

Genomic elucidation of hybridization between *Liberica* and *excelsa* coffee and its implications for coffee crop development

Received: 16 February 2026

Accepted: 14 April 2026

Published online: 14 May 2026

Cite this article as: Davis A.P., Shepherd-Clowes A., Lee K.W.T. *et al.* Genomic elucidation of hybridization between *Liberica* and *excelsa* coffee and its implications for coffee crop development. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-49305-5>

Aaron P. Davis, Amelia Shepherd-Clowes, Kenny Wee Ting Lee, Diana M. Jitam, Alasdair Clayre, Akshay Dashrath & Juan Viruel

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Title: Genomic elucidation of hybridization between Liberica and excelsa coffee and its implications for coffee crop development

Davis AP¹, Shepherd-Clowes A¹, Wee Ting Lee, K², Jitam D.M.³, Clayre A⁴, Dashrath A⁵, Viruel J^{1, 6, 7}

Author list:

Author	ORCID ID	e-mail address
*Aaron P. Davis	0000-0001-9213-4353	a.davis@kew.org
Amelia Shepherd-Clowes	0009-0002-9883-6908	a.shepherd-clowes@kew.org
Kenny Wee Ting Lee		kenny@earthlings-coffee.com
Diana M. Jitam	0009-0005-6561-3136	dianajitam@sarawak.gov.my
Alasdair Clayre	0000-0002-5202-0607	alsadairclayre@gmail.com
Akshay Dashrath	0009-0001-9104-6455	akshay@sicc.coffee
Juan Viruel	0000-0001-5658-8411	viruel@unizar.es

Affiliations:

- 1 Royal Botanic Gardens, Kew, UK
2. Earthlings Coffee, Kuching, Sarawak, Malaysia
3. Agriculture Research Centre Semonggok, KM20 Jalan Borneo Heights, 93250 Kuching, Sarawak, Malaysia.
4. University of Malaysia Sarawak, Institute of Borneo Studies, 94300 Kota Samarahan, Sarawak, Malaysia
5. SICC Labs, Suntikoppa, 571237, Kodagu District, Karnataka, India
6. Technological College. University of Zaragoza, Huesca, 22071, Spain
7. Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza. Mariano Esquillor, Edificio I+D. Zaragoza, 50018, Spain

ABSTRACT

The development and establishment of a broader range of coffee species and hybrids is likely to play a key role in coffee farming sustainability in an era of accelerated climate change. We investigated hybridization between *Coffea liberica* (Liberica) and *C. dewevrei* (excelsa) utilizing 7,618 SNPs from 113 accessions, sampled across three continents. Our analyses demonstrated that these two species have hybridized readily in cultivation and produced fertile progeny with a wide range of genomic admixture. We revealed extensive genomic admixture in farmed accessions from Sarawak. The identified hybrids exhibited intermediate characteristics and overlapping values for key agronomic traits, including seed size and parchment thickness. These traits, which increase yield and outturn and are likely to facilitate post-harvest processing, can be transferred via hybridization and are influenced by the proportion of genomic admixture. The hybrids also have the potential to broaden the climate envelope for successful coffee cultivation and perhaps transfer disease resistance. Improved genotypes resulting from hybrids between *C. liberica* and *C. dewevrei* could be brought into production relatively quickly, as a means of developing both species. We propose the formal name *Coffea* × *libex* for the interspecies hybrid.

KEY WORDS

Genomics, Hybrids, Excelsa Coffee, Liberica Coffee, Malaysia, SNPs

INTRODUCTION

The development and establishment of a broader range of coffee crop species and hybrids is likely to become a key requirement for achieving coffee farming sustainability in an era of accelerated climate change¹⁻³. Future coffee crops will need to be commercially viable across the value chain, operate under elevated temperatures, a broader range of annual precipitation patterns, including changes in timing and duration, and increasing frequency and severity of extreme weather events. Arabica (*Coffea arabica*) and robusta coffee (*C. canephora*) currently account for more than 99.99% of the world's coffee⁴, and are not expected to require substantial replacement or intervention at a global level over the short-term⁵. However, as climate change progresses, the development and deployment of alternative, climate-appropriate coffee crops for communities where Arabica and robusta farming are already faltering or failing. Given the limitations for agricultural operability of Arabica and robusta under certain climate parameters, an expanded and diversified coffee portfolio will need to include additional *Coffea* species, to introduce the traits (and numerous genes) that are required for meaningful climate resilience¹⁻³. Some diversification is already underway, with increased utilization of once more commonly grown species, including Liberica (*C. liberica*) and excelsa (*C. dewevrei*)⁶, Congo (*C. congensis*), and to a lesser extent Ibo/Zanzibar (*C.*

zanguebariae), *racemosa* (*C. racemosa*), and Nandi (*C. eugenioides*) coffees^{3,4,6-9}. Sierra Leone coffee (*C. stenophylla*) has recently attracted attention due to its ability to produce high-quality coffee, with an Arabica-like flavour, at substantially higher temperatures than Arabica, and under a longer dry season than robusta^{2,10,11}. Although *stenophylla* coffee is not yet established within the value chain, it has become a focal species for producers seeking a climate resilient alternative to *C. arabica*.

Interspecies hybrids, other than the well-known hybrid *C. arabica* × *C. canephora*, are generally rare in coffee farming. However, hybridization has been utilised in several breeding contexts. In India crosses and backcrosses involving *C. canephora* × *C. congensis* have been developed and are in use^{3,12,13}, as well as *C. arabica* × *C. liberica*¹⁴. In Brazil, breeding programmes have used *C. arabica* × *C. racemosa* hybrids¹⁵⁻¹⁷. Other interspecies hybrids have been reported, including *C. canephora* × *C. liberica*¹⁸, *C. liberica* × *C. eugenioides*^{19,20}, and *C. canephora* × *C. racemosa*^{12,21}.

Recent genomic and morphological research has clarified species delimitation within *C. liberica sensu lato*, demonstrating that it comprises three distinct species: *C. liberica* (Liberica), *C. dewevrei* (excelsa) and *C. klainei*⁴. Indigenous populations of *C. liberica* and *C. dewevrei* have discrete, geographical distributions and climate profiles, and possess distinct morphological and physiological traits relevant to crop development⁴. *Coffea dewevrei* possesses several attributes favourable for coffee production, including more flowers and fruits per branch/node, a smaller fruit, a thinner fruit pulp (mesocarp), thinner parchment (endocarp; i.e. the hard, crisp coating that surrounds the seed) and smaller Arabica-sized seeds, compared with *C. liberica*. These traits translate into higher fresh fruit yields, facilitate faster post-harvest processing, including reduced drying times, and improved outturns, relative to *C. liberica*. Outturn is defined as the conversion ratio of fresh fruit to clean coffee, which is a key factor in coffee farming profitability. Favourable traits for *C. liberica* include tolerance to high temperatures and successful cultivation across a broader range of precipitation regimes, including environments characterised by high rainfall and low elevation²². *Coffea liberica* may also tolerate longer and more severe dry periods than *C. dewevrei*^{4,22}. Both wild and cultivated *C. liberica* demonstrate substantially greater climatic tolerance compared with *C. arabica*, particularly with respect to temperatures and precipitation. This information suggests that hybridization between *C. liberica* and *C. dewevrei* may provide opportunities to combine favourable agronomic and climatic characteristics within a single crop lineage.

The aim of this study is to verify and quantify hybridization between *C. liberica* and *C. dewevrei*, and to provide an initial assessment of its potential value for coffee crop development. Although hybrids between these two species have been reported²³ and inferred⁴ based on morphology, their existence has not been confirmed by molecular data. Fruit and seed traits, specifically parchment (endocarp) thickness and seed dimensions, provide useful morpho-taxonomic markers for *C. liberica* and *C. dewevrei*⁴. Intermediate values for these characters have been observed in cultivated accessions, including farmed plants, suggesting the existence of interspecies hybrids. Hybridization has also been inferred in cultivated populations exhibiting high variability in leaf size

and shape; floral traits, including flower number (per node), flower size and shape, and corolla lobe number (per flower); and fruit size, shape and colour, and pulp (mesocarp) thickness. *Coffea liberica* and *C. dewevrei* do not overlap in their wild distributions⁴. Consequently, this study is focused on cultivated plants and farmed populations.

Our study has five main objectives: (1) to confirm the existence of hybridization between *C. liberica* and *C. dewevrei*; (2) to estimate the frequency of hybridization between these species; (3) to quantify levels of genetic admixture in hybrid individuals; (4) to assess the influence of hybridization on morphology, with emphasis on seed and parchment traits of agronomic relevance; and (5) to evaluate the potential usefulness of hybridization and introgression for the breeding and improvement of *C. liberica* and *C. dewevrei*.

To identify hybrids we employed genomic data generated using the Angiosperms353 target capture kit²⁴⁻²⁶ analysing 7,618 SNPs from the exon regions²⁷ across 113 accessions, sampled from three continents. The dataset includes predominantly cultivated material from farms and germplasm collections, with wild accessions used as reference standards. Morphological data on parchment (endocarp) thickness and seed dimensions were analysed to evaluate the phenotypic consequences of hybridization for these agronomic traits.

RESULTS

Angiosperms353 data recovery

Across all accessions, we recovered between 72.4 and 91.1% (mean 87.4%) of the Angiosperms353 genes length, representing 208,305–262,158 (mean 251,364) bp per sample. Overall, there were 2.22% missing data in the alignments.

SNP analyses

To investigate genetic structure and relationships between *C. liberica*, *C. dewevrei* and suspected hybrids we analysed 7,618 SNPs derived from exon regions²⁷ across 113 accessions (Supplementary Table 1). Evaluation of cluster number using ΔK supported $K = 2$ as the most likely value. STRUCTURE²⁸ results for $K = 2$ are shown in Fig. 1b, with corresponding admixture proportions provided in Supplementary Table S2. STRUCTURE results for $K = 3$ and $K = 4$ are presented in Supplementary Fig. S1; with admixture values given in Supplementary Table S3.

At $K = 2$, no genomic admixture was detected for wild-sourced accessions of either *C. liberica* or *C. dewevrei*. In contrast, genomic admixture was detected in 60 cultivated samples. Of these, 40 samples showed admixture proportions of 56.9–97.4 % from *C. liberica*, while 20 samples showed admixture proportions of 51.1–90.4 % from *C. dewevrei* (Supplementary Table S2). Six cultivated accessions identified as *C. dewevrei* showed low levels of admixture from *C. liberica*, ranging from 1–3%, including samples from Sierra Leone (1%

admixture, sample 23F68), Hawaii (1.1%, 23D58 and 3.4%, 23D59), India (1.9%, 24L56 and 1.3%, 24L82), and Costa Rica (2.2%, 23D49), as detailed in Supplementary Table S2. Similarly, three cultivated accessions identified as *C. liberica* showed low admixture from *C. dewevrei* ranging from 0.7–2%, originating from Dominica (1.9% admixture, sample 25C38), Cameroon (1.2%, 25C83), and Peninsula Malaysia (0.7%, 23D52).

Principal coordinate analysis (PCoA) (Fig. 1b) showed a clear separation between *C. liberica* and *C. dewevrei* along PC1, which explained 85.3% of the total variance. Admixture samples were positioned between the two parental clusters, with PC2 accounting for an additional 13% of variance. Two cultivated accessions of *C. liberica* from Dominica and Cameroon (25C38 and 25C83, respectively) were outliers relative to the otherwise tightly clustered *C. liberica* group, consistent with low levels of detected admixture.

A genetic distance-based phylogenetic tree reconstructed from pairwise genetic distances (Fig. 1a) supported the division of two principal groups. One comprised *C. dewevrei* and accessions with high proportions of *C. dewevrei* admixture, while the other comprised *C. liberica* and samples with high proportions of *C. liberica* admixture. Bootstrap support values across the tree were generally low.

SNP analyses for Malaysia

STRUCTURE²⁸ analysis revealed a particularly high incidence of *C. liberica*-*C. dewevrei* admixture among farmed *C. liberica* from Sarawak (Fig. 2). At $K = 2$, admixture proportions ranged from 2.8 to 59.9% *C. dewevrei*. Of the 45 Sarawak accessions labelled as *Liberica*, 40 showed more than 2.8% *C. dewevrei* admixture, and 28 exceeded 12% admixture, with a maximum of 59.90% admixture. Only five accessions were identified as genetically 'pure' *C. liberica* (< 1% *C. dewevrei* admixture). In contrast, of the eight accessions from Peninsular Malaysia, six were genetically 'pure' *C. liberica* (< 1% admixture), and two showed low admixture levels of 4.4% and 8.6% (Supplementary Table S4).

Outside Malaysia, admixture was also detected in cultivated accessions from multiple regions. Four accessions from Indonesia showed admixture values of 13.6%, 51.9%, 85.7% and 90.4% *C. dewevrei*. Nine accessions from India exhibited admixture values between 34.8–89.8%. Four accessions from Costa Rica showed admixture ranging from 64.1 to 78%, and a single accession from Uganda exhibited 53.4% *C. dewevrei* admixture.

Morphological analysis of seed and parchment

Seed size and parchment thickness differed markedly among *C. liberica*, *C. dewevrei*, and admixed accessions. For *C. dewevrei*, seed length and seed width measurements were 4.9–12.7 and 3.4–8.5 mm, respectively, with mean values of 8.4×6 mm. For *C. liberica*, seed length and seed width were 6.2–18.3 and 5.0–12.0 mm, with mean values of 11.7×7.7 mm. Admixed accessions spanned a broader range for seed length and seed width, of 3.4–16.7 and 4.7–10 mm, with

mean values of 10.5×7.1 mm. Parchment thickness measurements for *C. dewevrei* were 0.14–0.59 mm, with a mean of 0.27 mm. For *C. liberica*, parchment thickness was 0.43–0.77 mm, with a mean of 0.59 mm. Admixture accessions showed intermediate and overlapping values for parchment thickness of 0.18–0.77 mm, with a mean of 0.44 mm. One way ANOVA and TukeyHSD tests showed that *C. liberica*, *C. dewevrei* and *C. liberica* $\times$ *C. dewevrei* differed significantly for parchment thickness, seed length and seed width (Fig. 3, Supplementary Fig. S2, Supplementary Table S5 (with *p*-values given)).

Where corresponding genomic and morphological data were available, we plotted admixture proportions at $K = 2$ were against seed length, seed width and parchment thickness. Despite limited sample sizes, accessions with higher proportions of *C. dewevrei* admixture were associated with thinner parchment and smaller seeds (Fig. 4).

Other observations

Field observations in Sarawak indicated that genotypes with high levels of *C. dewevrei* admixture (i.e. 50–75%) displayed combinations of parental traits. These plants, typically, exhibited more flowers and fruits per node/axil, and thinner fruit pulp (mostly mesocarp), compared with *C. liberica*. Leaf morphology ranged from those indistinguishable from *C. liberica*, to very large leaves characteristic of *C. dewevrei*^A. Considerable variation was observed in flower size, shape and corolla lobe number, ranging from 5–12 lobes per flower. In *C. liberica*, individual flowers usually have 7–12 corolla lobes, and for *C. dewevrei* commonly 5 or sometimes 6. Fruit morphology among admixed plants was also highly variable (Fig. 5), including variation in fruit size, shape and colour. *Coffea liberica* fruits are typically red or rarely yellow, and *C. dewevrei* exhibits a wide range of colours including different shades and depths of yellow, orange, red, red-purple, and purple; highly admixed genotypes displayed especially diverse fruit colouration, including reds, oranges, and yellow, and purple (Fig. 5). Variation in nectary disc remains (the ‘woody’ area at the apex of the fruit) was observed across admixed and non-admixed populations alike. Large and small nectary disc remains are evident across admixture farm population, but this is also true of ‘pure’ wild and cultivated populations of *C. liberica* and *C. dewevrei*. Genotypes with low admixture of *C. dewevrei*, of around 10% or lower, tend to have a fruit and gross morphology that is identical to *C. liberica*, and vice versa.

Our measurement of 2,140 seeds (Supplementary Tables S7, S8) and observation of harvested seeds (coffee beans) across the full admixture spectrum, including those genotypes with around 50% admixture (at $K = 2$), showed negligible ($< 5\%$) to small percentages (10%, rarely 20%) of fruits with a single seed. These single seeds are rounded in shape, rather than flat on one side (the adaxial surface) and are usually referred to as peaberry or caracoli. Peaberry beans are formed due to the abortion of one of the ovules in the bi-locular ovary, resulting in a single, round or rounded, seed (coffee bean) per

fruit²⁹. In an average coffee crop the bearberry yield ranges from 10–30%, but can exceed 50%²⁹.

Discussion

Using genomic SNP data analysed with STRUCTURE²⁸ and principal coordinate analysis (PCoA) we confirmed the existence of hybridization between *C. liberica* and *C. dewevrei* (*C. liberica* × *C. dewevrei*). Admixture values of roughly equal proportions (e.g. 50–60%) likely represent F₁ hybrids, while values deviating from this expectation indicate subsequent generations, including backcrosses and introgression. Clarifying the precise generational structure and direction of these hybridization events would require controlled crossing experiments and additional genomic analyses. The wide geographic distribution of admixed accessions, including samples from Central America (Costa Rica), Africa (Uganda), India, Vietnam, and South-East Asia (Peninsular Malaysia, Sarawak, Indonesia), indicates that hybridization between *C. liberica* and *C. dewevrei* is not a localised phenomenon. This pattern suggests multiple independent hybridization events, although dissemination of hybridized seed and plants across continents is also a likely scenario.

Patterns of admixture varied markedly among regions. In Sarawak, admixture proportions were biased toward *C. liberica*, whereas in India the genomic contribution of *C. dewevrei* was generally higher. These differences may reflect founder effects, i.e., what genotypes were originally sent to farms and germplasm collections in receiving countries, the relative abundance of each species in cultivation, opportunities for backcrossing, and perhaps also selective pressures associated with local environments. In the Upper Baram Valley in Sarawak, the genomic data infers establishment using a mixture of ‘pure’ *C. liberica* and admixture genotypes of *C. liberica* × *C. dewevrei*. Another scenario is that *C. liberica* × *C. dewevrei* hybrids were introduced alongside already existing farmed populations of *C. liberica*. Even though the admixture proportions in Sarawak were biased towards *C. liberica*, plants of possible F₁ generation also exist. The origin of the seeds/plants of these coffees in the Upper Baram Valley has not been established. *Coffea liberica* was introduced to Malaysia multiple times from the 1870s onwards⁶ and throughout the latter part of the nineteenth century^{6,22}. There were also late nineteenth century introductions to many other parts of Southeast Asia^{6,22}. By contrast, *C. dewevrei* was not known to science until the end of the nineteenth century, and there are no records of dissemination to Malaysia during that period or in the early part of the twentieth century. However, *C. dewevrei* was distributed widely across Africa and to Vietnam during that time⁶. It should be noted that the climate of Sarawak is more conducive to *C. liberica*, compared with *C. dewevrei*^{4,22}, and thus selective environmental forces could have been acting in favour of *C. liberica*. Notably, ‘pure’ *C. dewevrei* has not been detected as a farmed plant in Peninsula Malaysia or Sarawak, via either morphological or genomic data, supporting the idea that hybrids were introduced rather than formed *in situ*. During the middle to latter part of the twentieth century coffee germplasm moved freely between the Americas, Africa and Asia due to institutional seed exchange programmes, and thus *C. liberica* × *C. dewevrei* or even *C. dewevrei* could have been introduced during this period. A more recent origin is also

possible, perhaps within the last few decades, from a neighbouring region. The Philippines is one possible option as both species are grown there with some frequency³⁰, although we also identified cultivated accession of *C. liberica* × *C. dewevrei* from Indonesia and Vietnam, providing other introduction origins. The likelihood of *in situ* hybridization in Sarawak cannot be ruled out with the data generated here. Given that *C. liberica* flowers throughout the year in Sarawak²², rather than over a specific period, opportunities for pollen transfer and thus outcrossing are certainly possible and even favoured. Wild hybrids between *C. liberica* and *C. dewevrei* are not a possibility, as the indigenous populations of these species are separated by at least 1000 km⁴ in Africa. In South India, admixture proportions were found to be biased towards *C. dewevrei*, possibly because this species is more frequent than *C. liberica* on farms in this part of India.

The mean values (and 1st and 3rd quartiles) for parchment thickness, seed length and seed width for *C. liberica* × *C. dewevrei* fall in-between those of the parental species (Fig. 3; Supplementary Tables S7, S8). Our preliminary analyses indicate that parchment thickness, seed length and seed width all decrease as the % admixture of *C. dewevrei* increases, and as *C. liberica* admixture % decreases (Fig. 4). Whilst the relationship between admixture percentage and these traits is not linear, it shows that introgression has a quantitative influence of agronomic traits. Morphological observation of cultivated plants in Sarawak with high percentages of *C. dewevrei* admixture (e.g. 50–75%) showed that they had a greater number of flowers and fruits per node/axil, and a thinner fruit pulp (mostly mesocarp), compared with *C. liberica*. *Coffea liberica* usually has 4–18 flowers per node (sometimes up to 20³¹), whereas *C. dewevrei* has 6–40; the pulp (mesocarp) of *C. liberica* is usually 4–9.5 mm thick, whereas in *C. dewevrei* it is 2–3.5 mm thick⁴. The leaves of the hybrid plants may be either undisguisable from *C. liberica*, or have much larger leaves, like *C. dewevrei*. We were unable to make assumptions as to the advantage of a specific leaf size; multilocation variety trials may help in this regard. Field observations also revealed that hybrid populations are substantially more morphologically diverse than either of the parental species, encompassing wide variation for vegetative, floral, and fruit traits. Such diversity is expected in recent hybrids and introgressed populations³² and contrasts with the more uniform appearance of genotypes with low levels of admixture (<10%), which tend to resemble their predominant genomic donor species. Importantly, even low admixture genotypes may retain quantitative contributions from the minor parent, e.g. for parchment thickness, seed length and seed width (Fig. 4).

Our data and observations demonstrate the potential for trait transfer from *C. dewevrei* to *C. liberica* via hybridization. Reduced parchment thickness and thinner fruit pulp directly increase outturn by reducing non seed (coffee bean) biomass per fruit, thereby improving the conversion of fresh fruit to clean coffee. This is particularly relevant for *C. liberica*, which is constrained by low outturns relative to other commercial coffees and *C. dewevrei*^{4,6}. Reported conversion ratios for *C. liberica* (Liberica) are 8.3:1–12.5:1^{33–35}; in practical terms this means that 8.3–12.5 kg of fresh fruit is required to produce 1 kg of clean coffee. This is unfavourable compared to the conversion ratios for *C. arabica* (Arabica), which has a conversion ratio of 5–6.25:1²⁹, and *C. dewevrei*

(excelsa), which has a outturn of 6:1–6.8:1^{33,34} or up to 8:1 and sometimes higher⁶. Improved outturn leads to increased yields per tree and per unit area, e.g. per acre or hectare. In addition to improving outturns, a thinner pulp and thinner parchment facilitate easier pulping and hulling and reduce drying times, thereby lowering labour and processing costs. If the coffee is transported in bulk as dried fruit (cherry) or parchment coffee, this would also mean reduced transport costs. Higher flower and fruit numbers per branch/node, as observed in hybrids with high *C. dewevrei* admixture, would also increase yield per tree and per unit area, compared with *C. liberica*. Collectively, these traits indicate that hybridization provides a practical route for improving the agronomic performance and economic viability of *C. liberica*.

Hybridization may also influence climatic suitability, compared with the parental species *C. dewevrei* and *C. liberica*. In Upper West Africa, *C. liberica* experiences a longer and more severe dry season with proportionally higher precipitation during the wetter/cooler months of the year, whereas much of the distribution range of *C. dewevrei* (in Central Africa) has a shorter dry season(s) with a more even annual distribution of precipitation⁴. The modelled mean annual temperature values for wild *C. liberica* and *C. dewevrei* are nearly identical, at 24.4 °C and 24.6 °C, respectively⁴. However, study of farmed *C. liberica*²² shows that it may be successfully cultivated over a greater range of temperature and precipitation values, compared with the species in the wild⁴. This includes a higher mean annual temperature of 25.7°C, and a very low precipitation seasonality (Bio15³⁶) which means that it has a near-consistent rainfall throughout the year. Reported Bio15 values for *C. liberica* are 45 for cultivated²² and 67 for wild⁴. A general, but frequent observation, is that *C. liberica* can be successfully farmed at low elevation locations (e.g. 10–100 m) and experience temperatures and precipitation values that would not be suitable for *C. dewevrei*, such as those in lowland Malaysia. Whilst the climate parameters for the successful cultivation of *C. dewevrei* have yet to be ascertained, it seems that this species is more suitable in slightly cooler climates with less precipitation seasonality. It is usually farmed at locations of 500–1500 m^{4,6}. Thus, it is possible that *C. liberica* × *C. dewevrei* hybrids may be appropriate for cultivation under climates where either of the parents may not be suitable. Adding a proportion of *C. liberica* genome to *C. dewevrei* may enable *C. dewevrei*-like plants to survive in warmer, lower elevation locations with longer dry seasons, for example. Long-term common garden experiments for two species of poplar (*Populus fremontii* and *P. angustifolia*), the F₁ hybrid, and backcross genotypes, showed that climate suitability traits can be transferred via hybridization in the F₁ and in backcrossed genotypes³⁷. Generally, introgressive hybridization has been important in the generation of new morphologies and adaptations to novel environments and climates compared with the parental species³². These factors need to be considered when identifying suitable environments for farming F₁ hybrids and introgressed genotypes of *C. liberica* × *C. dewevrei*, and in general when developing new, climate-suitable coffee crops. Multilocation (and multiclimate) field trials, or common garden experiments, need to be undertaken to elucidate climate suitability for *C. liberica* × *C. dewevrei*.

Further research is required to understand pests and disease tolerance and resistance in *C. liberica* and *C. dewevrei*, and *C. liberica* × *C. dewevrei*.

Tolerances and resistance could be gained, enhanced, or lost via introgression. Fungal pathogen resistance would be a key research focus, particularly for coffee leaf rust (*Hemileia vastatrix*), coffee wilt disease (*Gibberella xylarioides*), coffee berry disease (*Colletotrichum kahawae*) and various anthracnose species (*Colletotrichum* spp.). Early literature reported that *C. liberica* was not fully resistant to coffee leaf rust^{34,38,39,23,40,41}, although other reports stated that the rust itself caused little harm to *C. liberica*^{42,43}. The inconsistency in historical reports is likely due to, at least in part, to taxonomic and identification confusion for *C. liberica* and *C. dewevrei*. Unless the coffee species was identified with certainty, these reports must be regarded as unreliable. Observations made by us on farms in Sarawak, showed zero to negligible incidence of coffee leaf rust on *C. liberica* and *C. liberica* × *C. dewevrei*. At 12–566 m elevation, the farms in question experience high mean annual temperatures (25.7°C), high mean annual precipitation (3633 mm), low precipitation seasonality²² and year-round high humidity. It has been shown that a thermal regime (TR) of 27–22°C resulted in 2000 times higher sporulation of coffee leaf rust than with a TR of 23–18°C⁴⁴. The germination of coffee leaf rust urediniospores and subsequent penetration into the leaf via stomata is highly dependent on leaf surface wetness^{45,46}. Both of these conditions are prevalent and occur year-round in Sarawak, where *C. liberica* and *C. liberica* × *C. dewevrei* is grown²². *Coffea liberica* is the source of the gene SH3, which provides resistance to coffee leaf rust except for rust races with different combinations of virulence genes⁴⁷. Indeed, *C. liberica* has been used for the breeding of coffee leaf rust resistant *C. arabica* cultivars⁴⁸. Field observations by us indicate low to mild incidence of coffee leaf rust on *C. dewevrei* in Uganda, South Sudan, and India, but further research concerning coffee leaf rust resistance for this species is urgently required. In Sarawak we have observed a high incidence of anthracnose (*Colletotrichum* spp.) on *C. liberica* and *Coffea liberica* × *C. dewevrei*, which could be exacerbated by the warm, wet climate and high humidity⁴⁹. Farmer and producer feedback from Uganda and South Sudan (2018–2022) report either insignificant or zero susceptibility to coffee berry disease (*Colletotrichum kahawae*) for *C. dewevrei*⁶. Coffee wilt disease (*Gibberella xylarioides*) was first reported on cultivated excelsa (*C. dewevrei*) in the Central African Republic in 1927⁴⁹ and later caused widespread damage to farmed excelsa and robusta (*C. canephora*) across large areas of tropical Africa⁵⁰. Coffee wilt disease is still a major constraint for robusta production but it has not been reported as a disease of excelsa during recent field surveys in Uganda and South Sudan⁶.

The coffee flavour profile of *C. dewevrei* (excelsa) is dominated by notes of dark, dried fruits (currant, prune, plum, fig), chocolate (cacao nibs, dark and milk chocolate), and spices; the coffee usually has a low to medium acidity, low bitterness, and medium to high sweetness. The sweetness is akin to Demerara sugar, maple syrup, or stevia. The coffee flavour profile of *C. liberica* (Liberica) is usually dominated by bold tropical fruit flavours, with notes of jackfruit, mango, passion fruit, longan, and various green teas; the coffee has low acidity and high sweetness. The sweetness is akin to light honey, or white sugar. *Coffea liberica* × *C. dewevrei* is rare in the global coffee value chain and currently only available in Sarawak, where it is currently sold as *C. liberica*. The flavour of this coffee, as sampled from Sarawak, combines features of the parental species, with perhaps a more accessible flavour profile and an improved balance of fruit

and other flavour notes, compared with *C. liberica*. Our preliminary evaluations indicate that *Coffea liberica* × *C. dewevrei* hybrids could be of value for flavour, quality and diversity enhancements for both *C. liberica* and *C. dewevrei*. Rigorous sensory evaluation across multiple environments and processing methods will be required to substantiate these initial impressions.

A key observation from this study is the negligible (<5%) to low incidence (10–20%) of single-seeded fruits with rounded seeds, often known as peaberry or caracoli^{29,51}. A high proportion of peaberry (50% or more) is usually linked to genetic factors, possibly including pollen-stylar incompatibility⁵² and pairing complications during meiosis. This is a common feature in interspecies and intraspecies F1 and F2 coffee hybrids⁵¹. There may be other causes for the formation of peaberry, aside from genetic issues, such as poor plant health and unsuitable weather and climate conditions²⁹. The low incidence in *C. liberica* × *C. dewevrei* indicates genomic compatibility, probably due to the close evolutionary relationship between the two species⁴. A high proportion (50%, or more) of peaberry, as detected in some interspecies *Coffea* hybrids is considered as a major constraint in coffee breeding, especially when associated with low yields. High yields in interspecies coffee hybrids, suggest genomic compatibility, as found in the interspecies hybrid *C. canephora* × *C. congensis*^{13,53}. *Coffea liberica* × *C. dewevrei* and *C. canephora* × *C. congensis* hybrids are the result of crossing between closely related species^{4,54,55}. Reproductive compatibility, and the resulting low proportion of peaberry formation means that *C. liberica* × *C. dewevrei* hybrids could be readily deployed as crop plants.

Improved genotypes resulting from *C. liberica* × *C. dewevrei* hybrids could be brought into production relatively quickly by clonal propagation, via cuttings, micropropagation (*in vitro*), or by grafting on to selected rootstocks, such as *C. liberica* and *C. dewevrei*. Pure breeding lines are groups of individuals with reduced genetic variation that produce offspring via seed, with consistent traits representing uniform phenotypes. This might be an option for *C. liberica* × *C. dewevrei*, although it is unknown how many generations it would take to achieve a workable level of variation to produce a pure line cultivar derived from *C. liberica* × *C. dewevrei*.

To facilitate communication and recognition of this interspecies hybrid, we propose the formal name *C. × libex* for hybrids between *C. liberica* and *C. dewevrei*. This follows established conventions for naming interspecies hybrids{Turland, 2018 #718}, such as the interspecies hybrid *C. canephora* × *C. congensis* which is formally named as *C. × crameri* and also referred to as 'Congusta'. Cultivars arising from *C. × libex* could be designated by using a cultivar name placed in single quotes⁵⁶, as in the example of *Coffea* 'Congusta'. For the methodology employed here, we suggest an admixture cut-off of 5% for the recognition of *C. liberica* × *C. dewevrei* (*C. × libex*). For some of our data, and when using other genomic methods, such a small proportion of genomic admixture might represent noise but in some cases it could represent the transfer of meaningful genetic material and thus identifiable and useful traits.

METHODS

DNA sampling and sequencing

We sampled 113 accessions of *C. liberica*, *C. dewevrei*, and putative hybrids, sourced primarily from cultivated material, including farms and germplasm collections. Wild sourced accessions were included as reference standards. Of these, seven accessions of *C. liberica* and 20 of *C. dewevrei* had been unambiguously identified to species in a previous genomic morphological study, using a comprehensive sampling of related *Coffea* species⁴. Owing to the high frequency of suspected hybrids more rigorous sampling was undertaken in Sarawak. Sampling details, other accession information and sequence information are provided in Supplementary Table S1. Accepted botanical names and authorities follow the International Plant Names Index (<https://www.ipni.org>).

Total DNA was extracted from herbarium leaf tissue, silica gel dried leaves, or seeds using a modified CTAB protocol for herbarium specimens⁵⁷. Sequence target capture data were generated using the universal Angiosperms353 target capture kit developed to retrieve 353 nuclear genes across the angiosperms^{24,25}. The Angiosperms353 target capture kit, is a genomic tool that has been used to resolve relationships at various levels of taxonomic hierarchy in flowering plants^{26,27,58}, including at the population scale⁵⁹⁻⁶¹.

Genomic libraries were constructed using an optimized protocol⁶² for half volumes of the NEBNext Ultra II DNA Library Prep kit for Illumina (New England Biolabs) and purified using AMPure XP magnetic beads and multiplexed using NEBNext Multiplex Oligos for Illumina (Dual Index Primer Sets I and II). Pools containing 113 genomic libraries mixed in equimolar conditions were enriched with half reactions of the Angiosperms353 probe kit following the myBaits kit manual v.3.02 (Arbor Biosciences), using an optimized protocol⁶³. The DNA concentration and fragment size distribution were calculated using a Quantus fluorometer (Promega Corp.) and an Agilent 4200 TapeStation (Agilent Technologies), respectively. Sequencing was performed on a HiSeq (Illumina Inc.) by Macrogen, producing 2× 150 bp paired-end reads. Raw reads were submitted to the European Nucleotide Archive (<https://www.ebi.ac.uk>). ID codes are given in Supplementary Table S6.

SNP calling production and genomic analyses

Trimmomatic v.0.35⁶⁴ was used to discard low-quality reads and trim adaptors based on the reports generated by FastQC v.0.11.7⁶⁵ and HybPiper v.2.3.0⁶⁶ to retrieve the 353 nuclear loci using a combination of map to reference and de novo assembly methods for all samples.

The 113 accessions were divided into three groups, (1) *C. liberica* and (2) *C. dewevrei*, based on key morpho-taxonomic characters^{4,6}, and (3) suspected hybrids between these two species. To generate SNP data for these three groups, we used the framework developed by DePristo et al.⁶⁷ using GATK⁶⁸ following the pipelines established for Angiosperms353 data^{27,59}. This process combined aligned and unaligned reads to a reference built with the longest

exon obtained for all samples. We removed duplicate sequences and performed joint genotype calling for all samples after initially generating variants for each sample individually⁵³ in a variant call format (VCF) file. The initial VCF file was processed with a stringent filter ($QD < 5.0 \parallel FS > 60.0 \parallel MQ < 40.0 \parallel MQRankSum < -12.5 \parallel ReadPosRankSum < -8.0$), removing indels and SNPs with missing data using GATK and eliminating linked SNPs with PLINK⁶⁹. Base quality score recalibration was performed in GATK, followed by a repeated variant calling step.

To exclude the likelihood of genome admixture (via hybridization) from other coffee species, we first ran all 113 accessions against a Liberica alliance SNP dataset⁴, and an unpublished data set of multiple *Coffea* species, which showed that all accessions aligned with either *C. liberica* or *C. dewevrei* (results not shown).

To examine genetic differentiation patterns for the accession of *C. liberica*, *C. dewevrei*, and potential hybrids we used STRUCTURE²⁸ to determine genetic clusters in the dataset. The online tool, StructureSelector⁷⁰ was used to calculate the most likely value of K for various estimators, including ΔK (REF). A K value of $K = 2$ was manually set based on the optimal recovered K value and the number of species; $K = 3$ and $K = 4$ were also examined. Each K was analysed with ten replicates, using 100,000 burn-in iterations followed by 1,000,000 Markov chain Monte Carlo repetitions. The most likely number of clusters was identified using PLINK outputs and were converted into STRUCTURE-compatible files using PGDSpider⁷¹. The results from STRUCTURE were visualized with StructRly⁷² (Fig. 1b, Supplementary Fig. S1). Genetic differentiation was explored using a genetic distance-based phylogenetic tree (Fig. 1a) using upgma with poppr 2.9.6, adegenet, ape 5.7.1 and MetBrewer⁷³ (Fig. 2) and PCoA in R 4.3.0⁷⁴, with the adegenet and ggplot2 packages, retaining three principal components to investigate genetic groupings (Fig. 1c).

Morphological analysis of seed and parchment

Parchment (endocarp) thickness and seed size values provide useful characters for the differentiation of *C. liberica* and *C. dewevrei*⁴. Reported values for parchment thickness are: *C. liberica* 0.36–0.77 mm (mean value 0.57 mm) and *C. dewevrei* 0.22–0.41 mm (mean value 0.31 mm). Seed length and width for *C. liberica* is 9.5–18.3 and 6.5–12 mm, respectively (mean values 12.6×8.4 mm) and *C. dewevrei* 7.7–11.3 and 5.4–8 mm, respectively (mean values 9.3×6.6 mm)⁴. We measured the parchment thickness of 477 samples (29 for *C. liberica*, 244 for *C. dewevrei*, and 204 for the putative hybrid accessions); accession details are provided in Supplementary Table S8. We measured the seed length and seed width of 2,140 seeds (467 for *C. liberica*, 781 for *C. dewevrei*, and 892 for the hybrid accessions); accession details are given in Supplementary Table S8. Measurements of parchment thickness was made using a Mitutoyo 193-111 0–25 mm (0.001 mm) micrometer, and for seed length and width a 0–150 mm (0.01 mm) Mitutoyo 500-196-30 digital Vernier calliper. One way ANOVA was used to examine the variance for each of the three characters, followed by a TukeyHSD test to identify if the means for specific

pairs, of the three study groups, were significantly different. Box plots for parchment thickness, seed length, and seed width are given in Fig. 3; a scatter plot of seed length vs. seed width is given in Supplementary Fig. S2. Descriptive statistics and the results of the statistical tests are given in Supplementary Tables S5. Parchment thickness, seed length and seed width were plotted against admixture % (Fig. 4) to determine the relationships between these physical traits and admixture.

ARTICLE IN PRESS

References

- 1 Davis, A. P. *et al.* High extinction risk for wild coffee species and implications for coffee sector sustainability. *Science Advances* **5**, eaav3473 (2019). <https://doi.org/10.1126/sciadv.aav3473>
- 2 Davis, A. P., Mieulet, D., Moat, J., Sarmu, D. & Hagggar, J. Arabica-like flavour in a heat-tolerant wild coffee species. *Nature Plants* **7**, 413–418 (2021). <https://doi.org/10.1038/s41477-021-00891-4>
- 3 Bertrand, B., Davis, A. P., Maraval, I., Forestier, N. & Mieulet, D. Potential beverage quality of three wild coffee species (*Coffea brevipes*, *C. congensis* and *C. stenophylla*) and consideration of their agronomic use. *Journal of the Science of Food and Agriculture* **103**, 3602–3612 (2023). <https://doi.org/10.1002/jsfa.12347>
- 4 Davis, A. P. *et al.* Genomic data define species delimitation in Liberica coffee with implications for crop development and conservation. *Nature Plants* **11**, 1729–1738 (2025). <https://doi.org/10.1038/s41477-025-02073-y>
- 5 Kath, J. *et al.* Vapour pressure deficit determines critical thresholds for global coffee production under climate change. *Nature Food* **3**, 871–880 (2022). <https://doi.org/10.1038/s43016-022-00614-8>
- 6 Davis, A. P., Kiwuka, C., Faruk, A., Walubiri, M. J. & Kalema, J. The re-emergence of Liberica coffee as a major crop plant. *Nature Plants* **8**, 1322–1328 (2022). <https://doi.org/10.1038/s41477-022-01309-5>
- 7 Davis, A. P. *et al.* Hot coffee: the identity, climate profiles, agronomy, and beverage characteristics of *Coffea racemosa* and *C. zanguebariae*. *Front. Sust. Food Syst.* **5**, 740137 (2021). <https://doi.org/10.3389/fsufs.2021.740137>
- 8 Davis, A. P., Kiwuka, C., Faruk, A., Mulumba, J. W. & Kalema, J. A review of the indigenous coffee resources of Uganda and their potential for coffee sector sustainability and development. *Front. Pl. Sci.* **13**, 1057317 (2023). <https://doi.org/10.3389/fpls.2022.1057317>
- 9 Davis, A. P. *et al.* The wild coffee resources of Uganda: a precious heritage. *Royal Botanic Gardens, Kew*, 1–44 (2023). <https://doi.org/10.34885/fdtw-m431>
- 10 Davis, A. P., Gargiulo, R., Fay, M. F., Sarmu, D. & Hagggar, J. Lost and found: *Coffea stenophylla* and *C. affinis*, the forgotten coffee crop species of West Africa. *Front. Plant Sci.* **11** (2020). <https://doi.org/10.3389/fpls.2020.00616>
- 11 Jan-Smith, E. *et al.* Metabolomic insights into the Arabica-like flavour of stenophylla coffee and the chemistry of quality coffee. *npj Sci. Food* **9**, 33 (2025). <https://doi.org/10.1038/s41538-025-00398-8>
- 12 Sureshkumar, V. B., Selvakumar, M., Santaram, A. & Devasia, J. Performance evaluation of interspecific hybrids of *Coffea racemosa* and Sln. 35 (*Coffea congensis* × *Coffea canephora*) under Indian conditions. *Proceedings of the 20th International Conference on Coffee Science, 11–15 October 2004, Bangalore.* (2004).
- 13 Bharatha Nandhini, R. M., Rahul, R. N., Thilaga, S., Surya Prakash Rao, N. & Ganesh, D. Molecular distinction of C×R hybrid (*Coffea congensis*×*Coffea canephora*) from morphologically resembling male parent using *rbcl* and *matK* gene sequences. *S. African J. Bot.* **88**, 334–340 (2013). <https://doi.org/10.1016/j.sajb.2013.08.011>
- 14 Surya Prakash, N., Combes, M.-C., Somanna, N. & Lashermes, P. AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica* **124**, 265–271 (2002). <https://doi.org/10.1023/A:1015736220358>

- 15 Guerreiro Filho, O. *Coffea racemosa* Lour. a review. *Café, Cacao, Thé* **36**, 171–186 (1992).
- 16 Maciel, A. L. d. R., Rodrigues, F. A., Pasqual, M. & de Carvalho, C. H. S. Acclimatization of coffee (*Coffea racemosa* × *Coffea arabica*) somaclones obtained from temporary immersion bioreactor system (RITA®). *Austr. J. Crop Sci.* **10**, 169–175 (2016).
- 17 Medina Filho, H. P., Carvalho, A. & Medina, D. M. Germoplasma de *C. racemosa* e seu potencial no melhoramento do cafeeiro. *Bragantia* **36**, XLIII–XLVI (1977).
- 18 Chinnappa, C. C. Interspecific hybrids of *Coffea canephora* and *C. liberica*. *Genetica* **41**, 141–146 (1970). <https://doi.org/10.1007/BF00958900>
- 19 Ganesh, D. *et al.* Evaluation of *Coffea liberica* × *Coffea eugenoides* and its progenies for yield, leaf rust tolerance and quality. *Proceedings of the 15th Plantation Crops Symposium Placrosym XV, Mysore, India, 10–13 December, 2002*, 72–77 (2002).
- 20 Reddy, A. G. S., Raju, K. V. V. S. N. & Dharmaraj, P. S. Breeding behaviour of Ligenioides — a spontaneous amphidiploid between *Coffea liberica* and *C. eugenoides* *J. Coffee Res.* **15**, 33–37 (1985).
- 21 Sureshkumar, V. B., Prakash, N. S. & Mohanan, K. V. A study of *Coffea racemosa* × *Coffea canephora* var. *robusta* hybrids in relation to certain critically important characters. *Int. J. Plant Breed. Genet.* **4**, 30–35 (2010). <https://doi.org/10.3923/ijpb.2010.30.35>
- 22 Wild, I., Davis, A. P., Wee Ting Lee, K. & Moat, J. Climate requirements for cultivated Liberica coffee (*Coffea liberica*) and consequences for its use and development as a crop species. *Plants People Planet* (2026). <https://doi.org/doi.org/10.1002/ppp3.70183>
- 23 Cramer, P. J. S. *A Review of Literature of Coffee Research in Indonesia*. (Turrialba: SIC Editorial, Inter-American Institute of Agricultural Sciences, 1957).
- 24 Johnson, M. G. *et al.* A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* **68**, 594–606 (2018). <https://doi.org/10.1093/sysbio/syy086>
- 25 Baker, W. J. *et al.* A comprehensive phylogenomic platform for exploring the Angiosperm Tree of Life. *Systematic Biology* **71**, 301–319 (2021). <https://doi.org/10.1093/sysbio/syab035>
- 26 Baker, W. J. *et al.* Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. *Amer. J. Bot.* **108**, 1059–1065 (2021). <https://doi.org/10.1002/ajb2.1703>
- 27 Slimp M, Williams L.D., Hale H & Johnson M.G. On the potential of Angiosperms353 for population genomic studies. *Appl. Plant Sci.* **9**, e11419 (2021). <https://doi.org/10.1002/aps3.11419>
- 28 Pritchard J. K., Stephens M & Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000). <https://doi.org/10.1093/genetics/155.2.945>
- 29 Wintgens, J. N. Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers. *WILEY-VCH Verlag GmbH & Co. KGaA. Pp 976* (2004).
- 30 Baltazar, M. D. & Buot, I. E. Leaf architectural analysis of taxonomic confusing coffee species: *Coffea liberica* and *Coffea liberica* var. *dewevrei*. *Biodiversitas* **20**, 1560–1567 (2019).
- 31 Mukhoyyaroh, N. I., Azrianingsih, R. & Luchman Hakim, L. Ecological factors and morphological traits of *Coffea liberica* Hiern in Kalipuro District,

- Banyuwangi Regency. *Plant Sci. Today* **11**, 527–536 (2024).
<https://doi.org/10.14719/pst.2939>
- 32 Grant, P. R. & Grant, B. R. Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences* **116**, 23216–23224 (2019).
<https://doi.org/10.1073/pnas.1913534116>
- 33 Wellman, F. L. *Coffee: Botany, Cultivation and Utilization*. (London: Leonard Hill [Books] Limited, 1961).
- 34 Cramer, P. J. S. Gevens over de variabiliteit van de in Nederlandsch-Indië verbouwde koffie-sorten. *Mededeelingen uitgaande van het Departement van Landbouw*. **1**, 1–696 (1913).
- 35 Ghawas, M. Yield performance and selection of potential Liberica coffee clones. *J. Trop. Agric. Fd. Sc.* **34**, 1–6 (2006).
- 36 Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 170122 (2017).
<https://doi.org/10.1038/sdata.2017.122>
- 37 Hord, A. M. *et al.* Hybrid introgression as a mechanism of rapid evolution and resilience to climate change in a riparian tree species. *Commun. Biol.* **8**, 1173 (2025). <https://doi.org/10.1038/s42003-025-08410-3>
- 38 McCook, S. *Coffee Is Not Forever. A Global History of the Coffee Leaf Rust*. (Ohio University Press, 2019).
- 39 Burkill, I. H. *A Dictionary of the Economic Products of the Malay Peninsula. Volume 1 (A–H)*. (Oxford: University Press, 1935).
- 40 Anon. Coffee cultivation in Java. *Trop. Agriculturist* **76**, 4 (1931).
- 41 McCook, S. Ephemeral Plantations: The Rise and Fall of Liberian Coffee, 1870–1900. In: Comparing Apples, Oranges, and Cotton. Environmental Histories of the Plantation (ed. Uekötter, F.). Frankfurt/New York: Campus Verlag. 85–112 (2014).
- 42 Cheney, R. H. *A Monograph of the Economic Species of the Genus Coffea L.* (The New York University Press, 1925).
- 43 Hooker, J. D. Report on the progress and condition of the Royal Gardens at Kew, during the year 1876. Interchange of plants and seeds. *Kew Reports*, 7–26 (1877).
- 44 Toniutti, L. *et al.* Influence of environmental conditions and genetic background of Arabica coffee (*C. arabica* L.) on leaf rust (*Hemileia vastatrix*) pathogenesis. *Front. Pl. Sci.* **8**, 2025 (2017).
<https://doi.org/10.3389/fpls.2017.02025>
- 45 Nutman, F. J., Roberts, F. M. & Clarke, R. T. Studies on the biology of *Hemileia vastatrix* Berk. & Br. *Transactions of the British Mycological Society* **46**, 27–44 (1963). [https://doi.org/10.1016/S0007-1536\(63\)80005-4](https://doi.org/10.1016/S0007-1536(63)80005-4)
- 46 Bebbber, D. P., Castillo, Á. D. & Gurr, S. J. Modelling coffee leaf rust risk in Colombia with climate reanalysis data. *Phil. Trans. Roy. R. Soc. B, Biol. Sci.* **371**, 20150458 (2016). <https://doi.org/10.1098/rstb.2015.0458>
- 47 Rodrigues Jr, C. J. & Eskes, A. B. in *Coffee: Growing, Processing, Sustainable Production* (ed J.N. Wintgens) 557–570 (Wiley-VCH, 2012).
- 48 Narasimhaswamy, R. L. Arabica selection S.795 - its origin and performance - a study. *Indian Coffee* **24**, 197–204 (1960).
- 49 Gaitán, A. L., Cristancho, M. A., Castro Caicedo, B. L., Rivillas, C. A. & Cadena Gómez, G. *Compendium of Coffee Diseases and Pests*. (APS Press, 2015).
- 50 Rutherford, M. A. Current knowledge of coffee wilt disease, a major constraint to coffee production in Africa. *Phytopathology* **96**, 663–666 (2006). <https://doi.org/10.1094/PHYTO-96-0663>

- 51 Carvalho, A. & Monaco, L. C. Relaciones geneticas de especies seleccionadas de *Coffea*. *Cafe* **9**, 3–19 (1968).
- 52 Wrigley, G. *Coffee – Tropical Agriculture Series*. 369 (Longman Scientific & Technical, 1988).
- 53 Poplin R *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* **201178** (2018). <https://doi.org/10.1101/201178>
- 54 Davis, A. P., Tosh, J., Ruch, N. & Fay, M. F. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot. J. Linn. Soc.* **167**, 357–377 (2011). <https://doi.org/10.1111/j.1095-8339.2011.01177.x>
- 55 Charr, J.-C. *et al.* Complex evolutionary history of coffees revealed by full plastid genomes and 28,800 nuclear SNP analyses, with particular emphasis on *Coffea canephora* (robusta coffee). *Mol. Phyl. Evol.* **151**, 106906 (2020). <https://doi.org/10.1016/j.ympev.2020.106906>
- 56 Turland, N. J. *et al.* *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. (Glashütten: Koeltz Botanical Books, 2018).
- 57 Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin of the Botanical Society of America* **19**, 11–15 (1987).
- 58 Zuntini, A. R. *et al.* Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024). <https://doi.org/10.1038/s41586-024-07324-0>
- 59 Liang, J. K. *et al.* Conservation implications for the Iberian narrowly endemic *Androsace cantabrica* (Primulaceae) using population genomics with target capture sequence data. *Ecol. Evol.* **15**, e71901 (2025). <https://doi.org/10.1002/ece3.71901>
- 60 Viruel, J. *et al.* Phylogenomic barcoding of soil seed bank-persistent and wind-dispersed non-native plant species in South Georgia. *Mol. Ecol. Resour.* **26**, e70068 (2026). <https://doi.org/doi.org/10.1111/1755-0998.70068>
- 61 Razanamaro, O. H. M. *et al.* Integrating genomics and habitat surveys to uncover population structure and regeneration challenges in *Adansonia suarezensis* (Malvaceae). *Ann. Bot.* (2025). <https://doi.org/10.1093/aob/mcaf320>
- 62 Viruel, J. *et al.* A target capture-based method to estimate ploidy from herbarium specimens. *Front. Plant Sci.* **10** (2019). <https://doi.org/10.3389/fpls.2019.00937>
- 63 Brewer, G. E. *et al.* Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Front. Plant Sci.* **10** (2019). <https://doi.org/10.3389/fpls.2019.01102>
- 64 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). <https://doi.org/10.1093/bioinformatics/btu170>
- 65 Andrews, S. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (Accessed on 7 June 2024). (2010).
- 66 Johnson, M. G. *et al.* HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* **4**, 1600016 (2016). <https://doi.org/10.3732/apps.1600016>

- 67 DePristo M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011). <https://doi.org/10.1038/ng.806>
- 68 McKenna A *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). <https://doi.org/10.1101/gr.107524.110>.
- 69 Chang C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742–13015–10047–13748 (2015). <https://doi.org/10.1186/s13742-015-0047-8>
- 70 Li, Y.-L. & Liu, J.-X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Res.* **18**, 176–177 (2018). <https://doi.org/10.1111/1755-0998.12719>
- 71 Lischer, H. E. L. & Excoffier, L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012). <https://doi.org/10.1093/bioinformatics/btr642>
- 72 Criscuolo N.G. & Angelini C. StructuRly: A novel shiny app to produce comprehensive, detailed and interactive plots for population genetic analysis. *PLoS ONE* **15**, e0229330 (2020). <https://doi.org/10.1371/journal.pone.0229330>
- 73 Mills, B. R. MetBrewer: Color Palettes Inspired by Works at the Metropolitan Museum of Art. R package version 0.2.0. <https://CRAN.R-project.org/package=MetBrewer> (2022).
- 74 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL <https://www.R-project.org/> (2025).

Acknowledgements

We thank the individuals and organizations who contributed cultivated and trade samples of *Liberica* and *excelsa* coffee, including A. Kawabata, the Department of Tropical Plant and Soil Sciences, Kealahou, Hawaii; G. Hernandez, at *Coffea diversa*, Costa Rica; D. Sarmu, Coffee Culture, Sierra Leone; Slow Foods, Italy; E. Selam, House of Kendal, Malaysia; and J. Liew, MyLiberica, Malaysia. We thank the coffee growers of Sarawak who allowed access to their farms and coffee samples. We thank the herbarium curators at the herbaria listed herein, and E. Messenger at Kew's Economic Botany Collection for access to collections for DNA sampling.

Author contributions

A.P.D., A.S.-C. and J.V. designed the study. DNA, parchment and seed sampling was undertaken by A.P.D., A.S.-C., A.D., K.W.T.L., A.C. and D.M.J. DNA extractions and genomic library preparations were performed by A.S.-C. Pre-analysis bioinformatics, phylogenomic analyses, SNP production and analyses were undertaken A.S.-C. and J.V. Assembly and assessment of morphological and distribution data was made by A.P.D., A.D., A.S.-C., K.W.T.L., A.C. and D.M.J. The paper was written by A.D., A.S.-C. and J.V. All authors contributed to manuscript revisions and approved the paper. A.P.D. and A.S.-C. contributed equally to the study.

Funding statement

Funding was provided by the Calleva Foundation and the Amar-Franes Foster-Jenkins Trust. J.V. received funding from RYC2023-042611-I by MCIU/AEI/10.13039/501100011033 and FSE+.

Data availability

Raw reads for Angiosperms353 sequence data are available at the European Nucleotide Archive (<https://www.ebi.ac.uk>) under project no. [to update, post review]; ID codes are given in Supplementary Table S6.

Additional Information

The authors declare no competing interests.

Supplementary Information

Supplementary figures

Supplementary Fig. S1. Alternative STRUCTURE K values for *C. liberica*, *C. dewevrei* and *C. liberica* × *C. dewevrei*.

Supplementary Fig. S2. Seed length versus seed width for *C. liberica*, *C. dewevrei* and *C. liberica* × *C. dewevrei*.

ARTICLE IN PRESS

Supplementary tables

Table S1. List of DNA samples with key accession data and sequencing details.

Table S2. Admixture percentages for $K = 2$.

Table S3. Admixture percentages for $K = 3$ and $K = 4$.

Table S4. Admixture percentages $K = 2$ for Malaysian accessions.

Table S5. Descriptive statistics and details of ANOVA and TurkeyHSD test for parchment thickness, seed length, and seed width.

Table S6. European Nucleotide Archive ID and file codes.

Table S7. List of parchment samples with accession details, DNA codes, admixture percentages ($K = 2$) and parchment thickness measurements.

Table S8. List of seed samples with accession details, DNA codes, admixture percentages ($K = 2$) and seed measurements.

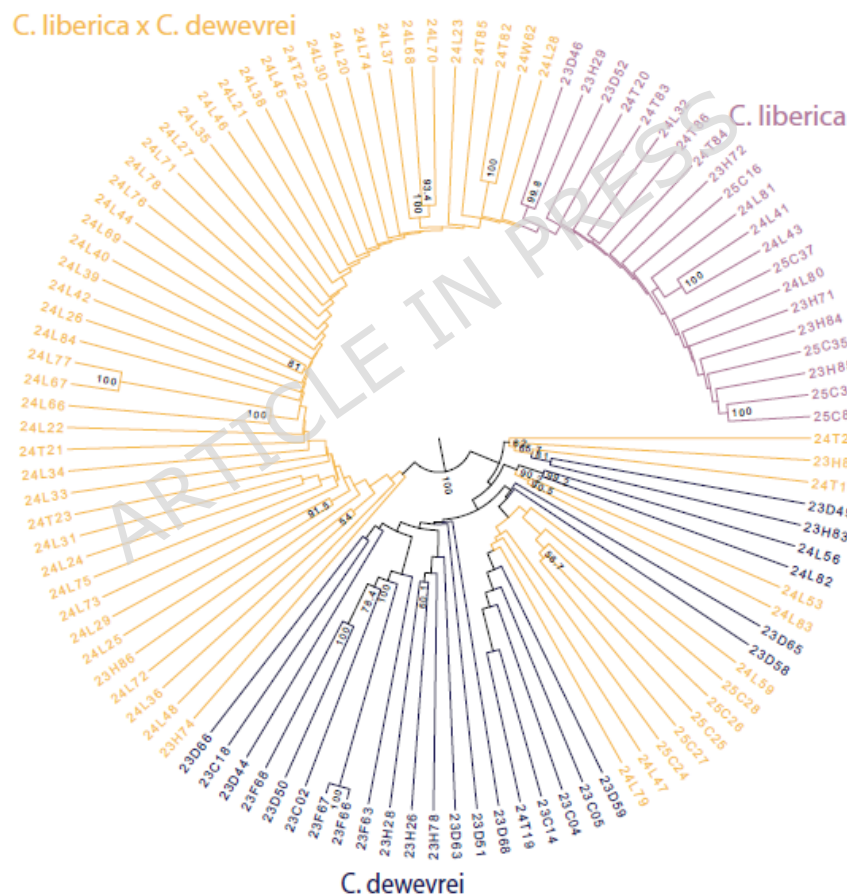
ARTICLE IN PRESS

Figures and figure legends

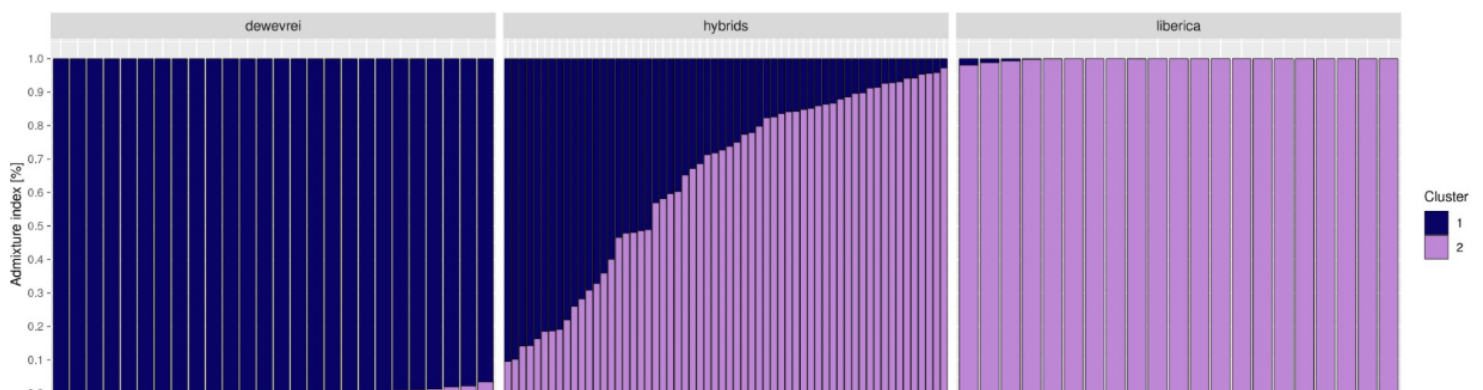
Fig. 1. Genetic structure and relationships between and within *C. liberica* (Liberica), *C. dewevrei* (excelsa) and *C. liberica* × *C. dewevrei*. Based on 7,618 exon region SNPs.

a, A genetic distance phylogenetic tree reconstructed with pairwise genetic distances (the proportion of loci that are different). The figures above branches indicate the BS values (BS values <50 not shown). **b**, STRUCTURE²⁸ analysis with the *K* value set to *K* = 2 to match the number of parental species, representing two clusters (*C. liberica*, *C. dewevrei*) and admixture accessions (hybrids). See Supplementary Fig. S1 for alternative *K* value analyses, and Supplementary Table S3 for *K* = 3 and *K* = 4 admixture values. **c**, PCoA analysis. PC1 (85.3% variance) separates the two parental species, with intermediate admixture accessions (hybrids). Accession information is provided in Supplementary Table S1.

(a)



(b)



(c)

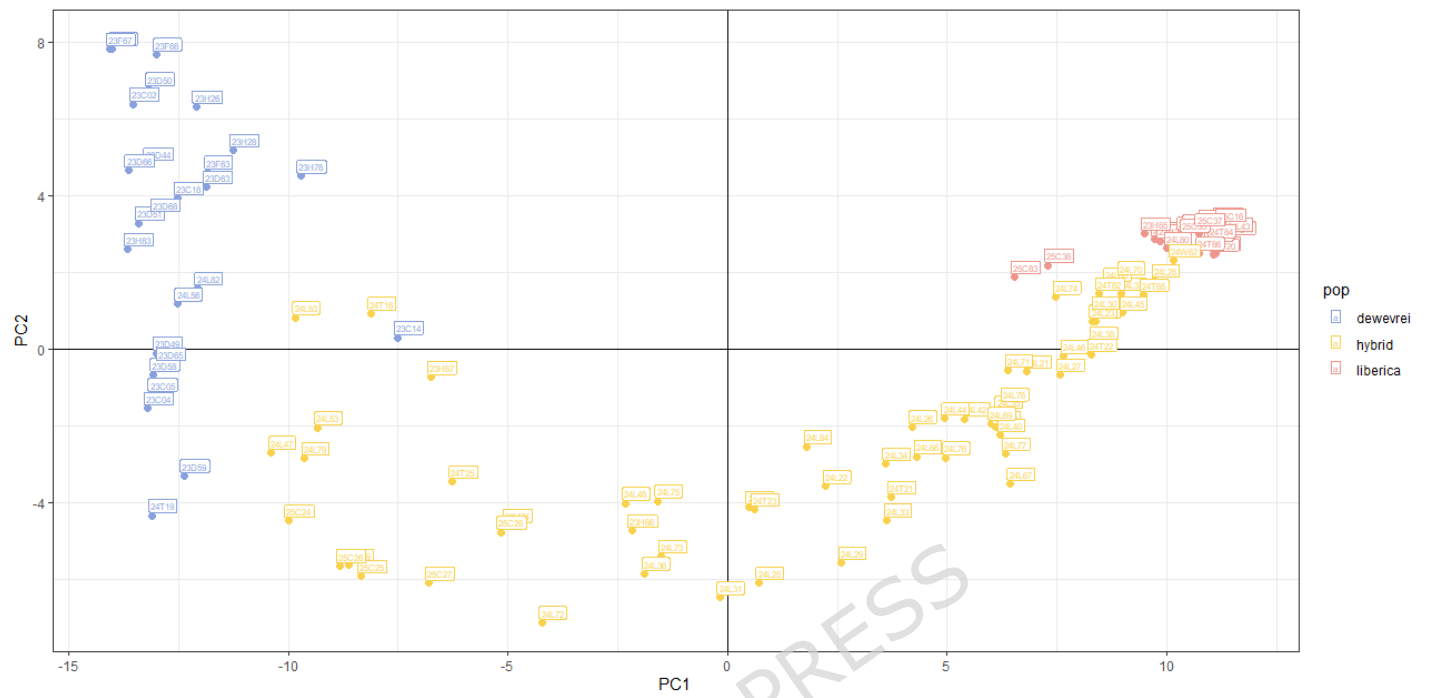


Fig. 2. Admixture pie-charts for *C. liberica* (Liberica), *C. dewevrei* (excelsa) and *C. liberica* × *C. dewevrei*, for Vietnam, and Malaysia accessions. Admixture based on STRUCTURE²⁸ analysis at $K = 2$ (see Fig. 1b) for 26 samples from Malaysia, with one representative sample of *C. dewevrei* from Vietnam. See Supplementary Tables S1, S4 for accession information, and Supplementary Table S2 for $K = 2$ values.

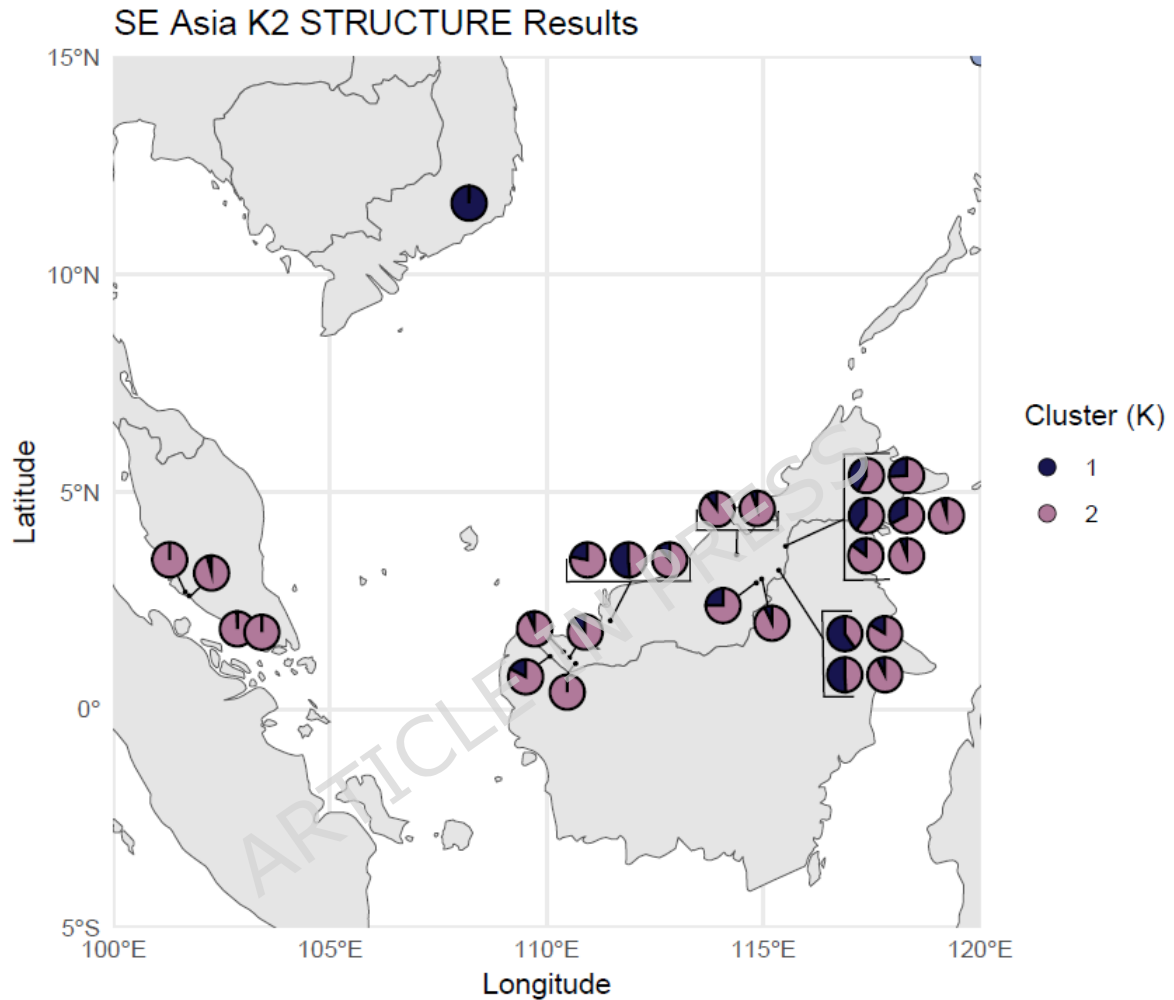
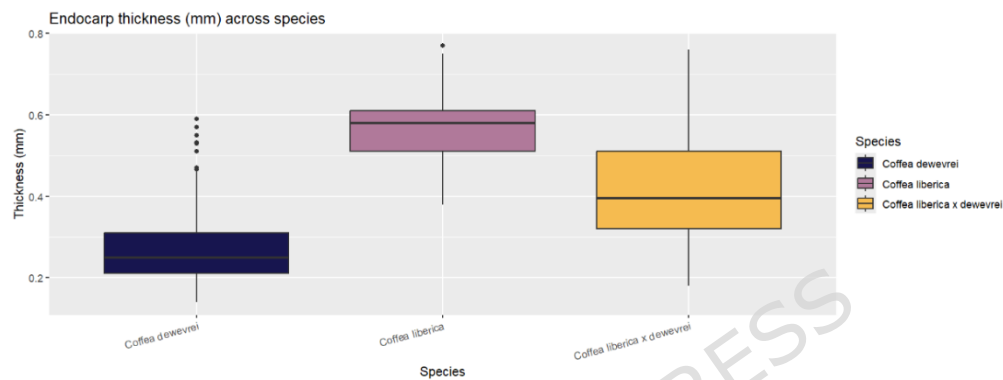
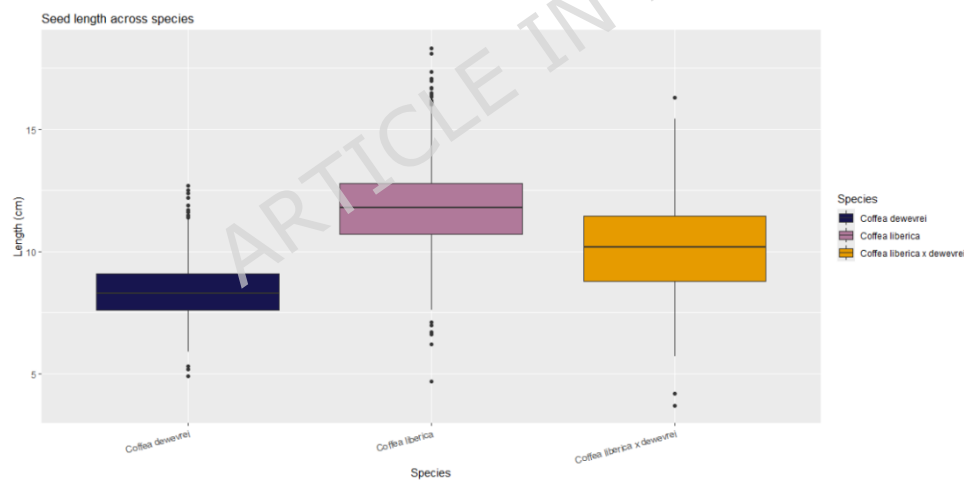


Fig. 3. Box and whisker plots for parchment (endocarp) thickness and seed length and width for *C. liberica* (Liberica), *C. dewevrei* (excelsa) and *C. liberica* × *C. dewevrei*. Mean values in parentheses; quartile values are given in Supplementary Table S5 .**a**, Parchment thickness (mm): *C. liberica* (0.59 mm), *C. dewevrei* (0.27 mm) and *C. liberica* × *C. dewevrei* (0.44 cm). **b**, Seed length (mm): *C. liberica* (11.85 mm), *C. dewevrei* (8.39 mm) and *C. liberica* × *C. dewevrei* (10.27 mm). **c**, seed width (mm): *C. liberica* (7.51) mm), *C. dewevrei* (6.02 mm) and *C. liberica* × *C. dewevrei* (7.15 mm). See Supplementary Table S5 for descriptive statistics and details of ANOVA and TurkeyHSD test.

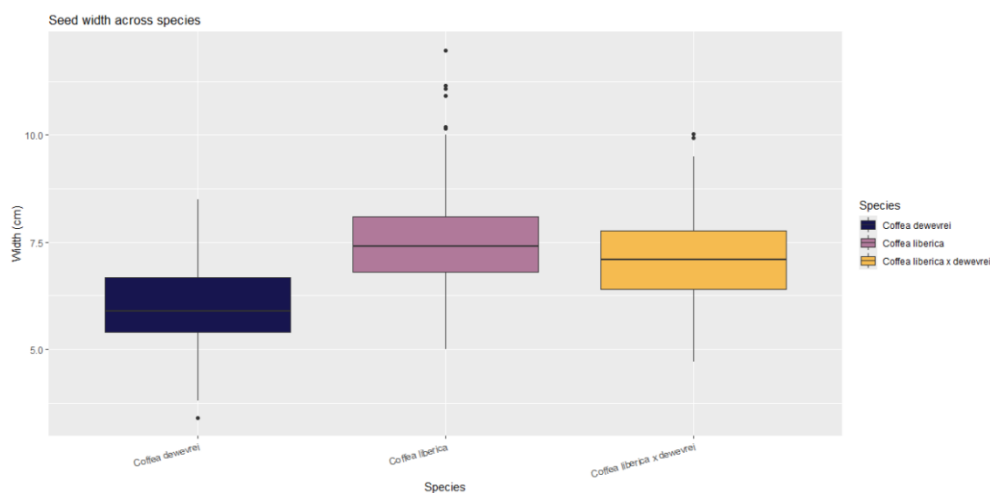
(a)



(b)



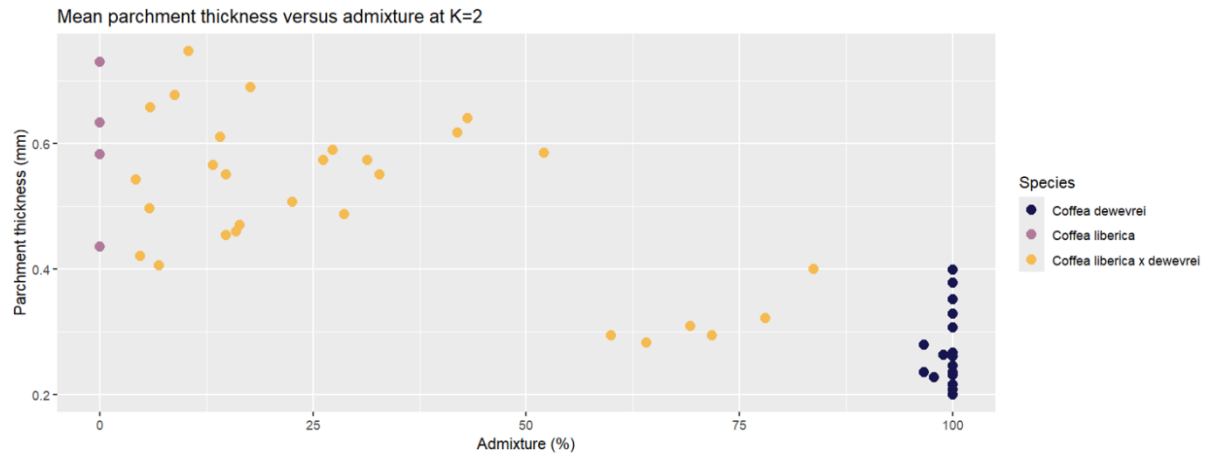
(c)



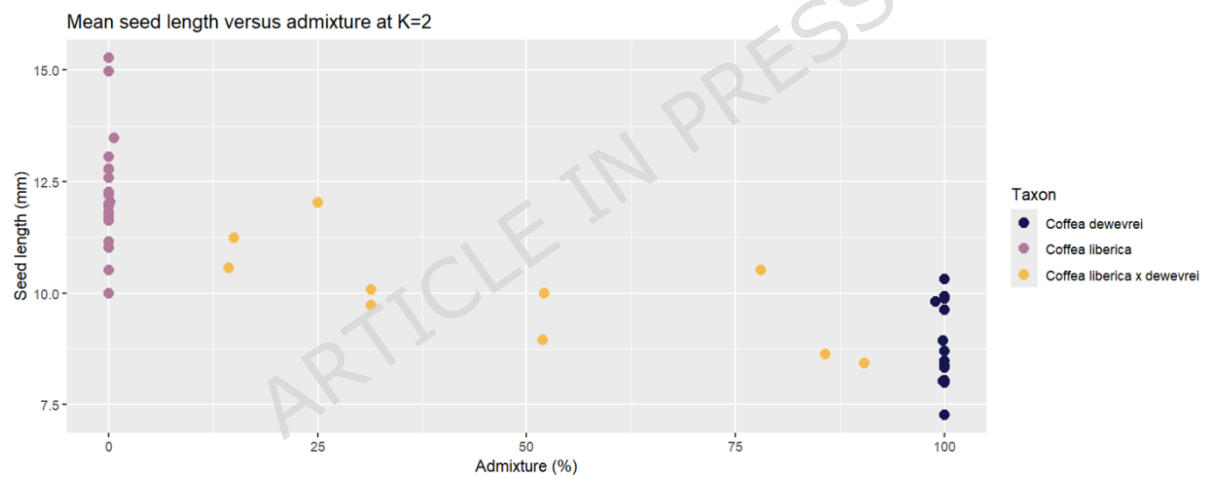
ARTICLE IN PRESS

Fig. 4. Mean parchment and seed measurements versus admixture at $K = 2$ for *C. liberica*, *C. dewevrei* and *C. liberica* $\times$ *C. dewevrei*. a, parchment thickness. b, seed length. c, seed width. See Supplementary Tables S7, S8 for accession details and mean values.

(a)



(b)



(c)

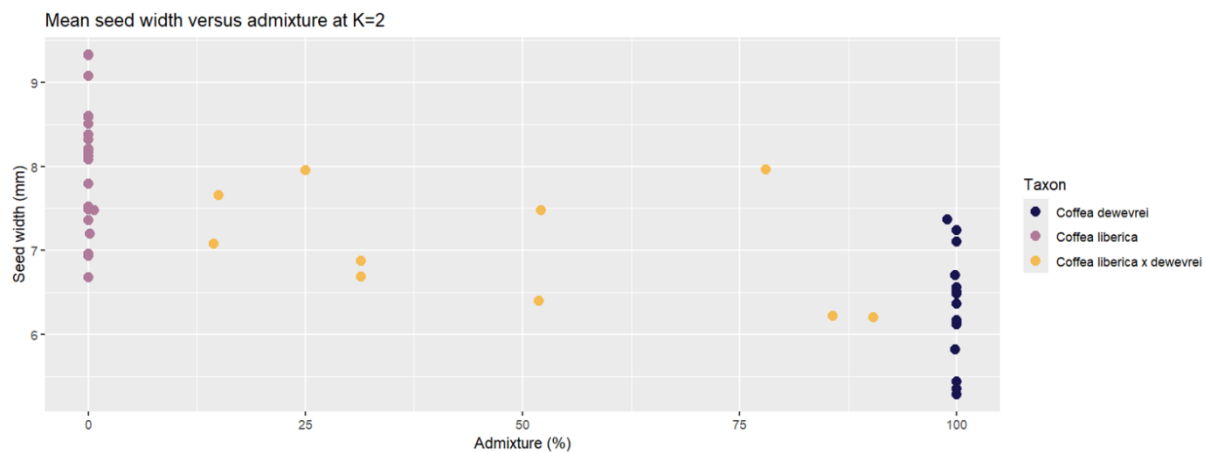
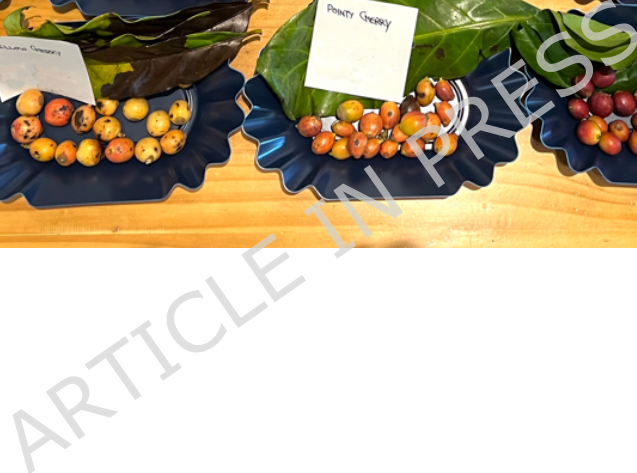
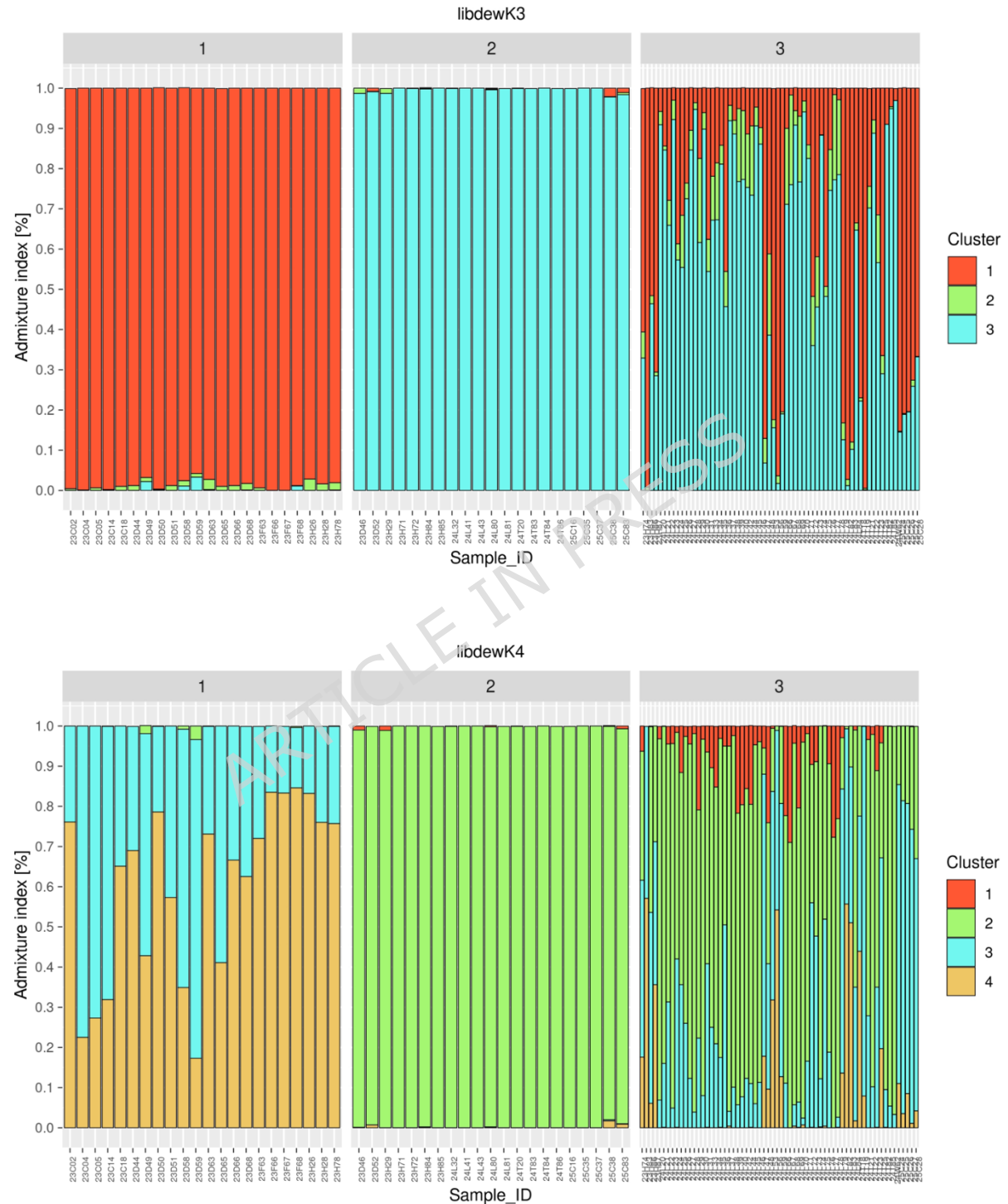


Image: Kenny Wee Ting Lee ©.



Supplementary Fig. S1. Alternative STRUCTURE K values for *C. liberica*, *C. dewevrei* and *C. liberica* × *C. dewevrei*. Analyses based on 7,618 exon region SNPs for 107 samples. Group 1 = *C. dewevrei*; group 2 = *C. liberica*; group 3 = *C. liberica* × *C. dewevrei*. K values of K = 3 and K = 4.



Supplementary Fig. S2. Seed length versus seed width for *C. liberica*, *C. dewevrei* and *C. liberica* × *C. dewevrei*. This scatter plot includes the measurement of 2,140 seeds (unroasted coffee) beans). Commercial samples are unscreened (i.e. not size-sorted by the producer). See Supplementary Table S8 for accession information.

