster during the domestication of citrus. *Nat. Plants* 4, 930–941 (2018).
52. Liang, M. *et al.* Evolution of self-compatibility by a mutant Sm-RNase in citrus. *Nat. Plants* 6, 131–142 (2020).
53. Yu, Y., Chen, C. & Gmitter, F. G. QTL mapping of mandarin (*Citrus reticulata*) fruit characters using high-throughput SNP markers. *Tree Genet. Genomes* 12, 77 (2016).
54. Zheng, X. *et al.* Natural variation in CCD4 promoter underpins species-specific evolution of red coloration in citrus peel. *Mol. Plant* 12, 1294–1307 (2019).
55. Allan, A. C. Domestication: colour and flavour joined by a shared transcription factor. *Curr. Biol.* 29, R57–R59 (2019).
56. Yu, Y. *et al.* Identification of QTLs controlling aroma volatiles using a 'Fortune' × 'Murcott' (*Citrus reticulata*) population. *BMC Genom.* 18, 646 (2017).
57. Nakanishi, H. Dispersal ecology of the maritime plants in the Ryukyu Islands, Japan. *Ecol. Res.* 3, 163–173 (1988).
58. Talon, M., Wu, G. A., Gmitter, F. G. & Rokhsar, D. S. in *The Genus Citrus* (eds M. Talon, M. Caruso, & F. G. Gmitter) 9–31 (Elsevier, 2020).
59. Ollitrault, P., Curk, F. & Krueger, R. in *The Genus Citrus* (eds M. Talon, M. Caruso, & F. G. Gmitter) 57–81 (Elsevier, 2020).
60. Gross, B. L. & Rieseberg, L. H. The ecological genetics of homoploid hybrid speciation. *J. Hered.* 96, 241–252 (2005).
61. Soltis, P. S. & Soltis, D. E. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588 (2009).
62. Burgarella, C. *et al.* Adaptive introgression: an untapped evolutionary mechanism for crop adaptation. *Front. Plant Sci.* 10, 4 (2019).
63. Janzen, G. M., Wang, L. & Hufford, M. B. The extent of adaptive wild introgression in crops. *N. Phytol.* 221, 1279–1288 (2019).
64. Grosser, J., Gmitter, F. & Bowman, K. New rootstocks in the citrus breeding pipeline. *Citrus Ind.* 101, 8–11 (2020).
65. Yamamoto, M., Nishiguchi, N., Shimada, A. & Matsumoto, R. Polymethoxylated flavone content of major cultivars and local accessions of citrus cultivated in Kagoshima, Japan. *Hortic. J.* 88, 320–328 (2019).
66. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).
67. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010).
68. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664 (2009).
69. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).
70. Bauscher, M. G., Singh, N. D., Lee, S. B., Jansen, R. K. & Daniell, H. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol.* 6, 21 (2006).
71. Jouanous, J., Long, W., Ragsdale, A. P. & Gravel, S. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics* 206, 1549–1567 (2017).
72. Xie, S. P., Manchester, S. R., Liu, K. N., Wang, Y. F. & Sun, B. N. *Citrus Linczangensis* Sp. N., a leaf fossil of *Rutaceae* from the Late Miocene of Yunnan, China. *Int. J. Plant Sci.* 174, 1201–1207 (2013).

73. Lee, W. C. Testing the genetic relation between two individuals using a panel of frequency-unknown single nucleotide polymorphisms. *Ann. Hum. Genet.* **67**, 618–619 (2003).
74. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

## Acknowledgements

We are grateful to Hiroshi Kobashigawa for sharing his expertise and hospitality, where the first *C. ryukyuensis* tree was observed, and to Yoshikatsu Yamakawa and Hiromitsu Yasumura from Katsuyama Shiiikuwasha Co., Ltd. for their assistance and providing the Katsuyama mandarin accession related to all shiiikuwasha. We thank Shuji Takino and Hideki Yamashiro for assistance in sample collection. We thank Nate Jameson for contributing his citrus expertise in the field in Okinawa, and John Willis for helpful discussions. We thank staff from the Okinawa Institute of Science and Technology (OIST) DNA sequencing section, Onna, Okinawa for their help with the DNA-sequencing. This study was supported by OIST Internal Funds (D.S.R.). The collaborative efforts of G.A.W., D.S.R., and F.G.G. are partially supported by a grant from the Citrus Research and Development Foundation (18-010), on behalf of the Florida citrus growers. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231 (G.A.W., D.S.R.). The work conducted by the Centro de Genómica at IVIA (Spain) is supported through grants No RTI2018-097790-R-100 (Ministerio de Ciencia, Innovación y Universidades/Fondo Europeo de Desarrollo Regional) and No 51915 (IVIA)(MT). D. S.R. is grateful for the support of the Marthella Foskett Brown Chair in Biological Sciences. We thank the DNA Sequencing Section, and the Information Service Section of OIST for technical support. We thank Shuji Takino for samples of TB0 and TB6 from Hashimoto and Shingu in Shingu City in Wakayama prefecture and for help during the sampling of TB3 and TB4 from Arima in Kumano city in Mie prefecture; Hirofumi Hamada for samples of TB7 from Suno in Kumano City in Mie prefecture; Naoe Ooe for samples of TB5, TB8 and YZN from Nikou in Nachikatsuura town in Wakayama prefecture; Motofumi Yamasaki for samples of YZS and DDS from Shingu in Shingu City in Wakayama prefecture; Hirofumi Higa for samples of SH0, SH1, SH2, SH3, OTO, KBU and TRG from Asahikawa in Nago City in Okinawa prefecture; Okinawa Prefectural Agricultural Research Center (OPARC) of Nago in Nago city in Okinawa prefecture for samples of SH4; Yoshikatsu Yamakawa from Katsuyama Shiiikuwasha Co., Ltd. of Katsuyama in Nago City in Okinawa prefecture for samples of RK3, RK4, RK5, SH5 and SH6; Eijyun Zamami for samples of R15 and ISH from Izumi in Motobu Town in Okinawa prefecture; Masakazu Nakazato for samples of R02 from Katsuyama in Nago City in Okinawa prefecture; Hiroshi Kobashigawa for samples of SHI, RK1, RK2, RK7, RK8, RK9, TB1, TB2 and KB2 from Oogimi in Oogimi Village in Okinawa prefecture; Sachiko Miyagi for samples of RK6 from Oku in Kunigami Village in Okinawa prefecture; Hideki Yamashiro from Katsuyama Community Center on Katsuyama in Nago City in Okinawa prefecture for samples of R01 and R19; Tetsu Yonamine for samples of R07 from Kaneshi in Nakijin Village in Okinawa prefecture; Katsurou Kinjyou from a theme park called 'Daiseikirinzan,' managed by Nanto Co., Ltd. of Ginama in Kunigami Village in Okinawa prefecture for samples of R05, R06 and R12; Kaori Miyazato for samples of R13 from Genka in Nago City in Okinawa prefecture; Yukihiko Shimabukuro for samples of R14 from Imadomari in Nakijin Village in Okinawa prefecture; Hideaki Oshiro for samples of R16, R17 and R18 from Katsuyama in Nago City in Okinawa prefecture; Chogi Miyagi for samples of R03 from Hama in Kunigami Village in Okinawa prefecture; Kiyoshi Miyagi for samples of R04 from Nuha in Oogimi Village in Okinawa prefecture; Noboru Teruya for samples of R00 from Tokijin in Nakijin Village in

Okinawa prefecture; Tetsuo Shimabukuro for samples of KB1 from Inamine in Nago City in Okinawa prefecture; Tomohiro Chinen from a theme park called 'Okinawa World,' managed by Nanto Co., Ltd. of Tamashiro in Nanjo City in Okinawa prefecture for samples of KB3, CAR, C10, C11, C12, OBN, TK0 and TK1.

## Author contributions

G.A.W., F.G.G. Jr, and D.S.R. conceived and led the project with C.S. and M.T. C.S. led and performed sampling, sequencing, and historical analysis. G.A.W. led and performed sequence analysis with input from D.S.R. and consultation with M.T. and F.G.G. Jr. C.S., H.K. and C.A. carried out historical and cultural background analysis. C.S., H.K., M.T. and F.G.G. Jr. contributed to the description of citrus accessions and discriminatory characteristics. M.T. and G.A.W. led the discussion of timing and dispersal of mandarins and consequences for citrus domestication in discussion with F.G.G. Jr. and D.S.R., C.S. and C.A. were responsible for logistics and communication with growers and sources. H.K. contributed samples. F.M. contributed seedless shiiikuwasha sample. G.A.W., C.S., M.T., F.G.G. Jr., and D.S.R. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24653-0>.

**Correspondence** and requests for materials should be addressed to G.A.W., F.G.G.Jr. or D.S.R.

**Peer review information** *Nature Communications* thanks Concetta Licciardello, Robert VanBuren, and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021