

understanding that an overall re-scaling factor (“fudge factor”) around one is implied in the speciation time estimates. As will be shown in **Supplementary Note 9**, comparison with multiple geological events suggests that the assumed citrus crown age of 8 Ma is quite reasonable and consistent with earlier estimates^{12,19}.

The *chronos* program from the R package *ape*³² was used for speciation time inference, with the parameter setting `model='correlated'` and `lambda=1`. The 95% confidence intervals for the speciation times are obtained from 200 bootstrap replicates generated with `seqboot` of *phylip*³³. The resulting chronogram is shown in **Fig. 1c**. With the citrus root age fixed at 8 Ma, the *Poncirus-Citrus* split is estimated at 9.1-9.2 Ma.

Supplementary Note 8.2 Distinct epochs of speciation for Asian and Australian citrus

The nuclear genome chronogram reveals two well-separated phases of species radiation associated with the diversification of Asian and Australian citrus respectively (**Fig. 1c**). With our sampled accessions, the Asian radiation (8-6 Ma) occurred in late Miocene and spans a period of 2 million years. It generated seven citrus species (*C. mangshanensis*, *C. ichangensis*, *C. micrantha*, *C. medica*, *C. maxima*, *C. reticulata* and *Fortunella margarita*) as well as the ancestor species that later diversified into the Australian limes.

The second phase of citrus speciation consists of the diversification of the Australian limes during early Pliocene. It spans the period 4.0-4.5 Ma, separated from the Asian citrus radiation epoch by about 1.5 Myr.

Both the nuclear genome phylogeny and the timing of speciation events point to an Asian origin for citrus, with Australian limes evolved later from an ancient citrus of Asian origin. The chloroplast genome phylogeny, though different from the nuclear genome phylogeny, is also consistent with an Asian origin for citrus. Both reject the previously proposed Australian origin³⁴.

Within the *C. reticulata* species, *C. tachibana* split from the mainland Asian mandarins during early Pleistocene (~2 Ma) and evolved in Taiwan, Japan and the Ryukyu Islands as a genetic isolate from the mainland Asian gene pool as revealed by the chloroplast phylogeny (**Extended Data Fig. 1b**), nuclear genome clustering (**Fig. 1a**), and haplotype sharing analysis (**Fig. 3a**). For this reason, although Swingle¹³ and Tanaka¹⁴ assigned Tachibana its own species (*C. tachibana*), we favor designating it instead as a subspecies of mandarin.

Supplementary Note 8.3 Comparison with chloroplast genome tree

Phylogenetic relationships inferred from the nuclear and chloroplast genomes (**Extended Data Fig. 1**) reveal three major discrepancies.

1) In the nuclear genome tree, *Poncirus* is an outgroup of citrus while in the chloroplast phylogeny *Poncirus* resides near the root of the tree. This observation is compatible with the view that *Poncirus* is likely a descendent of an ancient hybrid of citrus with an unknown parent, as suggested before¹².

2) The chloroplast tree clusters citrons with Australian species. By contrast, in the nuclear phylogeny, citrons cluster with pummelos whereas *Fortunella* clusters with Australian limes.

3) The branching of the three Australian species is not the same. In the chloroplast tree, *Microcitrus australasica* is an outgroup to the clade formed by *Eremocitrus glauca* and *Microcitrus australis*, while in the nuclear tree both *Microcitrus* (*australis* and *australasica*) cluster together with *E. glauca* as an outgroup.

Both nuclear and chloroplast phylogenetic trees are statistically highly supported indicating that inconsistencies are essentially biological and not due to stochastic errors. Incongruences between trees inferred from nuclear and chloroplast DNA are rather common in plants³⁵ and are explained in terms of convergent evolution, lineage sorting and/or reticulate evolution including horizontal gene transfer, hybrid speciation, introgression and chloroplast capture³⁶⁻⁴¹.

In general, as shown in many studies in both plant and animals^{42,43} nuclear phylogenies and networks agree with morphology-based taxonomy, while relationships inferred from chloroplast and mitochondria are more correlated with geographic proximity. Our data showed that the citrus nuclear genome phylogeny agrees in general terms, with both major morphological citrus characters as presented in **Extended Data Fig. 6** and also with geographic proximity (**Fig. 1d**). However, the chloroplast tree topology is rather incongruent in terms of fruit characteristics and less accurate regarding geographic distribution.

Concordance between citrus morphological characters and the nuclear genome phylogeny includes the presence of low number of fruit loculi and ovules in the clade containing *Fortunella* and Australian desert limes. **Extended Data Fig. 6** also suggests that mandarins do not share many fruit characteristics with *C. ichangensis* though they group together in the chloroplast genome tree.

The grouping of citrons and pummelos as revealed in the nuclear genome phylogeny has not been reported previously. While the relationship between citrons and pummelos is new, proximity of citrons and Australian limes has been reported in practically all chloroplast-derived trees presented previously. In a recent study¹², we reported chloroplast genes with unusually high number of SNPs that may be under positive pressure. A detailed analysis of the chloroplast SNP set shared exclusively by citrons and Australian limes revealed that more than half of these genes may be under positive selection.

On the other hand, citrons and pummelos share several significant morphological characteristics that are rather peculiar in the genus *Citrus* and certainly absent in the Australian limes. Overall, *C. medica* and *C. maxima* show complex floral vascular anatomies with large flowers and ovaries that contain joined stamens and many loculi producing large fruits with yellow or pale yellow peel, very thick rinds, a higher number of segments and larger columellas (**Extended Data Fig. 6**). In contrast, Australian limes exhibit small leaves, fruits and flowers with free stamens and fewer loculi. Major differences can also be found between Australian citrus species and citrons/pummelos in other major vegetative traits since Australian limes show characteristic patterns of dimorphic foliage, with coriaceous strongly veined leaves that are not typical in the rest of *Citrus*. Finally, citrons and pummelos have overlapping geographical distribution, as wild genotypes of these two species are mostly located in regions in close proximity, from India, Bhutan, Bangladesh, Myanmar and Indochina to Yunnan (**Supplementary Table 1**).

Supplementary Note 9. The origin, biogeography, and dispersal of citrus

The information reported in this work, *i.e.*, the mapping of the distribution of the 10 pure citrus species identified in this study, the phylogenetic relationship (**Extended Data Fig 1**) and chronogram (**Fig. 1c**) inferred from the analyses of the whole genome sequences together with the recent description of *Citrus linczangensis* from the late Miocene of Lincang³¹, leads us to propose that the center of origin of citrus was located in Southeast Himalaya, in a region including the eastern area of Assam, northern Myanmar and western Yunnan.

Supplementary Note 9.1. Biogeography of *Citrus* and related genera.

The genus *Citrus* and related genera (*Fortunella*, *Poncirus*, *Eremocitrus*, *Microcitrus*) are of wide distribution in Southeast Asia through northeastern Australia, New Caledonia, Melanesia and western Polynesia⁴.

Wild genotypes of several species of the genus *Citrus* have been reported growing freely around this entire region, although the pattern of distribution varies from species to species. Thus, *C. micrantha* has been reported in the Philippines⁴⁴; *C. ichangensis* in northeastern India, northern Myanmar and central and southwestern China, and Yunnan^{4,44}; *C. medica* in northeastern, central and southern India, Bangladesh, Myanmar, Bhutan and Yunnan⁴⁴⁻⁵⁰; *C. maxima* in Indochina, Malaysia, Yunnan and Hainan^{44,45,50,51}; *Citrus reticulata* (mandarins) are present in a wide area from northeastern India to southern and Southeast China^{4,46,52}; Sun Chu Sha Kat mandarin (*Citrus reticulata*) is naturally found in Assam, China and Japan⁴ while *Citrus tachibana* is widespread in southern Taiwan, the Ryukyu Islands and southern Japan⁵³ and *C. mangshanensis* in Hunan, China⁵⁴. In addition, wild hybrids between the species of the genus *Citrus* have also been found mostly in areas and habitats occupied by the parental genotypes. Thus, *C. x limonia* (*C. reticulata* x *C.*

medica) has been reported in eastern Guangxi but mostly in southern Tibet^{48-50,55} while *Citrus x jambhiri* (*C. reticulata* x *C. medica*) has been found growing in a wild condition in India⁵⁶. Sour orange, *C. x aurantium* (*C. maxima* x *C. reticulata*), is found in Nepal, northeast India, Garwal and Sikkim^{45,46}. Calamondin (*Fortunella sp* x *Citrus reticulata*) is widely cultivated in the Philippines and also in China.

Regarding the relevant cultivated hybrids such as *C. x sinensis* (sweet oranges), the presence of wild trees has been reported in tropical forests of northern Myanmar and in the Khasi Mountains in Assam⁵⁷. However, there are no reports on wild *C. x aurantifolia* (limes, *C. micrantha* x *C. medica*) and *C. x limon* (lemons, *C. x aurantium* x *C. medica*)⁴, although there is a general agreement that limes come from the Southeast Asian archipelago⁴ while lemons are native of India and northern Myanmar^{4,58}.

The genus *Fortunella* includes four species, all of them found south of Yangtze River in provinces such as Yunnan or Guizhou^{48,49}. The single *Eremocitrus* species, (*E. glauca*) and 5 of the 6 described species of *Microcitrus*, (including *M. australis* and *M. australasica*) are thought to be native of Queensland and northern New South Wales in eastern Australia^{4,59}. Regarding the 2 species of *Poncirus*, *P. trifoliata*, is native to eastern Asia^{60,61} while *P. polyandra* was identified in Yunnan Province⁶².

Supplementary Note 9.2. Citrus dating and fossils.

Previous estimations of citrus divergence times using molecular analyses have produced different results. Initial analyses based on partial chloroplast sequences estimated the age of citrus to be 22-18 Ma⁶³ while other more solid studies reported that citrus appeared about 7.0 Ma¹⁹. Using whole genome chloroplast sequences, we recently dated the emergence of citrus as *ca.* 8.0 Ma¹².

The first report of a citrus fossil (*C. meletensis*) concluded that the specimen was dated in the Pliocene of Europe and therefore outside of its native geographic distribution⁶⁴. While this finding is still not properly contextualized, the recent report on the identification of a new *Citrus* species, *Citrus linczangensis* sp. n., from the late Miocene (11.6-5.3 Ma) of Lincang³¹, provides definite evidence for the existence of *Citrus* within the province of Yunnan (China) since *ca.* 8 Ma.

The fossil leaves of *C. linczangensis*, which do not exactly resemble any particular extant citrus species in all characters, combine a number of features that are present in distinct phylogenetic clades as defined in **Extended Data Fig. 6**. These characteristics include the possession of an articulated, subcordate and broadly winged petiole as usually seen in *C. maxima*, some species of *Fortunella*, *C. micrantha* and in *C. ichangensis*. The fossil also exhibits intramarginal venation as found in *C. reticulata* and *C. aurantifolia* (*C. micrantha* x *C. medica*) and an entire margin as observed in *C. maxima*. The presence of a long winged petiole has traditionally been regarded as an old trait of ancestral citrus⁴.

The specimen, therefore, might well represent a common ancestor of the major citrus groups (papedas, pummelos, mandarins, *Fortunella* and *Micrantha*). This finding implies that southwestern China⁵⁰ in Late Miocene was a native habitat of citrus and therefore a potential region of early diversification, supporting the hypothesis of a Southeastern Asian origin of *Citrus*.

Supplementary Note 9.3. The center of origin of citrus

The center of origin of citrus has been matter of dispute during almost a century. The elegant but general pioneer visions of Vavilov⁶⁵ identified two centers of origin for citrus, the Indo-Burma center (Assam and Burma; oranges, mandarins and citrons), and the Indo-Malayan center (Indochina and the Malay Archipelago; pummelo). This vision was reformulated by Tanaka,¹⁴ who in subsequent work concluded that the primary center was placed within northeast India and northern Myanmar, from where citrus dispersed to a secondary center located in Indochina and Southeast China. The center of origin of citrus was also suggested to be in northeastern India and in the mountainous parts of southern China⁶⁶.

In their masterful review, Swingle and Reece⁴ indicated that *Citrus* is native to an extensive barrel-shaped area that has its long axis slanting from the northwest (northeastern India to north-central China) to the Southeast (east-central Australia to New Caledonia). Many other propositions have been advanced in more recent years suggesting e.g. that primary centers of citrus origin were in the southwestern mountains of China⁴⁸, in Yunnan and adjacent areas in northern east India, northern Myanmar and southwestern and southern China⁵⁰.

The analyses of the native habitats of citrus presented in this work (**Extended Data Fig. 1a**) based on documented reports (**Supplementary Table 1**) upon the presence of wild genotypes growing freely in non-cultivated areas reveals that the triangle limited by Eastern India, Northern Myanmar and Western Yunnan concentrated the highest number of wild citrus genotypes (*C. medica*, *C. maxima*, *C. reticulata*, *C. ichangensis*, *C. x limonia*, *C. x sinensis*, *C. x aurantium*, *C. x limon*). The identification of *C. linczangensis*³¹, on the other hand, provides definite evidence that in Late Miocene western Yunnan was a native habitat of citrus and therefore a potential region of early diversification. Furthermore, the phylogenetic relationship and estimation of divergence times (**Fig. 1c**) indicated that the ancestral citrus experienced a relatively fast radiation giving rise to all major citrus species in a period of about 2 Myr, which is consistent with a unique initial area of diversification. While our proposal offers partial support to some of the previous formulations conferring geographic accuracy to the otherwise broad and vague propositions, overall, it clearly precludes ideas based on primary centers located in Australia or nearby islands⁴, Malay Archipelago⁶⁷ or Thailand⁵¹.

Supplementary Note 9.4. Citrus rapid radiation and monsoon weakening

Both nuclear (**Extended Data Fig. 1c**) and chloroplast (**Extended Data Fig. 1b**) phylogenies and the estimation of divergence times (**Fig. 1c**) indicated that citrus underwent a rapid radiation during late Miocene (ca. 8.0-6.0 Ma). Rapid biological radiations in this region have been reported for practically all major groups of organisms^{68,69} including insects, fishes, crabs, amphibians, reptiles, birds, ferns⁷⁰ and plants such as the eudicot genera *Caragana*, *Rheum*, *Pedicularis*, *Saussurea*, *Rhododendron*, *Primula*, *Meconopsis*, *Rhodiola*, and many lineages of gymnosperms, *i.e.*, the conifer genus *Juniperus*. It is widely accepted that climatic oscillations provoked by monsoonal seasonality were the pivotal factors facilitating speciation and diversification in Southeast Asia. However, contrary to what is generally believed, current evidence clearly indicates that monsoons in this region arose at different times and are certainly unrelated to the Tibetan uplift⁷¹.

Monsoon regimes in South and East Asia were probably established in Early Miocene and experienced an extended period of intensification with strong summer monsoons in the Middle Miocene and reached maximum between 18 and 10 Ma⁷². After this phase of monsoon intensification, a period of monsoon weakening started around 10 Ma in East Asia and around 8 Ma in South East Asia^{72,73}. It is accepted that the rapid weakening of the monsoon provoked a sudden and drastic climate transition from wetter conditions to a drier climate with seasonal heavy rains^{72,74}. In Southeast Asia this dramatic alteration caused major biota changes including migration of mammals and a conspicuous substitution of evergreen tall tropical trees by tall grasses, a transition that in southeastern Himalaya took place 8.0-7.0 Ma^{74,75}. Thus, the rapid radiation of citrus is in accordance with the proposed phase of weakening of the East Asian monsoon that occurred around 8.0 Ma in Southeastern Himalaya^{72,73}.

Interestingly, the physiological and phenotypic adaptations of current citrus^{76,77} appear to still carry the signature of that ancient transition from wetter to dryer conditions, since continental citrus that can be defined as mesophytes exhibit intriguing xerophytic adaptations to cope with periods of water stress. Citrus, for instance, possesses efficient waxy coated leaves and fruit peels to reduce water losses; the fruit also develops individual juice sacs to protect water, and the tree shows low photosynthetic and transpiration rates associated with slower growth patterns. Citrus plants also produce lush foliage, and have high chlorophyll content and shallow root systems, as found in tropical understory bushes with lower light availability and poor organic material soils. Citrus also develop a unique spongy fruit albedo to cushion rapid volume alterations produced by sudden water inputs. Taken together, these observations are also compatible with the assumption that ancestral citrus were native of regions with “tropical monsoon or savanna climates” (according to the Köppen-Geiger climate classification⁷⁸) but evolved in a “humid subtropical climate” with more pronounced dry seasons.

Supplementary Note 9.5. Citrus dispersal

After species radiation during the late Miocene, we propose that the prevalent directions of citrus dispersal from the putative center of origin were: west (citrons), northeast (*C. mangshanensis* and *papedas*), east (*Fortunella* and mandarins) and southeast (pummelos and *micrantha*). Support for this proposal comes from studies on citrus biogeography and phylogeny, and on the paleogeography of the region, especially the geological history of Wallacea and Japan. According to this view, all citrus except citrons dispersed in a predominant west to east direction (**Fig. 1d**), as many groups of plants and animals did^{68,69,79}.

Furthermore, there is compelling evidence that the eastern edge of the Himalaya, the Hengduan range, comprising several mountain subranges in western Chinese provinces, including Yunnan, underwent rapid uplift only after the Miocene, reaching maximum elevations shortly before Late Pliocene, *ca* 3.6 Ma^{80,81}. Additional evidence based on the reconstruction of paleovegetation and paleoclimate in the Late Pliocene of west Yunnan indicates that uplift of Gaoligong and Nu Mountains (Hengduan range) and the eastern portion of the Tibetan Plateau (Western Yunnan) must have occurred during or after the late Pliocene^{82,83}. Therefore, the paleogeography of the region is compatible with the suggestion that during the period from Late Miocene to Late Pliocene, there was land continuity between current Northern Myanmar and Western Yunnan, i.e. there were not orogenic barriers blocking potential west to east routes allowing citrus dispersal from the center of origin.

Supplementary Note 9.6. Origin of Australian citrus

Diversification of Australian limes (4.0-4.5 Ma, **Fig. 1c**) was not driven by monsoons. Since these species have developed xerophytic structures and are found native in dry environments in north- and south-east Australia, the most logical explanation is that they adapted in Australia. This implies a migration from continental Southeast Asia to Australia in agreement with early plant botanists who suggested that the predominant dispersal direction across the Wallace's line was west to east⁷⁹.

Furthermore, a significant percentage of Australia's northern tropical flora is clearly derived from Southeast Asia, an idea supported by molecular and geographic analyses⁸⁴⁻⁸⁶ that have concluded that lineages of several angiosperm genera such as *Aglaia*, *Alocasia*, *Begonia*, *Pseuduvaria*, *Neonauclea* and *Uvaria* and some palm lineages displayed dispersal patterns largely consistent with initial diversification in continental Southeast Asia and subsequent dispersal to eastern Malesia or Australia.

There is also a general agreement⁸⁷ that most of western Malesia emerged throughout the Cenozoic while islands and lands east of Wallace's Line elevated above sea level only during the late Miocene and Pliocene. Hall⁸⁷ has indicated that most of Sumatra and Java, Sulawesi, parts of the Banda Arc, and the Moluccas were

elevated above sea level since 5 Ma while Seram and Timor have both emerged in the last 3 Ma. This regional plate reorganization that occurred in Wallacea in the last few million years⁸⁸ apparently provided potential stepping stones allowing plant dispersal to New Guinea and Australia as reported for the genera *Bridelia* which reached Australia *ca.* 2 Ma⁸⁹, two different genera of *Cucurbitaceae*, *Benincasa* and *Neoachmandra* that arrived at Australia from Southeast Asia *ca.* 5 and 1 Ma, respectively⁹⁰ and *Begonia* that dispersed six independent times from continental Asia and western Malesia to Wallacea and New Guinea dating from the late Miocene to the Pleistocene⁸⁴.

Likewise, we propose that the three Australian species studied here had an Asian continental ancestor that dispersed from west to east to reach Australia (**Fig. 1d**). This proposal is compatible with the citrus phylogenetic relationships and the inferred timing of Australian lime diversification (**Fig. 1c**). The proposal is also concordant with the predominant west to east dispersal trend observed in the region and with the spatio-temporal diversification patterns reported in other genera as noted above. Further supporting evidence comes from the paleogeography of the region, especially the geological history of Wallacea which postulates that during Late Miocene onwards extensive land masses and islands emerged providing potential stepping stones and allowing island-hopping dispersal²⁵. Lastly, New Guinea was very likely an intermediate stop on the way to Australia, as all the five known *Microcitrus* species found native in eastern Australia are very probably derived from *M. warburgiana*, a species exclusively found in New Guinea⁴. The paleogeographic reconstructions of this epoch also provide additional support for this suggestion⁸⁷.

Supplementary Note 9.7. Tachibana mandarin dispersal.

It has been well documented that Taiwan, the Ryukyu archipelago and Japan attained their flora and fauna from adjacent mainland through the emergence of land bridges that occurred mostly during the Pleistocene with the expansion of ice sheets^{91,92}, the route that probably was also followed by Tachibana during its migration in the early Pleistocene. According to wide palaeoceanographic evidence, the expansion of glaciers led to drastic reductions in the levels of the South China Sea⁹³, creating land bridges and providing major corridors between the islands. The bridges connecting the mainland with the islands occurred many times throughout the Quaternary⁹⁴⁻⁹⁶, including the period of Tachibana split from mainland Asian mandarins during the early Pleistocene, an epoch characterized by strong glacial maxima⁹⁷.

Supplementary Note 10 Pummelo admixture and citrus fruit size and acidity

Though the domestication process for mandarins and sweet orange is complex, two independent lines of investigation point to a strong connection between citrus domestication and pummelo introgression in the mandarin gene pool. For this

analysis, the recently sequenced Chinese accessions³ are not included because information on fruit size and acidity profile are not available.

Supplementary Note 10.1 Pummelo admixture correlates with fruit size

Using diameter as a measure of fruit size, we observe a strong correlation (Pearson correlation coefficient = 0.88) between pummelo admixture proportion and the fruit sizes of mandarins, oranges and grapefruit (**Extended Data Fig. 5a**). In particular, the two pure mandarins (TBM, SCM) are smaller than any other mandarins with pummelo admixture.

A simple linear regression fit shows outliers on both sides of the regression line. On the side of small fruits are the four acidic mandarins, which are either admixture free (TBM, SCM) or have small amount of pummelo admixture (CLP, SNK). By contrast, Ponkan mandarin (PKM) appears unusually large given its admixture size. These outliers suggest that certain genomic loci could be more significant than others in fruit size determination.

With the addition of two pummelos (LAP, CHP), the correlation between fruit size and pummelo admixture proportion becomes stronger (Pearson correlation coefficient 0.94). However, a polynomial regression of the 2nd degree provides a better fit than simple linear regression (adjusted $R^2=0.92$) (**Extended Data Fig. 5b**). Though the four acidic mandarins and Ponkan mandarin remain as outliers, the general trend in fruit size increase is clear from pure to admixed mandarins, and from oranges, grapefruit to pummelos. As larger fruit size, relative to pure mandarins, is a desirable trait, mandarins with pummelo admixture were probably selected and propagated during domestication.

Supplementary Note 10.2. Genome scan for citrus acidity/palatability association

It is very plausible that one of the pivotal drivers of fruit domestication is palatability. In citrus, palatability requires reduced acidity that is dependent upon citric acid accumulation in the vacuole⁹⁸. Thirty-seven citrus accessions with known acidity profile were divided into two groups for a case-control study of fruit acidity/palatability. *Poncirus* and *Severinia* were excluded from this study because they are not true citrus, while fruits from the recently sequenced Chinese accessions³, eremorange and *C. mangshanensis* were not available for acidity determination. The palatable group (case) was formed by 15 accessions including 9 mandarins, all 4 pummelos, sweet orange and grapefruit (**Supplementary Table 2**). The non-palatable group (control) consists of 4 acidic mandarins (TBM, SCM, CLP, SNK), sour orange, all 4 citrons, lemon and limes, as well as CAL, FOR, ICH, MIC, and 5 Australian limes.

To scan for candidate loci associated with fruit palatability, a genome wide association analysis was performed. Since the sample set was relatively small, GWAS

was used as a preliminary approach to reveal chromosome fragments and loci significantly associated with acidity for further manual evaluation. We used a mixed linear model as implemented in *gemma*⁹⁹ to correct for confounding effects from population stratification and sample relatedness. The quality control for the SNP set followed the criteria of **Supplementary Note 3.1**. In addition, we require a SNP call rate > 90% and minor allele frequency > 5%, resulting in 634,888 common SNPs. We used the standardized relatedness matrix and the likelihood ratio test to estimate P value⁹⁹, and applied the conservative Bonferroni correction at significance level of $\alpha = 0.05$ (i.e. $P < 7.9 \times 10^{-8}$).

The GWAS analysis, shown as Manhattan plot in **Extended Data Fig. 5c**, yielded 24 SNPs that exhibited genome-wide statistical significance (**Supplementary Table 4**). All 24 SNPs were manually examined in each of the 37 samples on a gene-by-gene basis as related to allele identity and functional impact. Using *Severinia* and *Poncirus* to define the ancestral state of an allele, palatability is found to be associated with derived alleles, suggesting that the ancestral citrus species were acidic. The results of the manual inspection are summarized in **Extended Data Table 3**. The table shows that SNPs located on chromosomes 2, 5 and 9 as well as three of the four SNPs on chromosome 1 are not discriminatory for acidic accessions versus the four pummelo accessions of the palatable group, though in the GWAS analysis significant allele frequency difference exists between the case and control groups. Based on these considerations, there remain two highly discriminatory loci for the case and control groups.

At the first locus (chr1:23512067; Target of EGR1 Protein 1, TOE1), the non-palatable accessions are mostly homozygous for the ancestral allele (T/T) with a few accessions being heterozygous (C/T). By contrast, all palatable accessions are homozygous in the derived allele (C/C) with the exception of King mandarin (C/T), an edible late variety that shows a marked delay in citric acid degradation.

The second locus spans a nearly 2 Mb region at the beginning of chromosome 8 (chr 8:0.3-2.2 Mb; **Extended Data Fig. 5d**) where pummelo admixture is present among all palatable mandarins but absent in the acidic mandarins, suggesting that the introgression of pummelo genes may have played a role in the domestication of mandarins. This region contains 15 significant and rather discriminatory SNPs. Among several potentially significant genes in this region, Ciclev10028714 and Ciclev10028121 have plausible relevance in the regulation of acidity. Ciclev10028714, a gene coding for mitochondrial NAD⁺ Isocitrate Dehydrogenase (NAD⁺ IDH) catalyzing the oxidative decarboxylation of isocitrate to α -ketoglutarate, a rate-limiting step in the tricarboxylic acid cycle and therefore in citric acid synthesis¹⁰⁰.

While the SNP specifically pinpointed by GWAS (chr 8: 325527) corresponds to a synonymous substitution, manual inspection of the IDH gene revealed three non-synonymous SNPs (positions chr 8: 324328, 326594, and 326608), two of which are

tri-allelic and did not enter the bi-allelic GWAS analysis (**Extended Data Table 4**). The SNPs at positions 326594 and 326608 correlate quite well with acidity except that the acidic sour orange shares the same genotypes as the palatable mandarins, an observation that may reflect the polygenic nature of acidity regulation. Besides the IDH gene, two variants located 1kb upstream of a glycosyl hydrolase gene (Ciclev10028121) are completely discriminatory for acidic versus the non-acidic accessions. Gene expression or protein activity of these two genes has recently been associated with acidity in citrus^{101,102}.

Supplementary Note 11 Acknowledgements

The authors acknowledge financial support from the following institutions and companies: Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, INIA, (Ministerio de Economía, Industria y Competitividad e Innovación, Spain) through grant # RTA-00071-C06-01 and from the Citrusseq-Citrusgenn consortium (Anecoop S. Coop., Eurosemillas S.A., Fundación CajaMar Valencia, GCM Variedades Vegetales A.I.E., Investigación Citrícola Castellón S.A. and Source Citrus Genesis) (MT); Agriculture and Food Research Initiative Grant no 2013-67013-21110 from the USDA National Institute of Food and Agriculture (MLR); Florida Citrus Production Research Advisory Council (FCPRAC), Florida Department of Agriculture and Consumer Services grant No. 013646, Florida Department of Citrus (FDOC) and Citrus Research and Development Foundation grant No. 71, on behalf of the Florida citrus growers (F.G., D.D.). 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 (GAW,DSR). Help and expertise of Isabel Sanchis and Angel Boix (IVIA) on lab and field tasks are also gratefully acknowledged.

Supplementary Tables

Supplementary Table 1. Biogeographic distribution of the genus *Citrus* in Southeast Asia and Australia. The occurrence of wild genotypes of pure citrus species, admixtures and relevant interspecific hybrids is reported. Classical taxonomy (Swingle¹³ and Tanaka¹⁴) has been maintained for clarity. Numbers indicate approximate locations in Extended Data Figure 1a.

Genera/Species	Geographical Region*	Report	Phenotype	References
Pure Citrus Species				
<i>Fortunella sp.</i>	1. Southern China 2. South of Yangtze river	Wild species	Freeze-hardy. Small fruit, acid pulp, edible skin.	Zhang, 1981 ⁴⁸ ; He et al. 1984 ⁴⁹ ; Zhou, 1990 ⁴⁴
Australian citrus	3. Eastern Australia	Wild <i>Eremocitrus glauca</i> <i>Microcitrus australasica</i> <i>Microcitrus australis</i>	Two ovules per locule (only <i>E.</i>) All tree, xerophytic adaptations, small leaves, fruits and flowers	Swingle and Reece, 1967 ⁴ ; Zhou, 1990 ⁴⁴ ; Sykes, 1997 ⁵¹
<i>Citrus micrantha</i>	4. Philippines	Wild <i>C. micrantha</i>	Small acid and bitter fruit. Shrubby tree.	Zhou, 1990 ⁴⁴ ; Ghosh, 1997 ¹⁰³
<i>Citrus ichangensis</i>	5. Northeastern India 6. Northern Myanmar 7. Central China 8. Southwestern China ⁵⁰	Wild <i>C. ichangensis</i>	Non-edible acid fruit. Cold-resistant.	Zhou, 1990 ⁴⁴ ; Gmitter and Hu, 1990 ⁵⁰ Ghosh, 1997 ¹⁰³ ; Swingle and Reece, 1967 ⁴
<i>Citrus medica</i>	9. Northeastern India	Wild species	Large acid fruit, persistent style.	De Candolle, 1883 ⁴⁵ ; Tanaka, 1961 ⁴⁶ ; Rajput

	10. Central India 11. Southern India 12. Bangladesh. 13. Myanmar 14. Bhutan 15. Yunnan.		Thick peel. Tendency to grow continuously.	and Haribabu, 1985 ⁴⁷ ; Zhang, 1981 ⁴⁸ ; He et al. 1984 ⁴⁷ ; Zhou, 1990 ⁴⁴ ; Gmitter and Hu, 1990 ⁵⁰ ; Yang et al. 2015 ¹⁰⁴
<i>Citrus maxima</i>	16. Indochina 17. Malaysia. 18. Yunnan 19. Hainan	Wild species	Large leaves, flowers and fruits. Thick fruit peel. Sweet or acidic.	De Candolle, 1883 ⁴⁵ ; Zhou, 1990 ⁴⁴ ; Gmitter and Hu, 1990 ⁵⁰ ; Scora 1988 ⁵¹
<i>Citrus reticulata</i>	20. Assam 21. China 22. Japan 23. Taiwan 24. Ryukyu Islands 25. Japan	Sun Chu Sha Kat Tachibana (<i>C. tachibana</i>)	Small-fruited, acidic-sweet, seedy deep red, mandarin. Small-fruited, acidic-sweet, seedy yellowish mandarin	Swingle and Reece, 1967 ⁴ Tanaka 1931 ⁵³
<i>Citrus mangshanensis</i>	26. Hunan	Wild <i>C. mangshanensis</i>	Similar to cultivated mandarins.	He et al. 1984 ⁴⁹ ; 1988 ¹⁰⁵
Citrus Admixtures				
<i>Citrus sp. (mandarins)</i>	27. Northeastern India 28. Southern China 29. Southeast China	Wild species of sp. mandarins	<i>Reticulata/maxima</i> admixtures. Acidic, acidic-sweet or palatable.	Tanaka 1961 ⁴⁶ ; Hodgson, 1967 ⁵²
<i>Citrus sinensis</i>	30. Northern Myanmar 31. Assam 32. Southern China	Probably feral <i>C. sinensis</i>	<i>Maxima/reticulata</i> admixtures Moderate freeze-hardy. Fruit of medium size. Palatable.	Swingle and Reece, 1967 ⁴ ; Cooper, 1990 ⁵⁷ ;

33. Indochina

<i>Citrus limon</i>	34. India	Wild specimens	<i>Maxima/reticulata/medica</i> admixture. Acidic fruits.	Bonavia, 1888 ⁵⁸ ; Swingle and Reece, 1967 ⁴ ; Webber et al. 1967 ⁶⁷
	35. Northern Myanmar	not found		

Interspecific Citrus Hybrids

<i>Citrus limonia</i>	36. Eastern Guangxi	Wild species	<i>Reticulata x medica</i> . Tolerant to stresses. Tart and acidic fruits.	Zhang, 1981 ⁴⁸ ; He et al. 1984 ⁴⁹ ; Gmitter and Hu, 1990 ⁵⁰ ; Chen, 1997 ¹⁰⁶	
	37. Southern Tibet				
<i>Citrus jambhiri</i>	38. India	Wild species	<i>Reticulata x medica</i> . Tolerant to stresses. Tart and acidic fruits.	Hodgson, 1937 ⁵⁶	
<i>Citrus x aurantium</i>	39. Nepal	Wild species	<i>Maxima x reticulata</i> . Tolerant to stresses. Tart and acidic fruits.	De Candolle, 1883 ⁴⁵ ; Tanaka, 1961 ⁴⁶	
	40. Northeast India				
	41. Garwal				
	42. Sikkim				
<i>Citrus aurantifolia</i>	43. Southeast archipelago	Asian	Wild specimens not found	<i>Micrantha x citron</i> . Freeze-sensitive. Mostly acidic fruits.	Swingle and Reece, 1967 ⁴
<i>Fortunella reticulata</i>	x C	44. China	Wild specimens not found	<i>Fortunella</i> sp x <i>reticulata</i> Small fruit, acid pulp, edible skin.	Swingle and Reece, 1967 ⁴
		45. Philippines	not found		

Supplementary Table 2. Accessions of the genus *Citrus* and related genera studied in this work. When possible, the Swingle (1943)¹³ and Tanaka (1954)¹⁴ taxonomic systems have been followed.

Code	Common name	Species name	Nuclear genome relationship (mat x pat)	Chloroplast Genome	Acidity
•COR	Corsican citron	<i>C. medica</i> L.	Pure citron	CI	Acidless
•VEU	Mac Veu citron	<i>C. medica</i> L. <i>Citrus lumia</i> Risso & Poit .	Pure citron	CI	Acidic
•BUD	Buddha's hand citron var. <i>Sarcodactylus</i>	<i>C. medica</i> L. (Noot.) Swingle	Pure citron	CI	Acidic
•HUM	Humpang citron	<i>C. medica</i> L.	Pure citron	CI	Acidic
•SCM	Sun Chu Sha Kat mandarin	<i>C. reticulata</i> (Blanco) <i>C. reticulata</i> var. <i>austera</i> (Swingle) <i>C. erythrosa</i> (Tanaka)	Pure mandarin	MA	Acidic
•TBM	Tachibana mandarin	<i>C. tachibana</i> (Mak.) Tan <i>C. reticulata</i> (Blanco)	Pure mandarin	MA	Acidic
•SNK	Sunki mandarin (sour mandarin, suanju)	<i>C. sunki</i> (Hayata, Hort. ex Tanaka <i>C. reticulata</i> (Blanco)	Reticulata/pummelo admixture (low % pummelo)	MA	Acidic
•CLP	Cleopatra mandarin	<i>C. reshni</i> (Hort. ex Tanaka) <i>C. reticulata</i> (Blanco)	Reticulata/pummelo admixture (low % pummelo)	MA	Acidic
•CSM	Changsha mandarin	<i>Citrus reticulata</i> (Blanco)	Reticulata/pummelo admixture	MA	Non-acidic
HLM	Huanglingmiao mandarin	<i>C. *reticulata</i> (Hort. ex Tanaka)	Reticulata/pummelo admixture (parent of UNS)	MA	Non-acidic
•KSH	a.k.a Mukakukishu Kinokuni mandarin	or <i>C. *kinokuni</i> (Hort. ex Tanaka)	Reticulata/pummelo admixture (mutant of Huanglingmiao)	MA	Non-acidic

•UNS	Satsuma (unshiu) mandarin, cv. Owari	<i>C. * unshiu</i> [(Mak.) Marc] <i>C. *reticulata</i> (Swingle)	Reticulata/pummelo admixture	MA	Non-acidic
PKM	Ponkan (Chinese honey orange)	<i>C. *reticulata</i> (Blanco, Swingle)	Reticulata/pummelo admixture (parent of DNC)	MA	Non-acidic
•DNC	Dancy mandarin, Dancy tangerine	<i>C. * tangerina</i> (Tanaka) <i>C. *reticulata</i> (Swingle)	Reticulata/pummelo admixture	MA	Non-acidic
WLM	Willowleaf mandarin	<i>C. x deliciosa</i> (Ten. Hort. ex Tanaka)	Reticulata/pummelo admixture (maternal parent of CLM)	MA	Non-acidic
CLM	Clementine mandarin, cv. Clementina de Nules	<i>C. x clementina</i> (Hort. ex Tanaka) <i>C.x reticulata</i> (Swingle)	Reticulata/pummelo admixture WLM x SWO	MA	Non-acidic
•KNG	King mandarin	<i>C. * nobilis</i> (Lour.) <i>C. *reticulata</i> (Swingle)	Reticulata/pummelo admixture	MA	Non-acidic
WMM	W. Murcott mandarin	<i>C. x reticulata</i> (Blanco)	Reticulata/pummelo admixture	MA	Non-acidic
•LMA	Rangpur lime	<i>C. x limonia</i> (Osbeck)	Reticulata x citron hybrid	MA	Acidic
•RRL	Red rough lemon	<i>C x jambhiri</i> (Lush)	Reticulata x citron hybrid	MA	Acidic
LAP	Low acid pummelo (Siamese Sweet)	<i>C. maxima</i> [(Burm.) Merr]. <i>C. grandis</i> (Swingle, Tanaka)	Pure pummelo (maternal parent of CHP)	PU	Acidless
CHP	Chandler pummelo	<i>C. maxima</i> [(Burm.) Merr]. <i>C. grandis</i> (Swingle, Tanaka)	Pure pummelo (Siamese Sweet (LAP) x Siamese Pink)	PU	Non-acidic
GXP	Guan-xi-mi-you pummelo	<i>C. maxima</i> [(Burm.) Merr]. <i>C. grandis</i> (Swingle, Tanaka)	Pure pummelo	PU	Non-acidic
STP	Sha-tian-you pummelo	<i>C. maxima</i> [(Burm.) Merr]. <i>C. grandis</i> (Swingle, Tanaka)	Pure pummelo	PU	Non-acidic
•PAR	Grapefruit, cv. Marsh	<i>C. x paradisi</i> (Macfadyen)	Pummelo x SWO	PU	Non-acidic

SWO	Sweet orange, cv. Washington Navel	<i>C. x sinensis</i> L. (Osbeck)	Pummelo/reticulata admixture (paternal parent of CLM)	PU	Non-acidic
SSO	Sour orange, cv. Seville	<i>C. x aurantium</i> L.	Pummelo x reticulata hybrid	PU	Acidic
•LIM	Lemon, cv. Eureka	<i>C. x limon</i> L. (Burm. f.)	SSO x citron hybrid	PU	Acidic
•ADR	Australian desert lime	<i>Eremocitrus glauca</i> (Lindl.) Swingle <i>C. * glauca</i> (Lindl.) Burkill <i>Eremocitrus glauca</i> (Lindl.) Swingle	Pure glauca	AU	Acidic
•ADL	Eremorange. Australian desert lime hybrid	N.A.	<i>Eremocitrus glauca</i> x <i>Citrus sinensis</i>	AU	---
•AFR	Australian finger lime	<i>Microcitrus australasica</i> (F. Muell.) Swingle <i>C. australasica</i> F. Muell.	Pure australasica	AU	Acidic
•AFL	Australian finger lime	<i>Microcitrus australasica</i> (F. Muell.) Swingle <i>C. australasica</i> F. Muell.	BC2 backcross	AU	Acidic
•ARR	Australian round lime	<i>Microcitrus australis</i> Swingle <i>C. australis</i> (A. Cunn. ex Mudie)	Pure australis	AU	Acidic
•ARL	Australian round lime	<i>Microcitrus australis</i> Swingle <i>C. australis</i> (A. Cunn. ex Mudie)	Pure australis	AU	Acidic
•FOR	Kumquat, Nagami	<i>Fortunella margarita</i> (Lour.) Swingle	Pure Fortunella	FO	Acidic
•CAL	Calamondin	<i>Citrus reticulata</i> Blanco var. <i>austera</i> Swingle ? x <i>Fortunella</i> sp. ? <i>C. madurensis</i> (Lour.)	Fortunella x mandarin hybrid	FO	Acidic
•MXL	Mexican lime	<i>C. * aurantifolia</i> (Christm.) Swingle	Micrantha x citron hybrid	MC	Acidic
•MIC	Micrantha, Biasong	<i>C. micrantha</i> (Wester)	Pure micrantha	MC	Acidic
CMS	Mangshan mandarin	<i>C. mangshanensis</i>	Pure mangshan	MS	---

•ICH	Ichang papeda	<i>C. ichangensis</i> (Swingle)	Pure ichangensis	IC	Acidic
•PON	Trifoliate orange	<i>Poncirus trifoliata</i> (L.) Raf.	Pure Poncirus	PT	Acidic
•SVR	Chinese box orange	<i>Severinia buxifolia</i> (Poir.) Tenore		SV	---
M01	Mandarin from south China (CYY)	<i>C. reticulata</i>	Pure mandarin	MA	---
M02	Mandarin from south China (HZ)	<i>C. reticulata</i>	Pure mandarin	MA	---
M03	Mandarin from south China (SJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M04	Mandarin from south China (SPG)	<i>C. reticulata</i>	Pure mandarin	MA	---
M08	Mandarin from Zhejiang province (20H)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M10	Mandarin from south China (HPJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M11	Mandarin from south China (YSJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M12	Mandarin from south China (NJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M14	Mandarin from south China (MSJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M15	Mandarin from south China (LYJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M16	Mandarin from south China (BTJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---
M17	Mandarin from south China (STJ)	<i>C. *reticulata admixture</i>	Admixed mandarin	MA	---

M19	Wilking (WLK)	<i>C. x reticulata</i>	King x willowleaf	MA	Non-acidic
M20	Kiyomi (KYM)	<i>C. x reticulata</i>	Satsuma x sweet orange	MA	Non-acidic
M21	Fallglo (QH117)	<i>C. x reticulata</i>	(Clementine x Orlando) x Temple	MA	Non-acidic
GF0	Cocktail grapefruit (14J)	<i>not assigned by Swingle or Tanaka.</i>	Low acid pummelo x Frua mandarin	PU	Non-acidic
SO5	Ambersweet orange (A20)	<i>not assigned by Swingle or Tanaka</i>	(Clementine x Orlando) x sweet orange	MA	Non-acidic
BO2	Sour orange from south China (CBSC)	--	Pummelo x pure mandarin	PU	---
BO3	Sour orange from south China (ZGSC)	--	--	PU	---

• = Genomes sequenced in this work; otherwise genomes were reanalyzed from original published sequences¹⁻³.

X = Hybrid origin previously known; * = Information generated in this work.

Chloroplast types: AU, Australian limes; CI, citron; FO, *Fortunella*; MA, mandarin; MC, *Micrantha*; MS, *Citrus mangshanensis*; IC, Ichang papeda; PT, *Poncirus*; PU, pummelo; SV, *Severinia*.

Non-acidic: palatable citrus fruit; usually with acid content < 10 g/l when overripe. Acidic: unpalatable citrus fruit; acid content > 20g/l when overripe.

Acidless: applies to Corsican citron and low acid pummelo, two varieties with reduced acid content < 2g/l.

Note that the last 19 accessions (beginning with M01) are from Wang et al.³ with their codes in parenthesis.

Supplementary Table 3. Sequencing statistics of the 30 new genomes reported in this work. When possible, the Swingle (1943)¹³ and Tanaka (1954)¹⁴ taxonomic systems have been followed.

Code	Common name	Species name	Total Reads	Coverage
CLP	Cleopatra mandarin	<i>C. x reshni</i> (Hort. ex Tanaka)	378265813	126X
KNG	King mandarin	<i>C. x nobilis</i> (Lour)	194708238	65X
DNC	Dancy mandarin	<i>C. x tangerina</i> (Tanaka)	180698706	60X
UNS	Satsuma mandarin	<i>C. x unshiu</i> (Marc)	188558275	63X
PAR	Grapefruit cv. Marsh	<i>C. x paradisi</i> (Macfadyen)	534428187	178X
MXL	Mexican lime	<i>C. x aurantifolia</i> (Christm.) Swingle	160670087	53X
LMA	Rangpur lime	<i>C. x limonia</i> (Osbeck)	177312373	59X
LIM	Lemon cv. Eureka	<i>C. x limon</i> L. (Burm. F.)	341041115	114X
PON	Trifoliolate orange	<i>Poncirus trifoliata</i> (L.) Raf.	410744971	137X
FOR	Kumquat, Nagami	<i>Fortunella margarita</i> (Lour.)	38133267	13X
CAL	Calamondin	<i>C. x madurensis</i> (Lour.)	47734506	16X
SVR	Chinese box orange	<i>Severinia buxifolia</i> (Poir)	145141150	48X
MIC	Micrantha, Biasong	<i>C. micrantha</i> (Wester)	169407177	56X
COR	Corsican citron	<i>C. medica</i> L.	59313741	20X
VEU	Mac Veu citron	<i>C. medica</i> L.	188510474	63X
SNK	Sunki mandarin	<i>C. x sunki</i> (hayata)	189645086	63X
ICH	Ichang papeda	<i>C. ichangensis</i> (Swingle)	187957308	17X
BUD	Buddha's hand citron	<i>C. medica</i> L.	182240696	62X
ADL	Eremorange	<i>Eremocitrus glauca</i> x <i>C. sinensis</i>	178340972	59X
HUM	Humpang citron	<i>C. medica</i> L.	208654674	69X
AFL	Australian finger lime	<i>Microcitrus australasica</i> (F Muell.)	203376652	68X
ARL	Australian round lime	<i>Microcitrus australis</i> Swingle	205354420	68X
ADR	Australian desert lime	<i>Eremocitrus glauca</i>	17607283	9.2X
AFR	Australian finger lime	<i>Microcitrus australasica</i> (F Muell.)	74982998	39X
ARR	Australian round lime	<i>Microcitrus australis</i> Swingle	48391420	25X
CSM	Changsha mandarin	<i>C. x reticulata</i> (Blanco)	186961315	46X
RRL	Red rough lemon	<i>C. x jambhiri</i> (Lush)	117836928	68X
SCM	Sun Chu Sha Kat	<i>C. reticulata</i> (Blanco)	72366584	38X
TBM	Tachibana mandarin	<i>C. tachibana</i> (Mak)	61787150	32X
KSH	Kishu mandarin	<i>C. x kinokuni</i> (Hort. ex Tanaka)	62761835	33x

Supplementary Table 4. Candidate SNPs associated with citrus acidity/palatability. The association study is based on a case-control GWAS analysis of n=37 accessions with known palatability, with conservative Bonferroni correction ($P=7.9 \times 10^{-8}$) at $\alpha=0.05$ significance level. SNPs in intergenic regions are labeled upstream (ups.) or downstream (downs.) of a gene if they are located within 1 kb from a neighboring gene.

Chr	Position	Ref:Alt	Beta	s.e.(beta)	P	Gene (region)	Annotation
1	415175	C:T	4.07E-01	5.24E-02	1.11E-09	Ciclev10008736 (intron)	COP9 SIGNALOSOME COMPLEX SUBUNIT 5
1	23512067	C:T	4.96E-01	3.55E-02	2.96E-08	Ciclev10007611 (exon)	TARGET OF EGR1 PROTEIN 1 (TOE 1)
1	23679916	C:A	-6.52E-01	9.14E-02	8.31E-09	Ciclev10007740 (intron)	RIBOSOMAL RNA METHYLTRANSFERASE NOP2-RELATED
1	24219222	A:G	-6.52E-01	9.14E-02	8.31E-09	Ciclev10010250 (downs.)	PTHR23155//PTHR23155:SF563 - LEUCINE-RICH REPEAT-CONTAINING PROTEIN
2	15484525	G:T	6.66E-01	1.03E-01	7.37E-08	Ciclev10015371 (intron)	S-ALKYL-THIOHYDROXIMATE LYASE SUR1-RELATED
2	15702160	G:T	5.27E-01	7.81E-02	2.85E-08	Ciclev10014095 (intron)	Phospholipid-translocating ATPase
5	35094706	G:A	3.87E-01	5.95E-02	6.30E-08	Ciclev10000105 (intron)	PTHR15245:SF20 - SYMPLEKIN
5	35098538	G:A	3.88E-01	5.98E-02	6.35E-08	Ciclev10000105 (intron)	PTHR15245:SF20 - SYMPLEKIN
8	325527	A:G	5.78E-01	8.66E-02	3.56E-08	Ciclev10028714 (exon)	NAD+-ISOCITRATE DEHYDROGENASE (IDH)
8	631678	T:C	6.48E-01	7.61E-02	1.13E-10	Ciclev10030330 (ups.)	PTHR31175:SF1 - SAUR-LIKE AUXIN-RESPONSIVE PROTEIN-RELATED
8	927020	C:T	6.61E-01	8.23E-02	5.05E-10	Ciclev10028228 (intron)	PEARLI 4
8	1149577	G:T	4.37E-01	6.72E-02	6.05E-08	Ciclev10028121 (ups.)	PTHR31490:SF3 - GLYCOSYL HYDROLASE FAMILY 10 PROTEIN
8	1149586	C:T	4.67E-01	6.31E-02	3.49E-09	Ciclev10028121 (ups.)	PTHR31490:SF3 - GLYCOSYL HYDROLASE FAMILY 10 PROTEIN
8	1174414	T:A	5.15E-01	6.89E-02	4.44E-09	Ciclev10028271 (ups.)	PTHR22950//PTHR22950:SF242 - AMINO ACID TRANSPORTER
8	1413967	A:G	4.66E-01	7.15E-02	5.81E-08	Ciclev10030436 (intron)	PROLINE IMINOPEPTIDASE
8	1651338	G:A	5.04E-01	7.53E-02	3.39E-08	Ciclev10027661 (exon)	SERINE/THREONINE-PROTEIN KINASE (MTOR)
8	1655701	G:T	5.04E-01	7.53E-02	3.39E-08	Ciclev10027661 (exon)	SERINE/THREONINE-PROTEIN KINASE (MTOR)
8	1722788	T:C	4.42E-01	6.77E-02	5.67E-08	Ciclev10027741 (intron)	PRE-MRNA-PROCESSING PROTEIN PRP40
8	2058824	C:T	4.24E-01	6.52E-02	6.15E-08	Ciclev10027948 (exon)	DNA POLYMERASE KAPPA
8	2060290	T:C	4.76E-01	6.18E-02	1.43E-09	Ciclev10027948 (intron)	DNA POLYMERASE KAPPA
8	2063416	T:C	4.49E-01	6.55E-02	2.02E-08	Ciclev10029201 (3'-UTR)	UNKNOWN
8	2137063	G:C	4.49E-01	6.55E-02	2.02E-08	Ciclev10028638 (intron)	SF18 - ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT
8	2174360	C:T	5.06E-01	7.00E-02	5.99E-09	Ciclev 10029274 (downs.)	PTHR31304:SF1 - LOB DOMAIN-CONTAINING PROTEIN 39
9	30789594	T:C	3.89E-01	5.06E-02	1.40E-09	Ciclev 10007189 (ups.)	KOG4524 - Uncharacterized conserved protein

References

- 1 Xu, Q. *et al.* The draft genome of sweet orange (*Citrus sinensis*). *Nature genetics* **45**, 59-66, doi:10.1038/ng.2472 (2013).
- 2 Wu, G. A. *et al.* Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature biotechnology* **32**, 656-662, doi:10.1038/nbt.2906 (2014).
- 3 Wang, X. *et al.* Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature genetics* **49**, 765-772 (2017).
- 4 Swingle, W. T. & Reece, P. C. in *The citrus industry, revised 2nd ed., vol. 1, History, world distribution, botany, and varieties* Vol. 1 (eds W. Reuther, H.J. Webber, & L.D. Batchelor) 190-430 (University of California, Berkeley, California, USA, 1967).
- 5 Frost, H. B. Seed reproduction: development of gametes and embryos. *The citrus industry* **2**, 290-324 (1968).
- 6 Webb, D. M. & Knapp, S. J. DNA extraction from a previously recalcitrant plant genus. *Plant Molecular Biology Reporter* **8**, 180-185 (1990).
- 7 Carrier, G. *et al.* An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods. *Am J Bot* **98**, e13-15, doi:10.3732/ajb.1000371 (2011).
- 8 Bausher, 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, doi:10.1186/1471-2229-6-21 (2006).
- 9 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
- 10 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
- 11 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307-321 (2010).
- 12 Carbonell-Caballero, J. *et al.* A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus *Citrus*. *Molecular biology and evolution* **32**, 2015-2035, doi:10.1093/molbev/msv082 (2015).
- 13 Swingle, W. The botany of *Citrus* and its wild relatives of the orange subfamily (Family Rutaceae, subfamily Aurnatioidideae). *The citrus industry* **1**, 399 (1943).

- 14 Tanaka, T. Species problem in Citrus. *Japanese Society for Promotion of Science* (1954).
- 15 Hodgson, R. W. *Horticultural varieties of citrus*. (Division of Agricultural Sciences, 1967).
- 16 Hirai, M., Mitsue, S., Kita, K. & Kajiura, I. A survey and isozyme analysis of wild mandarin, tachibana (*Citrus tachibana* (Mak.) Tanaka) growing in Japan. *Journal of the Japanese Society for Horticultural Science* **59**, 1-7 (1990).
- 17 Shimizu, T. *et al.* Hybrid Origins of Citrus Varieties Inferred from DNA Marker Analysis of Nuclear and Organelle Genomes. *PloS one* **11**, e0166969 (2016).
- 18 Nicolosi, E. *et al.* Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *TAG Theoretical and Applied Genetics* **100**, 1155-1166 (2000).
- 19 Pfeil, B. E. & Crisp, M. D. The age and biogeography of Citrus and the orange subfamily (Rutaceae: Aurantioideae) in Australasia and New Caledonia. *Am J Bot* **95**, 1621-1631, doi:10.3732/ajb.0800214 (2008).
- 20 Team, R. C. R: A language and environment for statistical computing. (2013).
- 21 Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *The American Journal of Human Genetics* **93**, 278-288 (2013).
- 22 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81**, 1084-1097 (2007).
- 23 Scora, R. W. Symposium on Biochemical Systematics, Genetics and Origin of Cultivated Plants .9. History and Origin of Citrus. *B Torrey Bot Club* **102**, 369-375, doi:Doi 10.2307/2484763 (1975).
- 24 Gulsen, O. & Roose, M. Chloroplast and nuclear genome analysis of the parentage of lemons. *Journal of the American Society for Horticultural science* **126**, 210-215 (2001).
- 25 Curk, F. *et al.* Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. *Ann Bot* **117**, 565-583 (2016).
- 26 Gillespie, J. H. *Population genetics: a concise guide*. (JHU Press, 2010).
- 27 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, doi:10.1046/j.1529-8817.2003.00063.x (2003).
- 28 Li, Y., Cheng, Y., Tao, N. & Deng, X. Phylogenetic analysis of mandarin landraces, wild mandarins, and related species in China using nuclear LEAFY second intron and plastid trnL-trnF sequence. *Journal of the American Society for Horticultural Science* **132**, 796-806 (2007).
- 29 Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* **61**, 539-542 (2012).
- 30 Sanderson, M. J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular biology and evolution* **19**, 101-109 (2002).

- 31 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, doi:10.1086/671796 (2013).
- 32 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
- 33 Felsenstein, J. PHYLIP-phylogeny inference package (version 3.2). *cladistics* **5**, 164-166 (1989).
- 34 Beattie, G., Holford, P., Mabberley, D., Haigh, A. & Broadbent, P. in *Orlando, Florida, USA: International Conference of Huanglongbing Florida*. 25-57.
- 35 Wang, W., Li, H. & Chen, Z. Analysis of plastid and nuclear DNA data in plant phylogenetics—evaluation and improvement. *Science China Life Sciences* **57**, 280-286 (2014).
- 36 Chat, J., Decroocq, S. & Petit, R. J. A one-step organelle capture: gynogenetic kiwifruits with paternal chloroplasts. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 783-789 (2003).
- 37 Kim, S.-T. & Donoghue, M. J. Incongruence between cpDNA and nrITS trees indicates extensive hybridization within Eupersicaria (Polygonaceae). *Am J Bot* **95**, 1122-1135 (2008).
- 38 Escobar, J. S. *et al.* Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). *BMC Evol Biol* **11**, 1 (2011).
- 39 Yu, W.-B., Huang, P.-H., Li, D.-Z. & Wang, H. Incongruence between nuclear and chloroplast DNA phylogenies in Pedicularis section Cyathophora (Orobanchaceae). *PLoS One* **8**, e74828 (2013).
- 40 Wang, Z.-H., Peng, H. & Kilian, N. Molecular phylogeny of the Lactuca alliance (Cichorieae subtribe Lactucinae, Asteraceae) with focus on their Chinese centre of diversity detects potential events of reticulation and chloroplast capture. *PLoS one* **8**, e82692 (2013).
- 41 Krak, K., Caklová, P., Chrtek, J. & Fehrer, J. Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (Hieracium, Asteraceae). *Heredity* **110**, 138-151 (2013).
- 42 Nauheimer, L., Boyce, P. C. & Renner, S. S. Giant taro and its relatives: a phylogeny of the large genus Alocasia (Araceae) sheds light on Miocene floristic exchange in the Malesian region. *Mol Phylogenet Evol* **63**, 43-51 (2012).
- 43 Rogers, J. & Gibbs, R. A. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature reviews. Genetics* **15**, 347-359, doi:10.1038/nrg3707 (2014).
- 44 Zhou, J. in *Proceedings of the International Citrus Symposium Guangzhou, China, Nov. 5-8, 1990/edited by Huang Bangyan, Yang Qian*. ([Beijing]: International Academic Publishers, c1991.).
- 45 De Candolle, A. *Origine des plantes cultivées*. Vol. 43 (G. Baillièrre et cie, 1883).
- 46 Tanaka, T. Citologia: semi-centennial commemoration papers on citrus studies. *Osaka: Citrologia Supporting Foundation* **114** (1961).
- 47 Rajput, C. & Hari Babu, R. *Citriculture*. (Kalyani, 1985).
- 48 Zhang, W. in *Proceedings of the International Society of Citriculture/[International Citrus Congress, November 9-12, 1981, Tokyo,*

- Japan; K. Matsumoto, editor]. (Shimizu, Japan: International Society of Citriculture, 1982-1983.).
- 49 He, S., Liu, G. & Xiang, D. in *Proc Int Soc Citriculture*. 100-103.
- 50 Gmitter, F. G. & Hu, X. L. The Possible Role of Yunnan, China, in the Origin of Contemporary Citrus Species (Rutaceae). *Econ Bot* **44**, 267-277, doi:Doi 10.1007/Bf02860491 (1990).
- 51 Scora, R. in *Citriculture: proceedings of the Sixth International Citrus Congress: Middle-East, Tel Aviv, Israel, March 6-11, 1988/scientific editors, R. Goren and K. Mendel, editor, N. Goren*. (Rehovot, Israel: Balaban, c1989.).
- 52 Hodgson, R. W. & Sciences, U. o. C. D. o. A. *Horticultural varieties of citrus*. (University of California, Division of Agricultural Sciences., 1967).
- 53 Tanaka, T. The discovery of Citrus tachibana in Formosa, and its scientific and industrial significance. *Studia Citrologia* **5**, 1-20 (1931).
- 54 Liu, G., He, S. & Li, W. Two new species of Citrus in China. *Acta Botanica Yunnanica* **12**, 287-289 (1990).
- 55 Chen, Z. in *Proc Citrus Germplasm conservation Workshop*. 85-95.
- 56 Hodgson, R. The citrus fruits of India. *Calif. Citrogr* **22**, 513-514 (1937).
- 57 Cooper, W. C. *Odyssey of the Orange in China*. (EO Painter Printing Company, 1990).
- 58 Bonavia, E. *The cultivated oranges and lemons, etc. of India and Ceylon*. (Applewood Books, 1888).
- 59 Sykes, S. in *Proceedings citrus germplasm conservation workshop*. 6-7.
- 60 Zhang, D. & Mabberly, D. Citrus (Rutaceae). *Fl. China* **11**, 90-96 (2008).
- 61 Fang, J., Wang, Z. & Tang, Z. *Atlas of woody plants in China: distribution and climate*. Vol. 1 (Springer Science & Business Media, 2011).
- 62 NESOM, G. Citrus trifoliata (Rutaceae): Review of biology and distribution in the USA. *Phytoneuron* **46**, 1-14 (2014).
- 63 Muellner, A., Vassiliades, D. & Renner, S. Placing Biebersteiniaceae, a herbaceous clade of Sapindales, in a temporal and geographic context. *Plant Systematics and Evolution* **266**, 233-252 (2007).
- 64 Fischer, T. C. & Butzmann, R. Citrus meletensis (Rutaceae), a new species from the Pliocene of Valdarno (Italy). *Plant systematics and evolution* **210**, 51-55 (1998).
- 65 Vavilov, N. I. in *Theoretical Basis for Plant Breeding* Vol. 1 17-75 (1935).
- 66 Tolkowsky, S. *Hesperides. A history of the culture and use of citrus fruits*. (J. Bale Sons & Curnow, London., 1938).
- 67 Webber, H. J. in *The Citrus Industry* Vol. 1 Ch. 1, 1-39 (Univ. California., 1967).
- 68 Wen, J., Zhang, J. Q., Nie, Z. L., Zhong, Y. & Sun, H. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in genetics* **5**, 4, doi:10.3389/fgene.2014.00004 (2014).
- 69 Favre, A. *et al.* The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev* **90**, 236-253, doi:10.1111/brv.12107 (2015).

- 70 Wang, L., Schneider, H., Zhang, X.-C. & Xiang, Q.-P. The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biol* **12**, 1 (2012).
- 71 Renner, S. S. Available data point to a 4 - km - high Tibetan Plateau by 40 Ma, but 100 molecular - clock papers have linked supposed recent uplift to young node ages. *J Biogeogr* (2016).
- 72 Clift, P. D., Wan, S. M. & Blusztajn, J. Reconstructing chemical weathering, physical erosion and monsoon intensity since 25 Ma in the northern South China Sea: A review of competing proxies. *Earth-Sci Rev* **130**, 86-102, doi:10.1016/j.earscirev.2014.01.002 (2014).
- 73 Clift, P. D. *et al.* Correlation of Himalayan exhumation rates and Asian monsoon intensity. *Nature Geoscience* **1**, 875-880 (2008).
- 74 Valdiya, K. S. Emergence and evolution of Himalaya: reconstructing history in the light of recent studies. *Prog Phys Geog* **26**, 360-399, doi:10.1191/0309133302pp342ra (2002).
- 75 Valdiya, K. Rising Himalaya: Advent and intensification of monsoon. *CURRENT SCIENCE-BANGALORE*- **76**, 514-524 (1999).
- 76 Iglesias, D. J. *et al.* Physiology of citrus fruiting. *Brazilian Journal of Plant Physiology* **19**, 333-362 (2007).
- 77 Tadeo, F. R. *et al.* Molecular physiology of development and quality of citrus. *Advances in Botanical Research* **47**, 147-223 (2008).
- 78 Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences discussions* **4**, 439-473 (2007).
- 79 Hooker, J. D. *On the Flora of Australia: Its Origin, Affinities, and Distribution, Being an Introductory Essay to the Flora of Tasmania.* (Lovell Reeve, 1859).
- 80 Golonka, J., Krobicki, M., Pajak, J., Nguyen, V. G. & Zuchiewicz, W. *Global plate tectonics and paleogeography of Southeast Asia.* (2006).
- 81 Sun, B.-N. *et al.* Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **304**, 328-336 (2011).
- 82 Kou, X.-Y., Ferguson, D. K., Xu, J.-X., Wang, Y.-F. & Li, C.-S. The reconstruction of paleovegetation and paleoclimate in the Late Pliocene of West Yunnan, China. *Climatic Change* **77**, 431-448 (2006).
- 83 Su, T. *et al.* Post-Pliocene establishment of the present monsoonal climate in SW China: evidence from the late Pliocene Longmen megaflora. *Climate of the Past* **9**, 1911-1920 (2013).
- 84 Thomas, D. C. *et al.* West to east dispersal and subsequent rapid diversification of the mega-diverse genus *Begonia* (Begoniaceae) in the Malesian archipelago. *J Biogeogr* **39**, 98-113, doi:10.1111/j.1365-2699.2011.02596.x (2012).
- 85 Richardson, J. E., Costion, C. M. & Muellner, A. N. in *Biotic Evolution and Environmental Change in Southeast Asia* (ed D.J. et al Gower) Ch. 6, 138-163 (Cambridge University Press, 2012).

- 86 Thomas, D. C. *et al.* Molecular phylogenetics and historical biogeography of the Meiohyne-Fitzalania clade (Annonaceae): Generic paraphyly and late Miocene-Pliocene diversification in Australasia and the Pacific. *Taxon* **61**, 559-575 (2012).
- 87 Hall, R. Southeast Asia's changing palaeogeography. *Blumea* **54**, 148-161, doi:10.3767/000651909X475941 (2009).
- 88 van Welzen, P. C., Slik, J. W. F. & Alahuhta, J. Plant distribution patterns and plate tectonics in Malesia. *Biol. Skr.* **55**, 199-217 (2005).
- 89 Li, Y., Dressler, S., Zhang, D. & Renner, S. S. More Miocene dispersal between Africa and Asia—the case of *Bridelia* (Phyllanthaceae). *Systematic Botany* **34**, 521-529 (2009).
- 90 Schaefer, H., Heibl, C. & Renner, S. S. Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 843-851 (2009).
- 91 Chiang, T.-Y. & Schaal, B. A. Phylogeography of plants in Taiwan and the Ryukyu Archipelago. *Taxon* **55**, 31-41 (2006).
- 92 Huang, S.-F. Hypothesizing Origin, Migration Routes and Distribution Patterns of Gymnosperms in Taiwan. *Taiwania* **59**, 139-163 (2014).
- 93 Wang, P. & Sun, X. Last glacial maximum in China: comparison between land and sea. *Catena* **23**, 341-353 (1994).
- 94 Dobson, M. & Kawamura, Y. Origin of the Japanese land mammal fauna: allocation of extant species to historically-based categories. *The Quaternary Research* **37**, 385-395 (1998).
- 95 Voris, H. K. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *J Biogeogr* **27**, 1153-1167 (2000).
- 96 Huang, S.-F. & Lin, T.-P. Migration of *Trochodendron aralioides* (Trochodendraceae) in Taiwan and its adjacent areas. *Botanical Studies* **47**, 83-88 (2006).
- 97 Gibbard, P. & Cohen, K. M. Global chronostratigraphical correlation table for the last 2.7 million years. *Episodes* **31**, 243-247 (2008).
- 98 Baldwin, E. in *Biochemistry of fruit ripening* 107-149 (Springer, 1993).
- 99 Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature genetics* **44**, 821-824, doi:10.1038/ng.2310 (2012).
- 100 Meléndez-Hevia, E., Waddell, T. G. & Cascante, M. The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *Journal of Molecular Evolution* **43**, 293-303 (1996).
- 101 Guo, L.-X. *et al.* Citrate Accumulation-Related Gene Expression and/or Enzyme Activity Analysis Combined With Metabolomics Provide a Novel Insight for an Orange Mutant. *Scientific Reports* **6** (2016).
- 102 Huang, D., Zhao, Y., Cao, M., Qiao, L. & Zheng, Z.-L. Integrated systems biology analysis of transcriptomes reveals candidate genes for acidity control in developing fruits of sweet orange (*Citrus sinensis* L. Osbeck). *Frontiers in plant science* **7** (2016).

- 103 Ghosh, S. P. in *Proc. Citrus Germplasm Conservation Workshop* 96-105 (Brisbane Australia, 1997).
- 104 Yang, X. *et al.* Genetic diversity and phylogenetic relationships of citron (*Citrus medica* L.) and its relatives in southwest China. *Tree Genetics & Genomes* **11**, 1-13 (2015).
- 105 He, S. W. L., G.; Wenpin, L. Wild mandarin oranges. *China Proc. Int. Soc. Citriculture*, 113-121 (1988).
- 106 Chen, Z. in *Proc Citrus Germplasm conservation Workshop* (ed Bevington KB Sykes SR, Hailstones D) 85-95 (Brisbane. Australia, 1997).