Ancient DNA reveals early use of melons in China's Song Dynasty

Nathanael Walker-Hale[1]*, Yunfei Zheng[2], Oscar Alejandro Pérez-Escobar[3], Abdelhafid Bendahmane[4], Michaela Preick[5], Michael V. Westbury[6,7], Michael Hofreiter[5], Hanno Schaefer[8], Xinyi Liu[9]*, Guillaume Chomicki[1]*, Susanne S. Renner[10]*

[1]Department of Biosciences, Durham University, Durham DH1 3LE, UK.

[2]Zhejiang Institute of Archaeology and Cultural Heritage, Hangzhou, Zhejiang Province, 310023, China.

[3]Royal Botanic Gardens, Kew, London, TW9 3AE, UK.

[4]Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, Université d'Evry, Gif-sur-Yvette, France.

[5]Faculty of Mathematics and Natural Sciences, Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany.

[6]Globe Institute, University of Copenhagen, Øster Farimagsgade 5, København K.

[7]Department of Health Technology, Section for Bioinformatics, Technical University of Denmark, Kongens Lyngby, Denmark.

[8]Technical University of Munich, Department Life Science Systems, 85354 Freising, Germany.

[9]Department of Anthropology, Washington University in St. Louis, St. Louis, MO 63130, USA

[10]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA

Corresponding authors: Nathanael Walker-Hale (nathanael.j.walker-hale@durham.ac.uk), Xinyi Liu (liuxinyi@wustl.edu), Guillaume Chomicki (guillaume.chomicki@durham.ac.uk) and Susanne S. Renner (srenner@wustl.edu)

Abstract

Melon (*Cucumis melo* L.) domestication is thought to have independently occurred once in Northeast Africa and twice in India. However, archaeobotanical seed remains suggest a possible additional domestication in China. Because *Cucumis* seeds are difficult to diagnose morphologically, genomic data from archaeological material are needed to test these scenarios and to infer ancient melon traits. We sequenced two Song Dynasty melon seeds from Shuomen Gugang (China), recovering 4.5x and 1.7x nuclear genome coverage. Nuclear and chloroplast analyses place both seeds within cultivated *C. melo* accessions from China, supporting introduction from the broader Asian domestication pool rather than an independent Chinese domestication event. To test whether Song Dynasty melon seeds carried trait associated with sweet dessert melons, we analyzed loci underlying fruit phenotypes. Both seeds lacked alleles associated with orange flesh; one carried an allele linked to yellow/orange peel coloration and the other possessed alleles associated with green flesh and reduced acidity. These results suggest that melons in Song Dynasty China were likely consumed as fresh or culinary fruits rather than as highly sweet dessert melons. Their flesh color echoes the Song-period aesthetic traditions, exemplified by the appreciation of jade-green celadon, which was often produced in melon-shaped forms. Thus, by anchoring East Asian archaeobotanical remains within modern melon genomic variation, our study provides a temporal framework for melon dispersal into China and demonstrates how ancient genomics can illuminate crop use in the past.

Main text

Melon (*Cucumis melo* L.) is an important crop from the Cucurbitaceae family, with worldwide cultivation exceeding 41.5 million tons in 2024 (FAO; http://faostat.fao.org/). Since its genome was sequenced in 2012 (Garcia-Mas et al., 2012), melon domestication has been extensively studied over the past decade, yielding valuable insights into the roles of selection and introgression (Endl et al., 2018; Zhao et al., 2019; Liu et al., 2020; Pereira et al., 2021; Lyu et al., 2023; Imoh et al., 2025). These efforts, however, have been limited by insufficiently broad geographic sampling, preventing a comprehensive understanding of the genetic basis for melon adaptation and breeding.

Molecular evidence suggests that melon domestication occurred independently once in Northeast Africa, and twice in India (Endl et al., 2018; Zhao et al., 2019; Liu et al., 2020). Archaeobotanical evidence suggested a potential independent fourth domestication in the lower Yangtze region of China (Zheng and Chen 2006; Fuller, 2006, 2012; Fuller et al., 2014). The Chinese archaeobotanical case relies mainly on seed macroremains and inferred trends (especially seed-size change), and morphology-based domestication is notoriously difficult in Cucurbitaceae (Wolcott et al., 2021). Although seed size can track fruit size, *Cucumis* seed characters overlap across domesticated, feral and wild forms, and preservation can obscure diagnostic features (Fuller, 2012; Fuller et al., 2014). This challenge is particularly acute in East Asia, where small-seeded archaeological finds have sometimes been linked to names used for “wild melons” in regional literature. Molecular data are needed to confirm any morphology-based identifications of ancient seeds and to test among alternative domestication scenarios for this species.

The inconsistency between genetic and archaeological evidence points to three possible scenarios: (a) independent domestication of *Cucumis melo* in ancient China; (b) early *Cucumis* use in East Asia associated with species other than *C. melo*; and (c) independent Chinese domestication that was later replaced by introduced Indian/African varieties. These gaps in our knowledge of *Cucumis* domestication highlight the importance of the Lower Yangtze environment, particularly given its continuous archaeological record since 7000 BP and the absence of comparable records in Africa or South Asia. Thus, a combined archaeo-genomic approach in the Lower Yangtze is timely.

Numerous *Cucumis* seed remains were recovered from archeological sites across China. We extracted DNA and obtained sequencing data from 24 seeds: 20 were from the Nanshan and

Tingshan sites in Shaoxing, dating from the Spring and Autumn period (770-476 BC), but we could not recover sufficient endogenous DNA. A further four were recovered from Shuomen Gugang (henceforth Gugang), an ancient port at Wenzhou. The site's heyday was during the Song Dynasty (960-1270 CE), when it served as a major medieval harbor connecting the Chinese economy with maritime trading systems across the eastern Pacific and Indian Oceans. Excavations in 2024 revealed archaeological features including city walls, defensive structures, a water gate, a pier, wooden plank roads, a sunken shipwreck, and many Longquan celadon ceramics, a key Chinese export of the period. Large numbers of faunal and macrobotanical remains were also recovered. The archaeobotanical assemblage, both charred and waterlogged, comprises abundant rice grains as well as peach (*Prunus persica*), apricot (*Prunus armeniaca*), Chinese bayberry (*Myrica rubra*), cherry (*Prunus avium*), grape (*Vitis vinifera*), olive (*Olea europaea*), jujube (*Ziziphus jujuba*), wax gourd (Benincasa hispida), and, notably, *Cucumis* seeds. Tropical fruits including lychee (*Litchi chinensis*) and kiwifruit (*Actinidia chinensis*) were also present, attesting to extensive maritime trade connections. Two of these seeds from waterlogged deposits dating to the Southern Song period provided sufficient endogenous DNA; we directly radiocarbon dated another from the same context to 930-800 cal. BP (c. 1020-c. 1150 CE).

Here, we generate and analyze ancient DNA from archaeological *Cucumis* remains from China to (i) confirm taxonomic identity of ancient melon seeds, (ii) test whether early Chinese material is consistent with an independent domestication event versus introduction from India followed by local selection, and (iii) evaluate the fruit traits of ancient Chinese melons. By anchoring East Asian archaeobotanical material directly within the genomic landscape of melon diversity, our study provides a critical temporal perspective on melon dispersal and domestication as well as insights into the traits, and hence likely uses, of melon in ancient China.

Results and discussion

We performed low coverage sequencing of 32 libraries prepared from 24 *Cucumis* seeds recovered from three sites in China, alongside extraction and library blanks (average coverage from unique hits 0.00017x, **Supplementary Information Online**). Two of these latter seeds from Gugang (GG1 and GG4, hereafter) yielded sufficient endogenous DNA for deep sequencing. We sequenced ~2.2 and ~2.5 billion single-end reads. After removing adapters, trimming reads, discarding short fragments, and marking and removing optical and PCR duplicates, we recovered 32 and 13 million unique fragments aligning to the *C. melo* DHL92 reference genome, yielding 4.5x and 1.7x coverage (138x and 64x chloroplast coverage), respectively. Misincorporation patterns at the ends of sequenced fragments suggested that our samples represented *bona fide* ancient DNA (**Supplementary Information Online**). We combined our ancient data with extant accessions of *C. melo* from Zhao et al. (2019) to infer pseudohaploid sequences and genotype likelihoods, and to place our ancient samples phylogenetically. Principal components analysis and phylogenetic analysis of four-fold degenerate sites suggest that our ancient seeds are most closely related to cultivated genotypes of *C. melo* from China in clade III (Zhao et al's '*agrestis*' clade) (**Fig. 1**), a result broadly consistent in chloroplast data and robust to missing data (**Supplementary Information Online**). Zhao et al. argued for three independent domestication events in groups corresponding to *C. melo* subsp. *meloides* (clade I), clade II (Zhao's '*melo*' clade), and clade III, and the placement of our ancient seeds therefore supports introduction from the broader Asian domestication pool in clade III, rather than an independent Chinese domestication event.

The genetically informed identification is consistent with the morphological observation of the waterlogged remains (**Fig. 1C**; the ancient seed) and raises further questions about the dispersal routes of cultivated *C. melo* originating in Africa and South Asia. The locality of Gugang and its archaeological contexts allude to maritime trading links with tropical regions. Beyond the lower Yangtze, morphologically attested *C. melo* has previously been reported from, notably, gut contents recovered in ancient burials at Chashancun (691 CE) and Astana (c. 700 CE) in northwestern China,

and from the royal tomb of a Han prince Haihunhou (59 BCE) in southern China (Chen et al. 2020; Wang et al. 2020; Lu et al. 2022). The latter example indicates that cultivated *C. melo* was introduced to China before the Han Dynasty, which is consistent with historical textual evidence. In Central Asia, the oldest credible record of *Cucumis* sp. is from Kainar, Tajikistan, dating to the fifth century CE and postdating Haihunhou (Nur-Makhamad et al. 2024). Morphologically identified *C. melo* are older along the Nile and Indus, dating to *c*. 5600 and 4000 BP, respectively (Fuller et al. 2025). It is not implausible that *C. melo* was first introduced to southern China during the Han period via routes along the Himalayan slopes rather than the northern Silk Route, with subsequent introductions through maritime networks during the Song. Such episodic dispersals resonate with the archaeological records but remain speculative. The relationship between our Song-dynasty seeds and the Neolithic *Cucumis* (c. 7000-4000 BP) in the lower Yangtze is equally ambiguous. It is possible that introduced melons around the time of the Han Dynasty (or indeed, Song) replaced local *Cucumis* cultivation. However, given the continuity of *Cucumis* in archaeobotanical records spanning the Neolithic and Bronze Age, a local domestication scenario (i.e., from a gene pool not represented by our ancient seeds or extant genotypes) cannot be ruled out. Further aDNA sampling targeting Neolithic or Bronze age seed remains would be timely, but given the low success rate of our sequencing efforts, this is likely to be challenging.

To infer the potential phenotypes of these ancient melons, which might inform how they were used, we compared read alignments in our ancient samples at key loci governing various fruit phenotypes in melon. *CmOR* is a locus involved in β-carotene biosynthesis in melon fruit flesh; the His^{108} allele conferring orange flesh is dominant over the Arg^{108} allele conferring white/green flesh (Tzuri et al. 2015). Both GG1 and GG4 showed evidence of the Arg^{108} allele, suggesting that they lacked orange flesh; in GG1 the probability of not observing the heterozygote if it was present, given the coverage, is very low (**Fig. 2F**). *Reduced Plastid Gene Expression 1* (*CmRPGE1***;** Valverde et al., 2025) encodes a protein involved in plastid development and chlorophyll maintenance in melon fruit flesh. Variation at this locus determines whether non-orange melons develop white flesh (chlorophyll loss) which is dominant to green flesh (chlorophyll retained). We found GG4 harboring the green flesh allele (**Fig. 2G**), consistent with a higher frequency of white/green flesh in clade III, but due to low coverage we cannot rule out heterozygosity at this locus. For peel coloration, GG1 lacked a white-peel-associated deletion in the 5′ UTR of *CmKFB* (Feder et al., 2015; **Fig. 2H**), suggesting that it had yellow peel. We further recovered a single read in GG4 supporting a 12bp duplication in *CmPH* known to control fruit acidity (Cohen et al. 2014), suggesting that this fruit may have been non-acidic (**Fig. 2I**). However, given that both acidic and non-acidic phenotypes can be heterozygous at this locus in modern populations, the genetic architecture of acidity is likely more complex (**Fig. 2I**, Zhao et al. 2019). In melons, fruit bitterness is mediated by cucurbitacin production, and in other cucurbit crops a truncation in the bitterness regulator *Bt* leads to loss of bitterness. We detected no evidence of such a truncation. However, in cultivated melons bitterness may instead be regulated by loss of *CmBt* expression through promoter variation (Zhao et al., 2017; Zhao et al., 2019), and we could not rule this out in our ancient samples.

Several extant melons combine the same trait suite inferred from our Song Dynasty seeds: non-orange flesh, often yellow peel, and sometimes white to light-green flesh. These are commonly used as fresh, mild fruits, vegetables or seeds. Using light-green melon flesh as vegetables or pickles is common culinary practice in the Lower Yangtze today. Elsewhere in East Asia, Korean “oriental melon” (makuwa/chamoe) is described as having a distinctive yellow peel and white flesh, with some landraces showing light-green flesh (Tuan et al., 2019; Jang et al., 2024). In Japan, pickling melons (e.g., conomon “Katsura-uri”) are not sweet and are traditionally used at immature/semi-ripened stages for pickles (Sasaki et al., 2017). In South India, “culinary melon” (acidulus) is widely grown and cooked in savoury dishes (“sambar cucumber” types) (Shet et al., 2022), while in Sudan, tibish is eaten as a vegetable and seinat is valued for roasted seeds (Imoh et al., 2025). Taken together, these modern analogues suggest that melons sharing the trait combination observed in our ancient samples (non-orange, white or potentially green flesh, and yellow peel) were likely used in Song Dynasty China as crisp, mildly sweet fruits, in fresh or processed forms, and potentially also as a source of edible seeds, rather than being selected exclusively for high sugar content. This interpretation is further supported by cultural considerations: “oriental melon” types with light-green flesh and yellow peel have long held symbolic significance in East Asia, where abundant seeds were associated with fertility and continuity, and green flesh resonated with established aesthetic values. In particular, green celadon ceramics produced in the Lower Yangtze, including Longquan celadon, were frequently fashioned in melon-shaped forms, with glazes echoing the jade-green tones of melon flesh, a typology that became especially prominent during the Song period.

Materials and Methods

Ancient DNA extraction, sequencing, and radiocarbon dating

We selected 24 *Cucumis* seed remains to attempt ancient DNA analysis: 20 from the Nanshan (NS) and Tingshan (TS) sites in Shaoxing, dating from the Spring and Autumn period (770-476 BC), and four from Shuomen Gugang (GG), an ancient port at Wenzhou dating from the Song Dynasty (960-1270 CE). Seeds were stored at 4°C prior to DNA extraction. The outer seed coat (testa) was separated from the inner tissues and used separately for extraction.

DNA was extracted from the seed coats and inner parts of GG1-4 and NS1-4, and seed coats of NS 5-10 and TS 1-10 following Dabney et al. (2013) and Wales et al. (2014). For digestion, a lysis buffer containing 0.5% (w/v) *N*-lauroylsarcosine (Sigma-Aldrich L9150-50G), 50 mM Tris-HCl (Thermo Fisher Scientific 15568025), 20 mM EDTA (VWR E177-500MLDB), 150 mM NaCl (Thermo Fisher Scientific AM9760G), 3.3% 2-mercaptoethanol (Sigma-Aldrich 63689-25ML-F), 50 mM DL-dithiothreitol (Sigma-Aldrich D9779-250MG), and 0.25 mg/mL proteinase K (Promega V3021) was applied to the powdered material as described in Wales et al. (2014). DNA purification followed Dabney et al. (2013) with reduced centrifugation speed (450 × g).

Library construction and sequencing were performed in the ancient DNA facility at the University of Potsdam; negative controls were included at all steps. DNA extracts were converted to Illumina sequencing libraries using the single-strand approach of Gansauge et al. (2017) using custom adapters from Gansauge & Meyer (2013). Quantitative PCR (qPCR) was performed on a PikoReal 96 Real-Time PCR machine (Thermo Fisher Scientific TCR0096) using 0.2% of each unamplified library with the following thermal profile: 10 min initial denaturation at 95°C, followed by 40 cycles of 15 s at 95°C, 30 s at 60°C, and 1 min at 72°C. Each 10 µL qPCR reaction contained 1 µL diluted library, 1× SYBR Green qPCR Master Mix (Applied Biosystems 4309155), and 0.5 µM each of primers IS7 and IS8; three technical replicates were run per library. Indexing PCR was performed using the cycle number determined by qPCR, adding 8-bp indices to both 5′ and 3′ adapters. Preliminary DNA sequencing was performed on an Illumina NextSeq 500 sequencing platform, using the 500/550 High

Output v2 kit (75 cycles, Illumina FC-404-2005), with a custom read-1 (Gansauge and Meyer 2013) and a custom index-2 (Paijmans et al. 2017) sequencing primer at the University of Potsdam, Germany. For deep sequencing of two successful seeds, pooled libraries (5:4 ratio) were sequenced on an Illumina NovaSeq S2 flowcell (two lanes, 100 cycles), using the same sequencing primers as above, at the Globe Institute, University of Copenhagen, Denmark.

All but two of these seeds (GG1 and GG4) did not yield sufficient endogenous DNA for deep sequencing. Both successful seeds derive from contexts dated to the Southern Song period (1127-1279 CE). We used another *Cucumis* remain (GG6) from the same context for radiocarbon analysis at Center for Applied Isotope Studies, University of Georgia, USA. The sample preparation followed well-established acid-base-acid (ABA) chemical pre-treatment, followed by combustion and graphitization prior to accelerator mass spectrometry (AMS) measurement. The result was calibrated by the OxCal (v4.4) and the Int Cal20 calibration curve. The result is 928–796 cal. BP (reported relative to 1950 CE at the 95.4% probability range), corresponding approximately 1020-1150 CE, placing it securely within the Southern Song period. GG1 and GG4 was not directly dated but derive from the same stratigraphic unit.

Ancient DNA read processing

Raw reads from the two ancient seeds were preprocessed and aligned using the BAM pipeline in PALEOMIX v1.3.10 (Schubert et al., 2014). AdapterRemoval v2.3.4 (Schubert et al., 2016) was used to remove adapter sequences, with a maximum mismatch threshold of 1/3, trimming Ns and low-quality bases and discarding fragments shorter than 30 bp. Trimmed reads were aligned to the *C. melo* DHL92 v4.0 nuclear reference (Castanera et al., 2020) and a chloroplast reference (NC_015983.1; Rodríguez-Moreno et al., 2011), from which one copy of the inverted repeat was removed, using BWA v0.7.18 (Li and Durbin, 2009) with the backtrack algorithm, which is appropriate for short reads. To assess characteristic patterns of post-mortem DNA damage, we estimated the proportion and position of misincorporations in aligned reads using mapDamage v2.2.2 (Jónsson et al., 2013). Based on these results, we trimmed the first and last three nucleotides from each read and removed fragments shorter than 30bp, prior to realignment with BWA backtrack. Quality scores were then rescaled to account for post-mortem damage using mapDamage, reads were realigned around indels with GATK v3.8.1 (Van der Auwera and O'Connor, 2020), and duplicates were marked with Picard MarkDuplicates implemented in GATK v4.6.2.

Raw read data processing

Sequencing data for 1,175 accessions of *Cucumis melo* (subsp. *melo* and *agrestis*) from Zhao et al. (2019) were downloaded from NCBI (BioProject PRJNA565104), and inbred lines were excluded. Cultivated accessions were further subsampled, yielding a final dataset of 305 samples. Read quality was assessed with FastQC v0.12.1 (Andrews, 2010), and Ns and low-quality bases were removed with Trim Galore! v0.6.10 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using Cutadapt v5.2 (Martin, 2011). Reads were aligned to the same references as above using BWA-MEM (Li, 2013) and similarly subjected to indel realignment and duplicate marking.

Ancient DNA pseudohaploidisation and error assessment

Due to low coverage, we employed a pseudohaploid approach following Pérez-Escobar et al. (2022): sample depths and nucleotide occurrences were calculated using ANGSD v0.940 (-doCounts 1 -minQ 30 -minMapQ 20 -dumpCounts 3). Pseudohaploid sequences for the chloroplast genome and each of the 12 nuclear chromosomes were inferred using Consensify v2.4.0, with maximum depth set to the upper 95th percentile of coverage for each sample (Barlow et al., 2020).

We assessed error rates by computing the excess of derived alleles in ancient relative to extant samples using the -doAncError function in ANGSD, with a modern *C. melo* subsp. *agrestis* consensus

sequence as reference (Pérez-Escobar et al., 2022). We created a reference FASTA from the MS-996 BAM with ANGSD (-doFasta 1 -doCounts 1 -minQ 30 -minMapQ 20 -setMinDepth 15), then used this together with the reference DHL92 genome to calculate excess substitutions as error rates for each ancient sample and three relatively high coverage accessions (MS-1001, MS-1002, and MS-1021) as controls, using ANGSD (-doAncError 1).

Phylogenomic placement of ancient samples

We used the pseudohaploid sequences from our ancient samples for phylogenomic inference. For extant accessions, pseudohaploid sequences were generated by randomly sampling a base at each position meeting quality and depth thresholds (-doFasta 1 -minQ 30 -minMapQ 20 -setMinDepth 5).

The chloroplast alignment was analyzed in a maximum likelihood framework using IQ-TREE v3.0.1 under a GTR+F+G model with 1,000 ultrafast bootstrap replicates and 1000 replicates for the SH-aLRT test (-m GTR+F+G -B 1000 -alrt 1000; Wong et al. 2025; Hoang et al. 2018). For nuclear data, four-fold degenerate sites were identified using degenotate v1.2.4 (Mirchandani et al., 2023) and extracted from each pseudohaploid genome. Invariable columns were removed with goalign v0.4.0 (--char MAJ --cutoff 1; Lemoine and Gascuel, 2021), and columns with >20% gaps were filtered using pxclsq from phyx (Brown et al. 2017). The final nuclear alignment was analyzed with IQ-TREE using the same settings as above, except that we added an ascertainment bias correction to account for the removal of fixed sites (-m GTR+F+G+ASC).

Population structure of Cucumis melo

To call genotype likelihoods, we used ANGSD with the GATK model, applying the default minimum Minor Allele Frequency cutoff of 0.05 and testing for polymorphism with a SNP p-value cutoff of $1x10^{-6}$, using the same quality thresholds as previously (-GL 2 -doMajorMinor 1 -doMaf 1 -doGlf 2 -SNP_pval 1e-6 -minQ 30 -minMapQ 20 -minMaf 0.05). To account for missing data, we repeated this process requiring that all individuals be present (-minInd 296). For Principal Components Analysis (PCA), we used pcangsd v1.36.4 (Meisner and Albrechtsen, 2018) to calculate the covariance matrix, setting a maximum of 1000 iterations (--iter 1000). PCA on the resulting matrix was conducted in R v4.5.0 (R Core Team, 2025) using the eigen() function. To calculate admixture proportions, we used NGSadmix v33 (Slotte et al. 2013), with K=2-8, setting the likelihood difference in 50 iterations to 0.05.

Comparative analysis of functional variants

Using IGV (Robinson et al. 2011), we manually inspected the alignments of our ancient DNA reads at loci known to determine fruit phenotypes in *Cucumis melo*, namely the bHLH regulator of bitterness CmBt (Zhou et al. 2017), the fruit color determinant CmOr (Tzuri et al. 2015), acidity regulator CmPH (Cohen et al. 2014), peel color regulator CmKFB (Feder et al. 2015), and the recently described locus determining white or green flesh in non-orange varieties, CmRPGE1 (Valverde et al. 2025). For this analysis, we attempted to increase informative data by reducing our fragment length filter to a minimum of 25bp after trimming each end. In each case, we searched read alignments for evidence of documented allelic variation corresponding to phenotypes.

Acknowledgments

We thank the UK Natural Environment Research Council for an Independent Research Fellowship (NE/S014470/3) and ERC/UKRI for a frontier research grant (EP/X026868/1) to G.C. A.B. is funded by NectarGland ERC Project (101095736). We thank Melissa Ritchey for assistance with imaging the archaeological seed remains.

References

1. J. Endl, E. G. Achigan-Dako, A. K. Pandey, A. J. Monforte, B. Pico, H. Schaefer, Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. *Am. J. Bot.* **105**, 1662–1671 (2018).
2. G. Zhao, Q. Lian, Z. Zhang, Q. Fu, Y. He, S. Ma, V. Ruggieri, A. J. Monforte, P. Wang, I. Julca, *et al.*, A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.* **51**, 1607–1615 (2019).
3. S. Liu, P. Gao, Q. Zhu, Z. Zhu, H. Liu, X. Wang, Y. Weng, M. Gao, F. Luan, Resequencing of 297 melon accessions reveals the genomic history of improvement and loci related to fruit traits in melon. *Plant Biotechnol. J.* **18**, 2545–2558 (2020).
4. Y. F. Zheng, X. G. Chen, Archaeological research on the origin of *Cucumis* seeds unearthed from the lower reaches of the Yangtze River. In: ZPICRA (ed). Zhejiang Provincial Institute of Cultural Relics and Archaeology Science (Series 8). *Proceeding of Symposium to Commemorate the 70th Anniversary Discovery of the Liangzhu Site*. Science Press, Beijing, pp 578-585 [in Chinese]
5. D. Q. Fuller, Agricultural origins and frontiers in South Asia: a working synthesis. *J. World Prehist.* **20**, 1–86 (2006).
6. D. Q. Fuller, "New archaeobotanical information on plant domestication from macro-remains: tracking the evolution of domestication syndrome traits" in *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*, P. Gepts, T. R. Famula, R. L. Bettinger, S. B. Brush, A. B. Damania, P. E. McGuire, C. O. Qualset, Eds. (Cambridge University Press, Cambridge, UK, 2012), pp. 110–135.
7. D. Q. Fuller, T. Denham, M. Arroyo-Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby, M. D. Purugganan, Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6147–6152 (2014).
8. K. A. Wolcott, G. Chomicki, Y. M. Staedler, K. Wasylikowa, M. Nesbitt, J. Schönenberger, S. S. Renner, Three-dimensional X-ray-computed tomography of 3300-to 6000-year-old *Citrullus* seeds from Libya and Egypt compared to extant seeds throws doubts on species assignments. *Plants People Planet* **3**, 694–702 (2021).
9. J. Garcia-Mas, A. Benjak, W. Sanseverino, M. Bourgeois, G. Mir, V. M. González, E. Hénaff, F. Câmara, L. Cozzuto, E. Lowy, *et al.*, The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11872–11877 (2012).
10. O. A. Pérez-Escobar, S. Tusso, N. A. S. Przelomska, S. Wu, P. Ryan, M. Nesbitt, M. V. Silber, M. Preick, Z. Fei, M. Hofreiter, *et al.*, Genome sequencing of up to 6,000-year-old *Citrullus* seeds reveals use of a bitter-fleshed species prior to watermelon domestication. *Mol. Biol. Evol.* **39**, msac168 (2022).
11. Y. Zhou, Y. Ma, J. Zeng, L. Duan, X. Xue, H. Wang, T. Lin, Z. Liu, K. Zeng, Y. Zhong, *et al.*, Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat. Plants* **2**, 16183 (2016).
12. S. Cohen, A. Itkin, D. Yeselson, G. Tzuri, M. Portnoy, A. Harel-Beja, S. Lev, U. Sa'ar, K. Davidovitz-Rikanati, A. Baranes, *et al.*, The PH gene determines fruit acidity and contributes to the evolution of sweet melons. *Nat. Commun.* **5**, 4026 (2014).
13. G. Tzuri, X. Zhou, N. Chayut, H. Yuan, V. Portnoy, A. Meir, U. Sa'ar, F. Bber, S. Aviber, T. Burger, *et al.*, A 'golden' SNP in *CmOr* governs the fruit flesh color of melon (*Cucumis melo*). *Plant J.* **82**, 267–279 (2015).
14. L. Valverde, S. Torres-Montilla, M. Pujol, J. Garcia-Mas, M. Rodriguez-Concepcion, A melon (*Cucumis melo*) homologue of REPRESSOR OF PHOTOSYNTHETIC GENES prevents chloroplast differentiation in the fruit flesh. *bioRxiv* [Preprint] (2025). https://doi.org/10.1101/2025.08.28.672820 (Accessed 16 January 2026).
15. A. Feder, J. Burger, S. Gao, N. Lewinsohn, N. Katzir, A. A. Schaffer, J. Meir, J. Fait, J. Tadmor, *et al.*, A Kelch domain-containing F-Box coding gene negatively regulates flavonoid accumulation in muskmelon. *Plant Physiol.* **169**, 1714–1726 (2015).
16. P. A. Tuan, J. K. Kim, S. Lee, S. C. Chae, S. U. Park, Carotenoid biosynthesis in oriental melon (*Cucumis melo* L. var. *makuwa*). *Foods* **8**, 77 (2019).
17. Y.-J. Jang, H.-S. Cho, Y.-J. Park, H.-L. Kim, S.-H. Lee, Quality evaluation and shipping export potential of 'Eliteggul' Korean melon (*Cucumis melo* var. *makuwa*) to Singapore using MAP technology. *Horticulturae* **10**, 1354 (2024).
18. A. Sasaki, Y. Nakamura, Y. Kobayashi, W. Aoi, T. Nakamura, K. Shirota, N. Suetome, M. Fukui, T. Shigeta, T. Matsuo, *et al.*, A new strategy to protect Katsura-uri (Japan's heirloom pickling melon, *Cucumis melo* var. *conomon*) from extinction. *J. Ethn. Foods* **4**, 44–50 (2017).

19. R. M. Shet, R. K. Hiremath, M. B. Chetti, S. S. Hebsur, Evaluation of postharvest shelf life of south Indian culinary melon (*Cucumis melo* var. *acidulus*) accessions. *J. Appl. Hortic.* **24**, 84–88 (2022).
20. O. N. Imoh, S. S. Renner, H. Schaefer, M. A. Leko, A. M. Mohammed, Y. S. A. Gorafi, H. Tsujimoto, Molecular polymorphisms of the nuclear and chloroplast genomes among African melon germplasms reveal abundant and unique genetic diversity, especially in Sudan. *Ann. Bot.* **135**, 1329–1344 (2025).
21. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, *et al.*, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013).
22. N. Wales, K. Andersen, E. Cappellini, M. C. Ávila-Arcos, P. T. Gilbert, Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS One* **9**, e86827 (2014).
23. M.-T. Gansauge, et al., Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res* **45**, e79 (2017).
24. M.-T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
25. J. L. A. Paijmans, R. Barnett, M. T. P. Gilbert, M. L. Zepeda-Mendoza, J. W. F. Reumer, J. de Vos, G. Zazula, D. Nagel, G. F. Baryshnikov, J. A. Leonard, *et al.*, Evolutionary history of saber-toothed cats based on ancient mitogenomics. *Curr. Biol.* **27**, 3330–3336 (2017).
26. M. Schubert, L. Ermini, C. Der Sarkissian, H. Jónsson, A. Ginolhac, R. Schaefer, M. D. Martin, R. Fernández, M. Kircher, M. McCue, *et al.*, Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
27. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
28. R. Castanera, J. Ruggieri, M. A. Arro, V. Sanseverino, Genomic diversity and population structure of a reference melon inbred line DHL92. *G3 (Bethesda)* **10**, 1399–1410 (2020).
29. L. Rodríguez-Moreno, V. M. González, A. Benjak, M. C. Martí, P. Puigdomènech, M. A. Aranda, J. Garcia-Mas, Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics* **12**, 424 (2011).
30. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
31. H. Jónsson, A. Ginolhac, M. Schubert, P. L. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
32. G. A. Van der Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, Sebastopol, CA, 2020).
33. S. Andrews, FastQC: a quality control tool for high throughput sequence data (2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed January 2026).
34. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
35. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint] (2013). https://arxiv.org/abs/1303.3997.
36. M. Stiller, R. E. Green, M. Ronan, J. F. Simons, L. Du, W. He, M. Egholm, J. M. Rothberg, S. G. Keates, N. D. Ovodov, *et al.*, Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13578–13584 (2006).
37. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
38. A. Barlow, S. Hartmann, J. Gonzalez, M. Hofreiter, J. L. A. Paijmans, Consensify: a method for generating pseudohaploid genome sequences from palaeogenomic datasets with reduced error rates. *Genes* **11**, 50 (2020).
39. T. K. F. Wong, C. Jermiin, L. S. Jermiin, L. Nguyen, IQ-TREE version 3: advancing phylogenetic inference with enhanced models, methods, and tools. *Mol. Biol. Evol.* **42**, msaf012 (2025).
40. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

41. C. D. Mirchandani, A. J. Shultz, G. W. C. Thomas, S. J. Smith, M. Bayber, B. Arnold, R. Corbett-Detig, E. Enbody, S. Satta, P. Wadsworth, *et al.*, A fast, reproducible, high-throughput variant calling workflow for population genomics. *Mol. Biol. Evol.* **41**, msad270 (2023).
42. F. Lemoine, O. Gascuel, Goalign: a comprehensive tool for sequence alignment manipulation. *NAR Genom. Bioinform.* **3**, lqab075 (2021).
43. J. W. Brown, J. F. Walker, S. A. Smith, Phyx: phylogenetic tools for Unix. *Bioinformatics* **33**, 1886–1888 (2017).
44. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
45. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2025). https://www.R-project.org/.
46. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).
47. J. Meisner, A. Albrechtsen, Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* **210**, 719–731 (2018).
48. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics* **195**, 693–702 (2013).
49. J. T. Robinson, *et al.*, Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).

Figures

Figure 1. Phylogenomic placement of ancient seeds. a) Map of China and surrounding countries, showing the location of the Gugang archeological site in Wenzhou. **b)** The Gugang archeological site. **c)** Ancient melon seed recovered from the Gugang site. **d)** Principal Components Analysis based on inferred genotype likelihoods of ancient seed genomes in the context of a diverse collection of 294 extant melon genotypes. **e)** placement of the ancient genomes in a maximum likelihood phylogeny with the same collection of genotypes and 9 outgroups, based on 9,627 four-fold degenerate sites. Colors at tips indicate the geographic origin and group according to the legend. Branch lengths are in expected substitutions per site. Estimated admixture proportions for K = 4 ancestral populations, based on the full dataset genotype likelihoods, with individual proportions aligned to the tips in the tree. For other values of K=2 to K=8, see **Supplementary Materials Online.**

Figure 2. Comparative analysis of ancient fruit phenotypes. **a)** Distribution of four phenotypes across tips in the inferred tree. Samples with no data for all traits were removed. **b-e)** Subtrees of the tips surrounding the inferred placement of our ancient samples, for **b)** fruit flesh color. **c)** fruit peel color. **d)** fruit acidity. **e)** fruit bitterness. **f-i)** Read pileups at loci known to control key phenotypes in *C. melo* subsp. *melo*: **f)** CmOR shows the Arg^{108} allele associated with non-orange flesh in GG1 and GG4. **g)** CmRPGE1 shows the non-truncated VED allele associated with green flesh in GG4. **h)** CmKFB lacks the 12bp deletion in the 5' UTR associated with white peel, suggesting it had yellow peel. **i)** CmPH shows some support for the 12bp duplication in CmPH associated with non-acidic fruit in GG4.

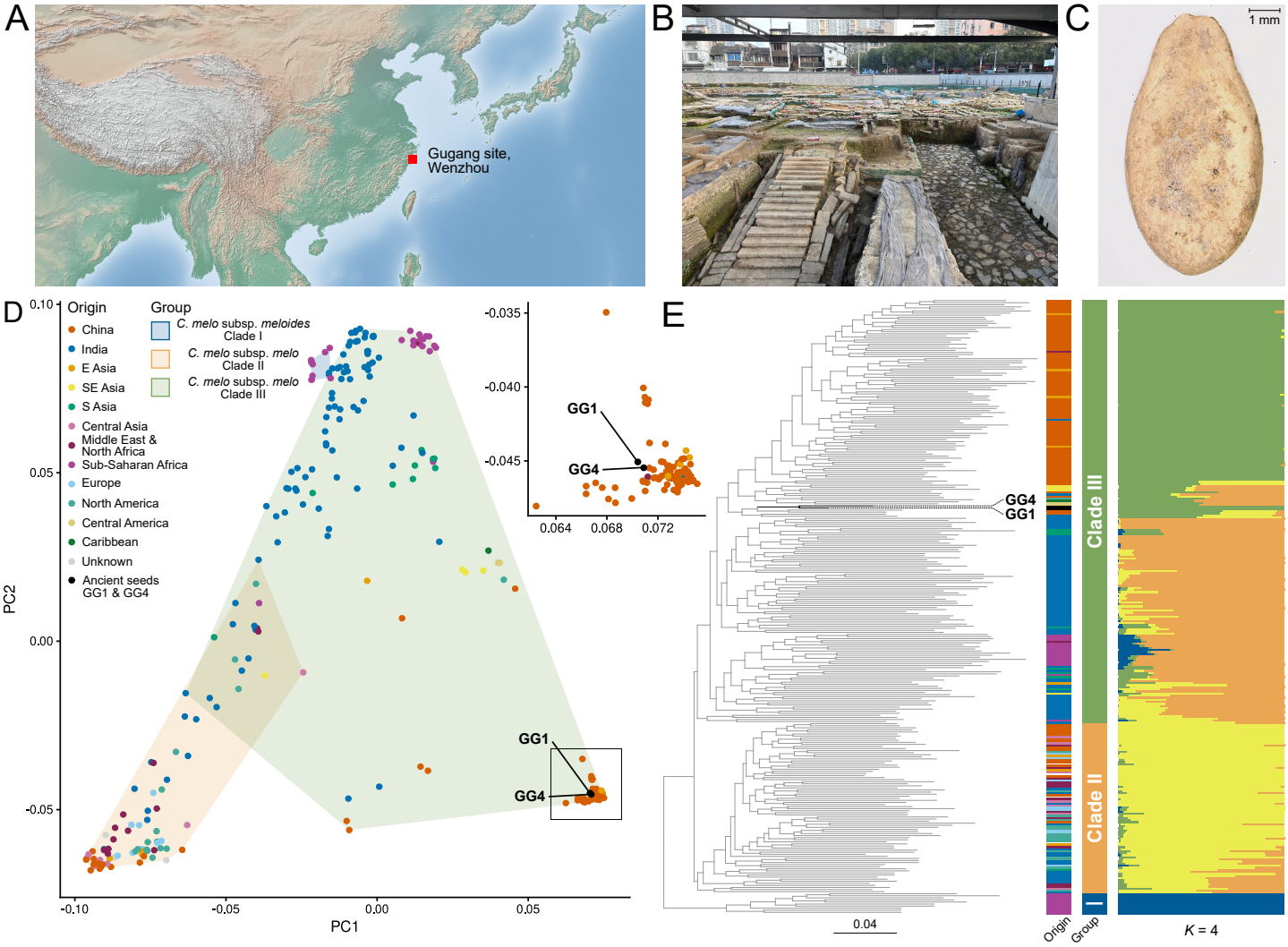

Figure 1. Phylogenomic placement of ancient seeds. a) Map of China and surrounding countries, showing the location of the Gugang archeological site in Wenzhou. **b)** The Gugang archeological site. **c)** Ancient melon seed recovered from the Gugang site. **d)** Principal Components Analysis based on inferred genotype likelihoods of ancient seed genomes in the context of a diverse collection of 294 extant melon genotypes. **e)** placement of the ancient genomes in a maximum likelihood phylogeny with the same collection of genotypes and 9 outgroups, based on 9,627 four-fold degenerate sites. Colors at tips indicate the geographic origin and group according to the legend. Branch lengths are in expected substitutions per site. Estimated admixture proportions for K = 4 ancestral populations, based on the full dataset genotype likelihoods, with individual proportions aligned to the tips in the tree. For other values of K=2 to K=8, see **Supplementary Materials Online.**

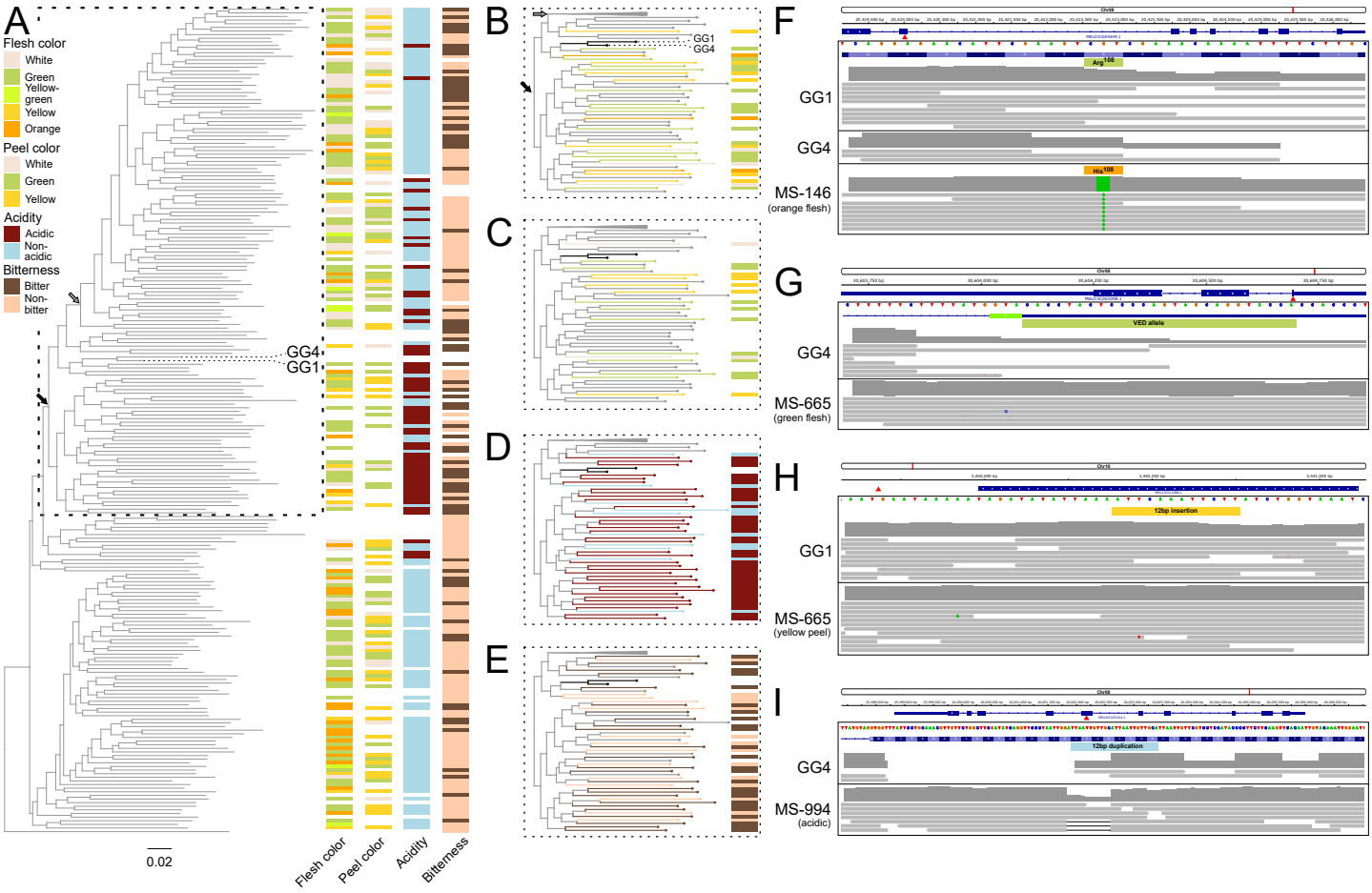

Figure 2. Comparative analysis of ancient fruit phenotypes. a) Distribution of four phenotypes across tips in the inferred tree. Samples with no data for all traits were removed. **b-e)** Subtrees of the tips surrounding the inferred placement of our ancient samples, for **b)** fruit flesh color. **c)** fruit peel color. **d)** fruit acidity. **e)** fruit bitterness. **f-i)** Read pileups at loci known to control key phenotypes in C. melo subsp. melo: **f)** CmOR shows the Arg^{108} allele associated with non-orange flesh in GG1 and GG4. **g)** CmRPGE1 shows the non-truncated VED allele associated with green flesh in GG4. **h)** CmKFB lacks the 12bp deletion in the 5' UTR associated with white peel, suggesting it had yellow peel. **i)** CmPH shows some support for the 12bp duplication in CmPH associated with non-acidic fruit in GG4.