The shape of aroma

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

RESEARCH ARTICLE

Plants People Planet PPP

The shape of aroma: Measuring and modeling citrus oil gland distribution

Erik J. Amézquita[1] | Michelle Y. Quigley[2] | Tim Ophelders[3,4] |
Danelle Seymour[5] | Elizabeth Munch[1,6] | Daniel H. Chitwood[1,2]

[1]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan, USA

[2]Department of Horticulture, Michigan State University, East Lansing, Michigan, USA

[3]Department of Mathematics and Computer Science, TU Eindhoven, Eindhoven, The Netherlands

[4]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

[5]Department of Botany and Plant Sciences, University of California, Riverside, California, USA

[6]Department of Mathematics, Michigan State University, East Lansing, Michigan, USA

Correspondence
Daniel H. Chitwood, Department of Horticulture, Michigan State University, 1066 Bogue St, East Lansing, MI 48824, USA.
Email: chitwoo9@msu.edu

Elizabeth Munch, Department of Computational Mathematics, Science and Engineering, Michigan State University, 428 S Shaw Ln, East Lansing, MI 48824, USA.
Email: muncheli@msu.edu

Funding information
Michigan State University AgBioResearch; National Science Foundation, Grant/Award Numbers: CCF-1907591, CCF-2106578, CCF-2142713; UC Riverside; Michigan State University; USDA National Institute of Food and Agriculture

Societal Impact Statement

Citrus are intrinsically connected to human health and culture, preventing human diseases like scurvy and inspiring sacred rituals. Citrus fruits come in a stunning number of different sizes and shapes, ranging from small clementines to oversized pummelos, and fruits display a vast diversity of flavors and aromas. These qualities are key in both traditional and modern medicine and in the production of cleaning and perfume products. By quantifying and modeling overall fruit shape and oil gland distribution, we can gain further insight into citrus development and the impacts of domestication and improvement on multiple characteristics of the fruit.

Summary

- Citrus come in diverse sizes and shapes, and play a key role in world culture and economy. Citrus oil glands in particular contain essential oils which include plant secondary metabolites associated with flavor and aroma. Capturing and analyzing nuanced information behind the citrus fruit shape and its oil gland distribution provide a morphology-driven path to further our insight into phenotype–genotype interactions.

- We investigated the shape of citrus fruit of 51 accessions based on 3D X-ray computed tomography (CT) scan reconstructions. Accessions include members of the three ancestral citrus species as well as related genera, and several interspecific hybrids. We digitally separate and compare the size of fruit endocarp, mesocarp, exocarp, and oil gland tissue. Based on the centers of the oil glands, overall fruit shape is approximated with an ellipsoid. Possible oil gland distributions on this ellipsoid surface are explored using directional statistics.

- There is a strong allometry along fruit tissues; that is, we observe a strong linear relationship between the logarithmic volume of any pair of major tissues. This suggests that the relative growth of fruit tissues with respect to each other follows a power law. We also observe that on average, glands distance themselves from their nearest neighbor following a square root relationship, which suggests normal diffusion dynamics at play.

- The observed allometry and square root models point to the existence of biophysical developmental constraints that govern novel relationships between fruit dimensions from both evolutionary and breeding perspectives. Understanding these biophysical interactions prompts an exciting research path on fruit development and breeding.

1 | INTRODUCTION

"Fairest of all God's trees, the orange came and settled here,... Alone and unmoving you stand: how can one not admire you? Deep-rooted, hard to shift: truly you have no peer! Alert to this world's ways you hold your ground, unyielding against the vulgar tide."
—from "In Praise of the Orange-Tree (Ju song)" by Qu Yuan (340–278 BC?) (Hawkes, 1985).

Citrus fruits and leaves have played a fundamental role across multiple aspects of human history including the development of modern nutrition and medical sciences. The aromatic and medicinal properties of mandarins and oranges have inspired delicate poetry since ancient times (Tseng, 1999; Vovin, 2016). Etrog citrons represent "the fruit of a goodly tree" during the Sukkot celebrations in the Jewish community (Isaac, 1959). The bael tree is considered sacred, and it is generally grown near Hindu temples (Sharma et al., 2007). The fruits, peels, and leaves of diverse citrus have been used as traditional medicine for millennia for a diverse array of maladies (Mahomoodally & Mooroteea, 2021; Shrestha & Dangol, 2019). Sour oranges and lemons inspired the first modern clinical trials in the 18th and 19th centuries to determine the causes and cure of scurvy thus paving the way to the eventual isolation and synthesis of the first vitamin, vitamin C (Baron, 2009; Magiorkinis et al., 2011).

Currently, there is a rising trend in global citrus production, with more than 143 million tonnes produced in 2019 alone (FAO, 2021). The citrus production is valued for more than 3.3 billion US dollars solely in the United States (NASS, 2021). Citrus-derived products are vital for other multibillion dollar industries as well, from orange juice in the food industry to essential oils in the perfume and cosmetics industry (Spreen et al., 2020). Essential oils in particular are extracted for their aromatic, flavoring, medicinal, and preservation properties useful in a variety of contexts (Mahato et al., 2019).

Before any human intervention, current paleobotanical evidence suggests that the common ancestor of citrus species originated more than 8 million years ago in the triangle defined by modern day northeastern India, northern Myanmar, and northwestern Yunnan (Talon et al., 2020). Throughout the next 2 million years, as monsoons weakened in southeastern Asia and climate transitioned to drier conditions, citrus underwent a rapid radiation and diversification. Citrons, pummelos, mandarins, kumquats, and papedas split from each other and

expanded across the south and southeast Asian peninsula, the Tibetan Plateau, and even Japan (Wu et al., 2018). Citrus underwent a second rapid radiation period 4 million years ago as they crossed the Wallace line and dispersed throughout Australia, giving rise to modern day microcitrus (Wu et al., 2018). Early civilizations in India and China domesticated some of these species and their interspecific hybrids, even as early as during the Xia Dynasty (2100–1600 BC) in Southern China (Deng et al., 2020). Through tribute, trade, and invasion, different cultures contributed to spread many of these citrus across the rest of the world over the next 3000 years (Langgut, 2017).

Citrus species are sexually compatible, and their ability to hybridize, combined with constant displacement and cultivation in multiple environments, produced a diversity of admixed accessions with a vast array of phenotypic traits (Gmitter et al., 2020; Luro et al., 2017; Wu et al., 2021). Asexual propagation is common in citrus, and the interaction between grafted individuals has led to novel phenotypes, including through the formation of graft chimeras, conglomerations of cells that originated from separate zygotes (Caruso et al., 2020). The first reported plant chimera, known as Bizarria, arose from a fortuitous graft junction of a Florentine citron and a sour orange in 1674 (Nati, 1674). Since then, chimeras have proved to be more common than originally thought, transforming our perception of the genetic heterogeneity of individuals and its impact on plant development and phenotype (Frank & Chitwood, 2016).

A phenotype of particular interest is shape. Specific combinations of shape features are used to distinguish diverse citrus varieties and have motivated various citrus taxonomic systems (Ollitrault et al., 2020). Leaf shape has been used to distinguish pummelos from sweet oranges among other different citrus genotypes and their respective environment interactions (Iwata et al., 2002). Root architecture is indicative of soil deficiencies for sour orange rootstocks (Mei et al., 2011). Morphological traits, such as fruit size and oil gland density, are used to infer genetic similarities between various mandarin cultivars (Pal et al., 2013). Oil gland size, structure, and distribution are associated with the fruit development of navel oranges (Knight et al., 2001) and grapefruits (Voo et al., 2012).

Here we study the shape of citrus fruits and fruits from close citrus relatives based on 3D X-ray computed tomography (CT) scan reconstruction of 166 different samples comprising 51 different accessions, including samples of the three fundamental citrus species (*Citrus medica*, *Citrus reticulata*, and *Citrus maxima*), accessions from related genera (*Poncirus*, *Fortunella*, and *Microcitrus*), and several

interspecific hybrids. Modern citrus types, including sweet oranges, mandarins, lemons, limes, and grapefruits, result from a series of interspecific hybridizations between the primary fundamental citrus species. For example, sweet oranges, mandarins, and grapefruits contain ancestry from both *C. reticulata* and *C. maxima* and lemons and limes are interspecific hybrids between *C. medica* and either sour orange (lemons) or *C. hystrix* (limes) (Wu et al., 2018). Interspecific hybridization is also commonly used to develop rootstock cultivars, with *P. trifoliata* being an important source of both biotic and abiotic stress resistance. The 51 accessions were selected to represent modern cultivated citrus types as well as the species contributing ancestry to each group. First, using the power of X-rays and image processing, we compared volume ratios between different tissues, including exocarp, endocarp, and oil gland tissue. Second, since citrus oil glands contain essential oils which include plant secondary metabolites associated with flavor and aroma, we examined the number of individual oil glands, their density, and their overall distribution across all fruits. We determine that the average distance between neighboring oil glands follows a square root model, which indicates that gland distribution might follow normal diffusion dynamics (Vlahos et al., 2008). Finally, based off a point cloud defined by the center of all individual oil glands, we model the fruit shape as an ellipsoidal surface, a sphere with its three main axes shrunk or stretched accordingly. Once the glands are considered points on an ellipsoid, we are able to apply multiple tools from directional statistics (Ley & Verdebout, 2017; Mardia & Jupp, 1999; Pewsey & García-Portugués, 2021), which allows us to study and infer possible statistical distributions on spherical surfaces. As an example of this mathematical machinery, we test whether the oil glands either follow a uniform or symmetric distribution across the fruit surface. To the best of our knowledge, the shape of citrus has not been explored with similar scanning technologies nor has it been analyzed with ellipsoidal and directional approximations. This morphological modeling will allow us to set a new exciting path to explore further the phenotype–genotype relationship in citrus.

2 | MATERIALS AND METHODS

2.1 | Plant material and scanning

We selected 51 different accessions of citrus and citrus relatives with diverse fruit morphologies and geographical origins for our analysis. Fruits were sampled from a single tree for each selected accession maintained in the University of California Riverside Givaudan Citrus Variety Collection. A total of 166 different individuals in total were sent for scanning at Michigan State University in December 2018 (Figure 1a and Table S1). These 166 samples were arranged into 63 raw scans, one scan per citrus variety containing all the replicates. Pummelo and citron samples were scanned individually due to the fruit size. The scans were produced using the North Star Imaging X3000 system and the included efX software, with 720 projections per scan, at three frames per second and with three frames averaged per projection. The data were obtained in continuous mode. The fruits were

placed as close as possible to the X-ray detector, provided all fruits could be scanned completely at once. The X-ray source was set to a current of 70 μA, voltage ranging from 70 to 90 kV, and focal spot sizes ranging from 4.9 to 6.3 μm. The 3D reconstruction of the fruit was computed with the efX-CT software, obtaining final voxel sizes ranging from 18.6 to 110.1 μm for different scans (Figure 1b and Table S2). Difference in voxel size corresponds to difference in fruit size.

The air and debris were thresholded out of each raw scan and individual replicates segmented into separate images. Based on density and location, for each fruit, we further segmented 3D voxel-based reconstructions of its central column, endocarp, mesocarp, exocarp, and oil glands (Figure 1c–g). The center of each oil gland was calculated as the center of mass of the voxels composing such gland. An in-house scipy.ndimage-based python script was used to process the images for all fruits and their tissues. These were later visually inspected to verify their correctness. All the Chinese box oranges (*Severinia buxifolia*) scans were discarded due to their poor quality. To highlight nuanced differences among certain citrus groups, scans were partly split into sensible clusters of morphological interest (Table 1).

2.2 | Allometric relationships

The total volume of fruits and their separate tissues was measured from the scans, as well as the number of individual oil glands. We studied the allometric relationships between these measurements; that is, the relative size of different tissues with respect to each other. These relationships in plants often follow a power law, so all the measurements were first log-transformed (Niklas, 2004) and a reduced major axis linear regression (Smith, 2009) was fitted considering all fruits. The slope, intercept, and R^2 coefficient of determination were recorded (Figures 2 and S1). The distribution of the residuals was compared against a normal distribution to determine the adequacy of the linear fit (Figures S2 and S3).

2.3 | Oil gland distribution

For each fruit, a point cloud, a collection of (x, y, z) coordinates was defined by the centers of all its individual oil glands. The 25 nearest neighbors, based on Euclidean distance, were computed for each point, so that distances are not affected by the fruit skin curvature. The average distance between each gland and its nearest neighbor, its second nearest neighbor, and so on were computed. The oil gland density was determined both in terms of volume and surface area, by dividing the number of glands by the volume of the whole fruit and by the surface area of the best-fit ellipsoid (discussed later in Section 2.4), respectively. As in the previous section, all these measurements were log-transformed, linear regressions fitted, parameters recorded, and residuals compared with a normal distribution (Figures 3a and S4). A root square model was fitted between the average nearest neighbor distance and the nearest neighbor index to describe how far apart glands spread from each other (Figure 3b).
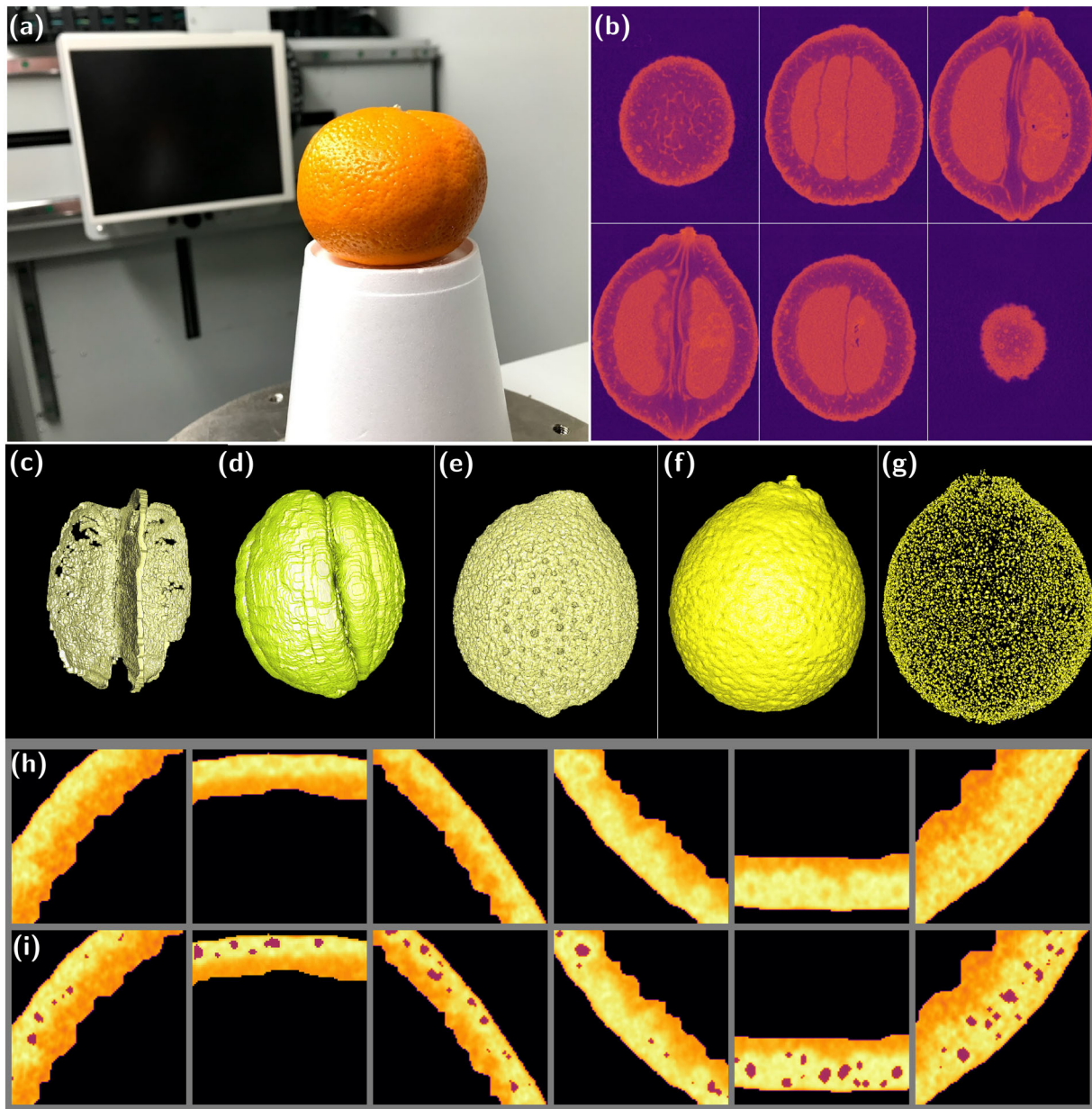
FIGURE 1 Depiction of our citrus scanning, data acquisition and image processing pipeline. A Willowleaf sour orange is used as an example for figures (b)–(i). Images not to scale. For illustration purposes only. (a) A diverse collection was scanned using X-ray computed tomography (CT) technology. (b) Slices of a raw scan. The image processing steps involved segmenting individual fruits and removing air and other debris. Then, individual tissues for each fruit were segmented such as the (c) central spine, (d) endocarp, (e) mesocarp, (f) exocarp and (g) oil glands. (h) Close-up of some X-ray slices of the exocarp. (i) Same figure as above, with the segmented oil gland tissue darkened for emphasis

2.4 | Modeling the whole fruit as an ellipsoid and computing its sphericity

The surface of most of citrus fruits and their relatives can be approximated by an ellipsoid, a sphere with its three main axes possibly shrunk or stretched. The three axes of symmetry of an ellipsoid delimit three line segments from the center of the ellipsoid to its surface. These are referred to as the *ellipsoid semi-axes*. Notice that a sphere is an ellipsoid with its three semi-axes of the same length. We will consider triaxial ellipsoids, where the length of each semi-axes can be different. An ellipsoid can also be represented as a quadratic equation surface which is both mathematically simple to manipulate (Harris & Stöcker, 1998 Ch. 8.12) and versatile enough to represent both the shapes of nearly spherical Valencia oranges and elongated finger limes given the right semi-axes lengths.

Each fruit is defined by a point cloud made by the centers of all its individual oil glands. The parameters of the best-fit ellipsoid for this point cloud are computed following the algorithm by (Li &

TABLE 1 A total of 51 citrus accessions were scanned to represent modern cultivated citrus types as well as the species contributing ancestry to each group

CVC name	Scientific name	S	-	CVC name	Scientific name	S
Kumquats						
Nagami kumquat	*F. margarita* (Lour.)	3				
Lemons and lemon hybrids						
Limon real	*C. excelsa* Wester	4		Lamas lemon	*C. limon* L. Burm.f.	3
Interdonato lemon	*C. limon* L. Burm.f.	3		Volckamer lemon	*C. volkameriana*	3
Frost nucellar Eureka lemon	*C. limon* L. Burm.f.	3				
Mandarins and mandarin hybrids						
Scarlet Emperor mandarin	*Citrus reticulata* Blanco	3		*	*Citrus reticulata*	3
Lee mandarin	*Citrus reticulata* Blanco RUTACEAE	3		Som Keowan	*Citrus reticulata* Blanco	3
Koster mandarin	*Citrus reticulata* Blanco	3		Cleopatra mandarin	*C. reshni* hort.ex Tanaka RUTACEAE	3
Beledy mandarin	*Citrus reticulata* Blanco	4		Fremont mandarin	*Citrus reticulata* Blanco RUTACEAE	3
USDA 88–2 (Lee X Nova)	*Citrus reticulata* Blanco	3		Kinkoji Unshiu	*C. neo-aurantium*	3
Microcitrus						
Australian finger lime	*M. australasica*	4				
Papedas						
*	*C. hanayu* Siebold ex Shirai	3		Kalpi, Nogapog	*C. webberii*	2
Makrut lime	*C. hystrix* DC	4				
Pummelos and pummelo hybrids						
Tahitian X Star Ruby	*C. paradisi* Macfadyen	3		Pomelit hybrid	*C. maxima* (Burm.) Merr	3
Kao Pan pummelo	*C. maxima* (Burm.) Merr. RUTACEAE	3		Hassaku (Beni) hybrid	*C. hassaku* hort.ex Tanaka RUTACEAE	3
Egami Buntan pummelo	*C. maxima* (Burm.) Merr. RUTACEAE	3				
Sour oranges and sour orange hybrids						
Willowleaf sour orange	*C. aurantium* var. salicifolia	3		Standard sour orange	*C. aurantium* L.	3
Konejime hybrid	*C. neo-aurantium*	4		Olivelands Sour orange	*C. aurantium* L.	3
Sweet oranges						
Cutter Valencia nucellar	*C. sinensis* L. Osbeck	3		Shamouti orange	*C. sinensis* L. Osbeck	3
Parent Washington navel	*C. sinensis* (L.) Osbeck	3		Argentina sweet orange	*C. sinensis* (L.) Osbeck RUTACEAE	3
Cara Cara pink navel	*C. sinensis* {L.} Osbeck	3				
Trifoliates and trifoliate hybrids						
Little leaf trifoliate	*P. trifoliata* (L.) Raf. RUTACEAE	3		Rubidoux trifoliate	*P. trifoliata* (L.) Raf. RUTACEAE	4
C-35 citrange	*X Citroncirus* spp.	3		Carrizo citrange	*X Citroncirus* spp. *RUTACEAE*	5
Swingle citrumelo	*X Citroncirus* spp. *RUTACEAE*	5				

Note: The following groups and varieties were highlighted for further shape and size analysis. **S** equals the number of pseudo-replicates scanned. The full list of citrus fruits and relatives scanned is found in Table S1. Names according to the University of California Givaudan Citrus Variety Collection (CVC). Asterisk denotes not available. Note that there are two columns in this table.

Griffiths, 2004), from which the semi-axes lengths, rotations, and center are determined (Panou et al., 2020). The fruit point cloud is then rotated and translated such that the best-fit ellipsoid is centered at the origin and its semimajor axes coincide with the proximal-distal axis of the fruit. Finally, the centers of the oil glands are projected to this ellipsoid via geocentric projection, where a ray from the center of the ellipsoid to the gland is drawn and its intersection with the ellipsoid is considered (Figure 4a–f).

This ellipsoid model summarizes important information of the overall shape of the fruit. As an example, we measure how sphere-like different citrus are. There is no unique way to measure sphericity; however, most of the commonly used formulas are based on the semi-axes lengths of the object (Blott & Pye, 2008; Clayton et al., 2009). We measured the sphericity of the resulting fruit-based ellipsoids using six different sphericity indices, all of them taking values between 0 (planes and lines) and 1 (perfect spheres) (Figure 4g and Table 2).
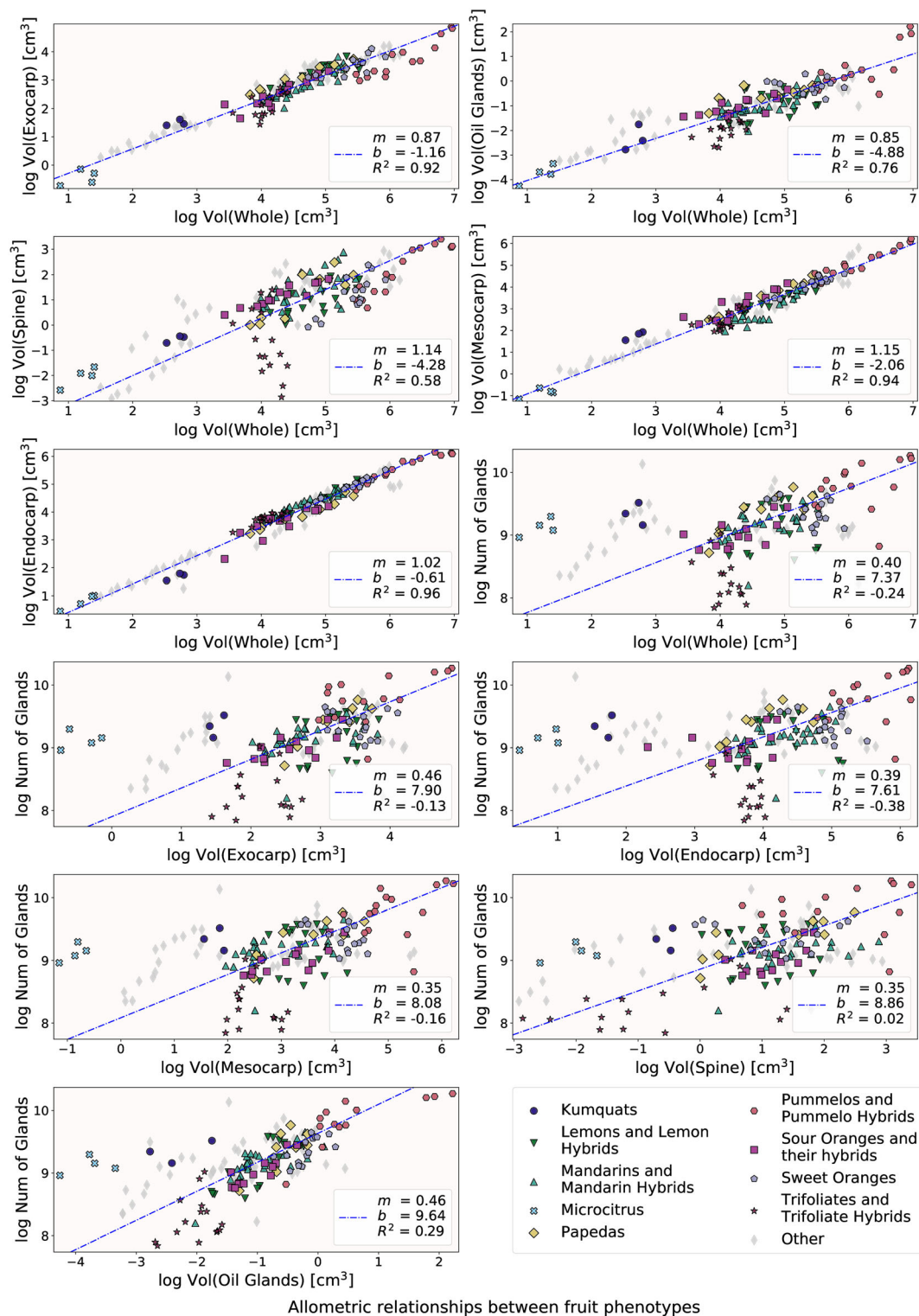
FIGURE 2 Various allometry plots to study the strong relationship between different tissue volumes compared with the total fruit volume across all citrus fruits. These relationships suggest that tissues grow relative to the whole following a power rule. The best-fit line is depicted by a dashed line in blue. However, this relationship is less clear when comparing the number of individual oil glands to different fruit tissue sizes. Several citrus groups are highlighted as in Table 1. The rest of citrus fruits are depicted by gray diamonds. "Vol" refers to the volume of the indicated tissue. For each plot, the slope, intercept and correlation coefficient are recorded as m, b and R^2, respectively. The linear relationship in the log–log plots suggests that fruit tissues may grow following a power law.
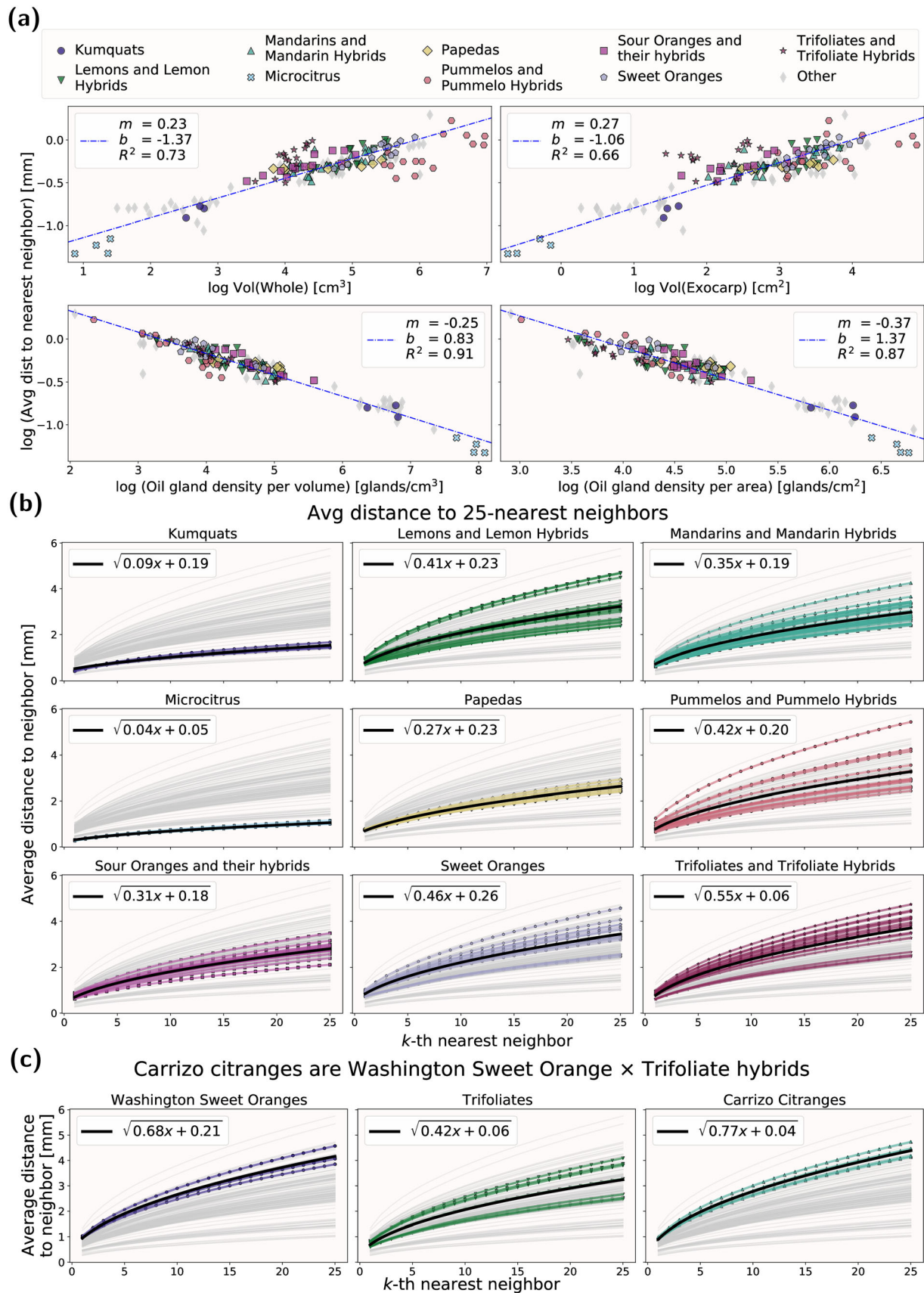
FIGURE 3 Legend on next page.

FIGURE 3   For every oil gland, the distance to the other 25 nearest glands was computed. Then the average distance between each oil gland and its nearest, second nearest, and so on neighboring gland was considered. (a) Allometric relationships are observed across all fruits when comparing the average distance between each gland to its closest neighbor with the overall size of the fruit. The overall linear trend is depicted by a dashed blue line. Several citrus groups are highlighted as in Table 1. The rest of citrus fruits are depicted by gray diamonds. "Vol" refers to the volume of the indicated tissue. The slope, intercept, and correlation coefficient are denoted by m, b and R^2, respectively. (b) For each oil gland, the average distance to its nearest neighbors follows a square root relationship. The average for each group is plotted as black, thick line. Trends for individual indicated fruits plotted as colored lines, while the rest of the individuals are plotted in gray. These square root models follow different parameters depending on the citrus group. The fruits that deviate the most from the average usually correspond to hybrids. (c) Carrizo citranges are Washington sweet orange × trifoliate hybrids. The average distance between oil glands increases at a faster rate for citranges than for their parents, which suggests hybrid vigor.

2.5 | Revisiting the distribution of the oil glands

The projected gland center locations on the ellipsoid were described in terms of longitude and latitude coordinates with respect to the ellipsoid (Diaz-Toca et al., 2020). We tested whether the gland point cloud follows a uniform distribution, where every unit area of the skin has the same probability of containing oil glands or if the underlying distribution is rotationally symmetric, where the oil glands pattern is symmetrical around a fixed direction. Uniformity was tested with Projected Anderson–Darling (PAD) test (García-Portugués, Navarro-Esteban, & Cuesta-Albertos, 2020) with the R package sphunif (García-Portugués & Verdebout, 2021). The rotational symmetry was tested with a scatter-location hybrid test with an unspecified direction of symmetry (García-Portugués, Paindaveine, & Verdebout, 2020) with the R package rotasym (García-Portugués et al., 2021). Additionally, we visually examined the distribution of oil glands for most fruits and compared them to simulated uniform and rotationally symmetric distributions, like von Mises–Fisher and Bingham distributions. These were simulated with the R package Directional (Tsagris et al., 2022). These oil glands and simulated distributions were projected to 2D via Lambert azimuthal equal-area projections (Mardia & Jupp, 1999 Ch. 9.1) from the North and South poles. Intuitively, these two projections flatten the sphere on a plane by pushing it from the North or South poles while minimizing the distortion seen on the north and south hemispheres, respectively (Figure 5).

3 | RESULTS

3.1 | Allometric relationships

The estimated volume of each tissue type and fruit follows the expected average fruit size of each genetic group, with the smallest fruit in the bottom left corners (microcitrus, kumquats) and large fruit in the top right corners (pummelos). Strong linear trends are observed when comparing most of the volume-related features of all the fruits, indicated by high R^2 coefficients of determination, usually above 0.75. There is a strong positive correlation between the total volume of oil glands and the volume of the whole fruit. The main exception is the central column tissue (Figures 2 and S1). Due to their thin size and scanning quality, these columns were difficult to identify and isolate,

especially in some trifoliates R^2 values. The residuals of the fitted linear regression tend to follow a normal distribution for most measurement pairs, suggesting that the linear fit is adequate (Figures S2 and S3). This linearity indicates that the tissues across all citrus fruits grow relative to each other following a power rule. For example, looking at the slope m values, both the exocarp and oil glands grow in volume at the same relative rate with respect to the whole fruit ($m = 0.85$). On the other hand, the total number of oil glands appears to be decoupled from all the measured size-related traits, as shown by low and negative R^2 values (Figure 2). In this case, a power law may not be an adequate model to describe the oil gland number with respect to tissue volume.

3.2 | Oil gland distribution

There is a strong positive linear relationship between the volume of the fruit and the average distance between an oil gland and its nearest neighbor, with R^2 coefficients of determination above 0.65. There is a stronger negative linear relationship when considering the overall oil gland density, reflected by R^2 coefficients above 0.85 (Figure 3a). The residuals follow normal distributions, indicating that the linear model is adequate (Figure S4). These allometric relationships suggest that for all citrus and relative fruits, the average distance between nearest oil glands follows a power law with respect to fruit volume and gland density. When considering fruit size, as expected, the samples distribute in a similar pattern as with most of the previous allometry plots. However, an inverse pattern is observed when considering oil gland density. In this case, the smallest fruits tend to report the highest number of oil glands per unit volume or unit area. Other than microcitrus and kumquats, the rest of highlighted citrus groups form a tighter cluster. This higher gland density could be partly affected by differences in scanning resolutions (Figure S5).

The average distance between an oil gland and its k-th nearest neighbor is modeled as

$$\text{Average distance}(k) = \sqrt{Mk + B},$$

where M is the rate of distance growth and B the line intercept (Figure 3b). As expected from their higher oil gland density, the average distance to the gland's nearest neighbors increases the
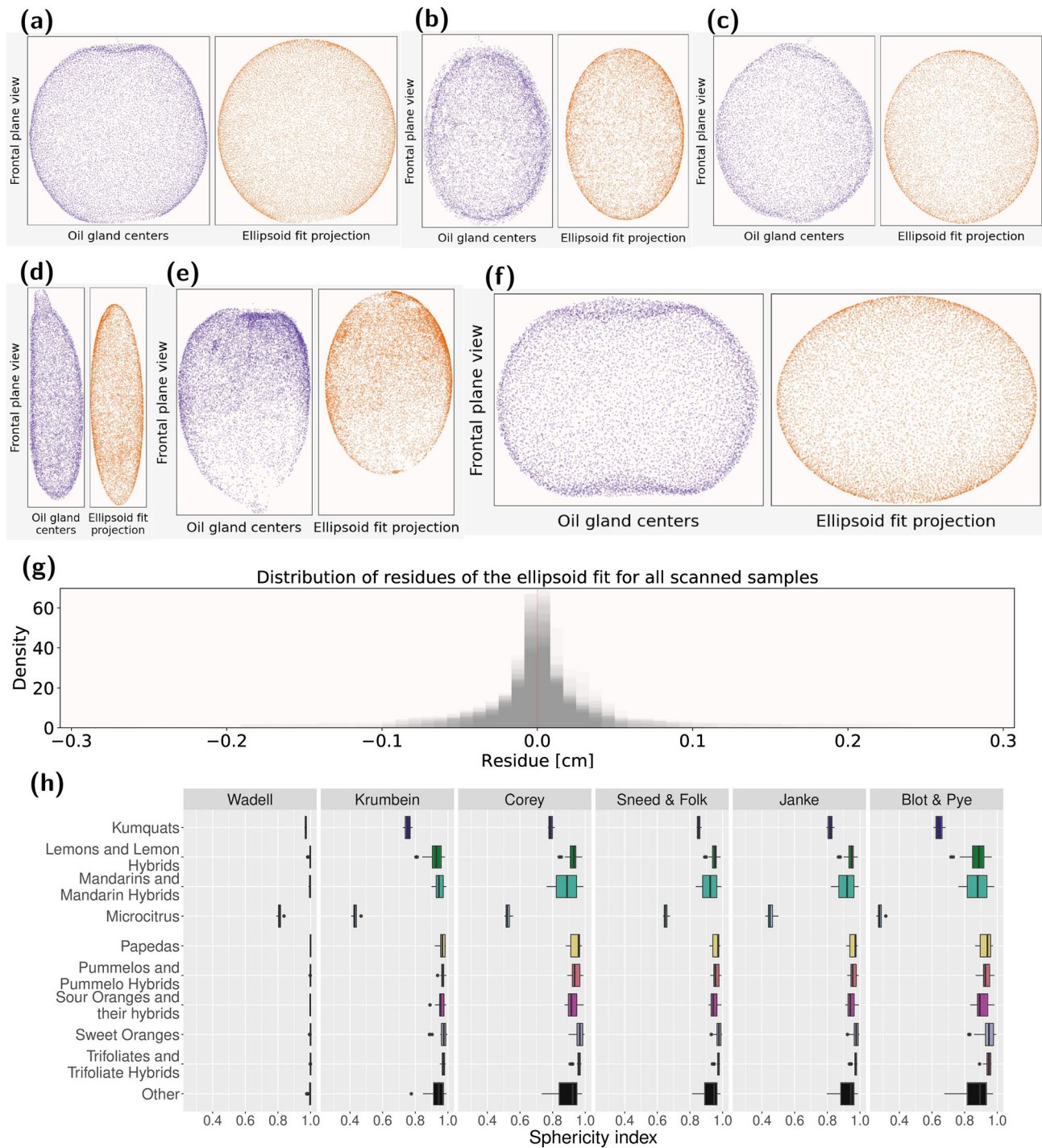
FIGURE 4 Modeling citrus as ellipsoids. The surface of different citrus was approximated with tri-axial ellipsoids. The glands were later centered at the origin, and the ellipsoid aligned with the proximal–distal, medial–lateral and adaxial–abaxial axes. Next, the oil glands were projected to this best-fit ellipsoid simply by projecting a ray from the origin—also known as the geocentric projection. Examples of a (a) Cutter nucellar Valencia orange, (b) Nagami kumquat, (c) Willowleaf sour orange, (d) Australian finger lime, (e) South Coast Field Station citron and a (f) Cleopatra mandarin. Figures (a)–(f) are not scaled. (g) Distribution of the residuals of the centers of the oil glands to the best-fit ellipsoid. The distributions for all the fruit scans are overlaid. (h) Various sphericity indices are computed and compared across different citrus groups. The indices are named according to their original reference in Table 2.

slowest for microcitrus, followed distinctly by kumquats. On the other hand, sweet oranges and trifoliates report the largest average distances between neighboring oil glands.

In general, all the samples of every accession follow the same growth model. Outliers in growth models are typically associated with hybrid accessions. For example, consider the Carrizo citrange, a

TABLE 2 Common sphericity indices were considered

Name	Formula	Reference
True sphericity	$\frac{A_s}{A_e} = \frac{\sqrt[3]{36\pi V_e^2}}{A_e}$	Wadell (1932)
Intercept sphericity	$\sqrt[3]{\frac{bc}{a^2}}$	Krumbein (1941)
Corey shape factor	$\frac{c}{\sqrt{ab}}$	Corey (1949)
Maximum projection sphericity	$\sqrt[3]{\frac{c^2}{ab}}$	Sneed and Folk (1958)
Janke form factor	$\frac{c}{\sqrt{\frac{1}{3}(a^2+b^2+c^2)}}$	Janke (1966)
Degree of equancy	$\frac{c}{a}$	Blott and Pye (2008)

Note: These indices are based off best-fit ellipsoid semi-axes lengths. The indices values are bounded between 0 (line or plane) and 1 (perfect sphere). The surface area, volume, and the largest, intermediate, and smallest semi-axes lengths of the ellipsoid are denoted by A_s, respectively. Also, V_e denotes the surface area of a sphere of volume V_e.

trifoliate × Washington sweet orange hybrid where sweet orange fruits are much larger than trifoliate fruits. We observed that the oil glands in the citrange grow on average farther apart from each other than in any of the parents (Figure 3c), which suggests that hybrid vigor might be at play.

3.3 | Ellipsoid modeling and sphericity of fruits

The best-fit ellipsoid successfully captures the overall shape of the citrus and relatives, with a negligible portion of gland centers differing by more than 2 mm from their ellipsoidal approximation (Figure 4g). This ellipsoidal model is flexible enough to capture both spherical and elongated fruit shapes, from sweet oranges to Australian finger limes (Figures 4a–f).

Most of the fruits report highly spherical indices, more than 0.9 for every sphericity index. Unsurprisingly, the elongated Australian finger limes are the least spherical and their shape is very distinct from the rest of the samples. The kumquats are less elongated than the finger limes, but they also remain highly distinguishable for most of the sphericity indices. Mandarins and their hybrids also tend to be slightly less spherical than the remaining groups of interest (Figure 4h).

3.4 | Oil glands revisited

Although the tools from directional statistics assume that the data points lie on a sphere rather than an ellipsoid, most of scanned fruits were very sphere-shaped according to a variety of sphericity indices (Figure 4h). Thus, shape information is not significantly altered when translating longitude and latitude coordinates from the best-fit ellipsoid to a sphere.

The uniform oil gland distribution hypothesis was strongly rejected for all scans, with all *p*-values below 0.015 for the PAD test and below 2.5×10^{-7} for 95% of all the point clouds (Figure S6a). The 10 samples for which the rotationally symmetric

hypothesis was not rejected were not concentrated in any citrus groups (Table S3 and Figure S6b). Upon closer visual examination, differences arise between the uniform distribution on a sphere and oil gland distributions. The oil gland distributions tend to have defined clusters and empty spots, which are not seen in typical uniform distributions. The northern and southern hemispheres are noticeably different from each other for the oil gland distributions, while these look roughly the same in uniform distributions (Figure 5).

4 | DISCUSSION

Measuring and understanding the shape is fundamental to extracting valuable information from data, and push further our insights. A vast number of biological-inspired shapes are intrinsically three dimensional, like citrus fruit, and capturing their shape as 3D voxel-based provides a faithful shape representation that allows accurate measurement of tissue volumes and modeling of gland distributions in space. Better fruit modeling is key to provide more accurate descriptions of fruit shape and oil gland content, as both are important traits for citrus scion improvement (Barry et al., 2020). Citrus shape impacts oil gland abundance and distribution, as the shape of the rind, the exocarp, and other tissues, along the distribution of the oil glands, affect the physics of citrus essential oil extraction and aroma dispersion (Smith et al., 2018).

When observing overall fruit tissue size trends, these correspond to known citrus genealogy. For example, when comparing the size of exocarp against size of endocarp, most of the sour and sweet oranges tend to lie between mandarins and pummelos, with sour oranges lying closer to mandarins, while the sweet oranges are closer to pummelos (Figure 2). A similar arrangement of citrus groups is observed when comparing the average distance between neighboring oil glands to either fruit volume or gland density. Moreover, it is observed that oil glands distance themselves from each other following a square root rate in general. The exact degree to which they push each other apart depends on the oil gland density, which in turn is partly affected by the citrus genealogy. For example, the average distances between oil glands in Carrizo citranges increase at higher rates than in either the Washington sweet oranges or the trifoliates, the citrange parents (Figure 4). We also observe that these distances in microcitrus and kumquats follow similar patterns, which might correspond to the fact that kumquats are the closest relative to Australian citrus (Wu et al., 2018). The square root suggests that the mechanics of oil gland displacement across the fruit could be partly governed by based on Brownian motion and normal diffusion interactions (Vlahos et al., 2008). However, the hypothesis of oil gland locations following either a uniform or symmetric distribution on the fruit surface is strongly rejected for all scans by the PAD and location-scatter hybrid tests, respectively. This discrepancy between our gland distribution observations could be explained by the fact that uniform distributions and diffusion processes assume that the data consists of point particles with no volume that can stand arbitrarily close to each other. For oil glands, this is obviously not the case, as they have volume and there are physical limitations on the proximity between glands, which
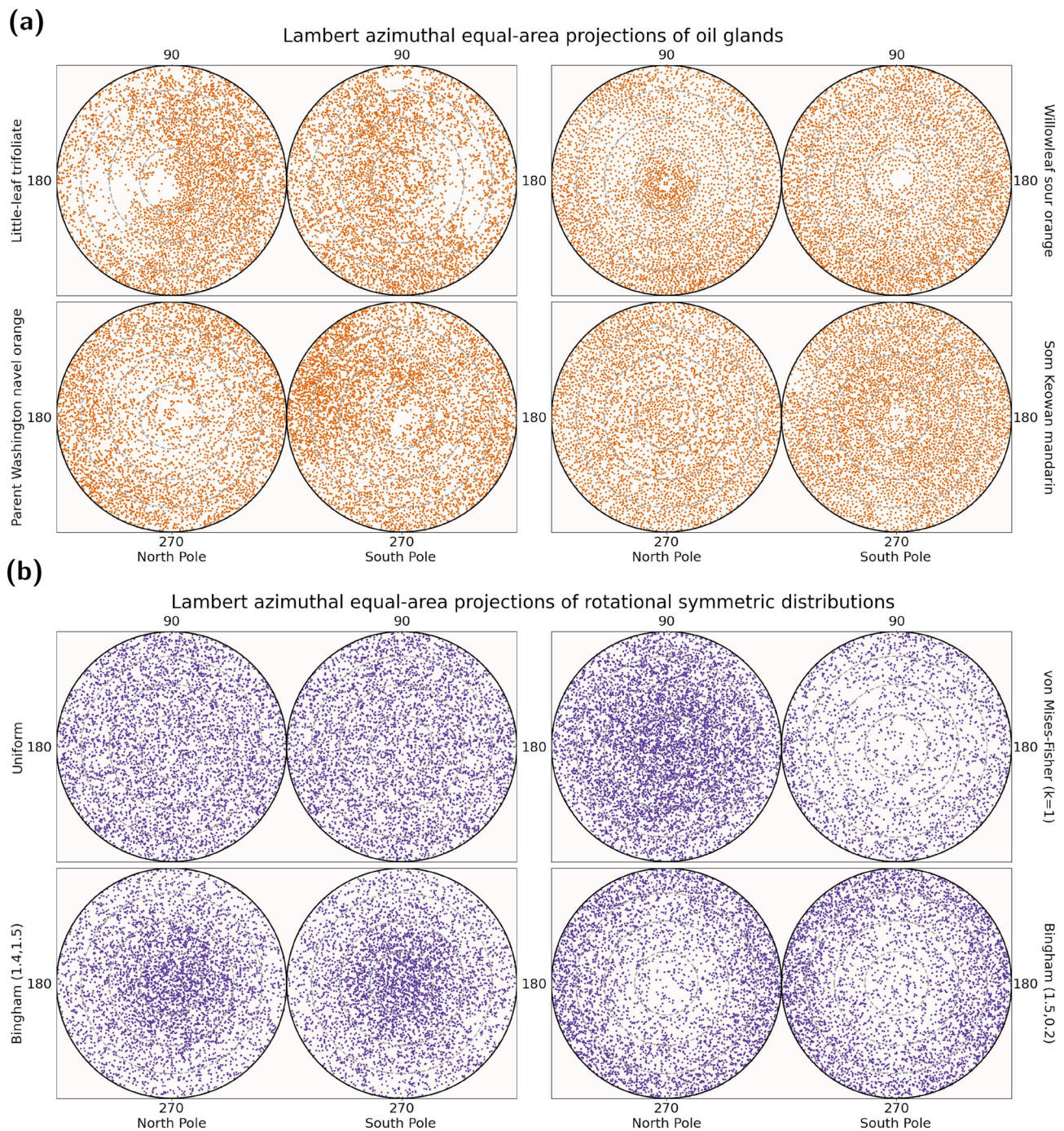
(a)

(b)

FIGURE 5 Distribution of oil glands is not uniform nor rotationally symmetric across the citrus exocarp. After modeling the fruits as ellipsoids and projecting their oil glands onto the ellipsoidal surface, longitude and latitude coordinates are computed as in Figure 4. These coordinates can be better visualized using two Lambert azimuthal equal-area projections, from the North and South poles which represent the northern and southern hemispheres, respectively, with minimal distortion. A battery of statistical tests strongly rejects the hypothesis of these glands being uniformly or symmetrically distributed over the ellipsoid surface. (a) Examples of oil gland distribution of a little leaf trifoliate, a parent Washington navel orange, a Willowleaf sour orange and a Som Keowan mandarin. (b) For comparison, a similar number of points are simulated following uniform, low-concentration von Mises–Fisher and low-concentration Bingham distributions. These three distributions are rotationally symmetric.

requires a more complex diffusion modeling. Higher-resolution scans might be able to capture better individual oil gland shape, rather than just its center. Individual oil glands then could be approximated by individual minimum volume enclosing ellipsoids (Todd &

Yıldırım, 2007), which could then pose more advanced distribution and diffusion models. At the same time, more advanced tools from directional statistics, like spherical kernel density estimators (Di Marzio et al., 2019; Vuollo & Holmström, 2018), might be able to

characterize the oil gland distributions in a nonparametric way, while allowing a more numerical comparison of their differences among distinct citrus groups.

All the studied citrus and related accessions exhibit allometric behavior in general across both their tissue volumes, and average distances between neighboring glands. This relative growth relationships suggest that tissue sizes are deeply linked, as the size of oil gland tissue in general may not be able to change without changing volume of both the endocarp and mesocarp. Moreover, there might be biophysical principles at play that govern different tissue development across all citrus fruits in general, just like normal diffusion might govern oil gland distribution. The determination of such biophysical constraints prompts future lines of exciting research.

There is rich shape information in the natural world to be captured, analyzed, and linked to biophysical developmental and evolutionary principles. Sound mathematical models are key to uncover these biophysical interactions at work. For example, even with a limited number of points, overall fruit shapes can be approximated with various quadratic surfaces like ellipsoids. Given the appropriate parameters, an ellipsoid can represent both nearly spherical navel oranges, and elongated finger limes. This quadratic surface approximation is mathematically versatile and computationally simple and can be applied to other round-shaped biologically motivated data. Moreover, ellipsoid coordinates can be translated naturally to longitudes and latitudes on a sphere, which opens the door to a wide array of mathematical tools from directional statistics. Some of those tools, like density estimations, hypothesis testing, and distribution fitting, allow us to quantify shape in a mathematically rigorous and comprehensive way. The quality of input data for our models is equally important. Through X-ray CT scanning technology, we have a novel way to observe, quantify, and analyze all the shape of citrus and their tissues in a comprehensive, automated, noninvasive, and nondestructive manner. With the right voltage and current, the 3D X-ray CT reconstructions can discern small, individual tissues, like oil glands, which enables us to analyze tissue shape and distribution at very granular levels. Capturing and analyzing this nuanced shape information for a wide array of data sets provide a morphology-driven path to further our insight into phenotype–genotype relationships. As stated by D'Arcy Thompson in his seminal biomathematical treatise *On Growth and Form*, "An organism is so complex a thing, and growth so complex a phenomenon, that for growth to be so uniform and constant in all the parts as to keep the whole shape unchanged would indeed be an unlikely and an unusual circumstance. Rates vary, proportions change, and the whole configuration alters accordingly" (1942).

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

EA, DS, EM, and DC conceived the experiment. DS selected the accessions to be scanned and collected the plant material, ensuring that citrus types were broadly represented. MQ and DC collected the digital data. EA and TO developed the necessary scripts to process the scans and extracted their shape descriptors. EA analyzed the data and wrote the manuscript. All authors contributed, reviewed, and revised the manuscript.

DATA AVAILABILITY STATEMENT

The processed and cleaned citrus X-ray CT 3D reconstructions can be found in the Dryad repository 10.5061/dryad.34tmpg4n6, along with their separated tissues and associated point clouds and ellipsoidal approximations.

All our code is available at the https://github.com/amezqui3/vitaminC_morphology repository. This includes the image processing pipeline to clean the raw scans and segment the fruit tissues, the computation of tissue volume, the best-fit ellipsoid, and the hypothesis testing of uniform and symmetric distributions on a unit sphere. A collection of Jupyter notebook tutorials is also provided to ease the usage and understanding of the different components of the data processing and data analyzing pipelines. All the image-related scripts are available in python, while the statistical analyses are in R.

ORCID

Erik J. Amézquita https://orcid.org/0000-0002-9837-0397
Michelle Y. Quigley https://orcid.org/0000-0001-5436-6532
Tim Ophelders https://orcid.org/0000-0002-9570-024X
Danelle Seymour https://orcid.org/0000-0003-2250-8377
Elizabeth Munch https://orcid.org/0000-0002-9459-9493
Daniel H. Chitwood https://orcid.org/0000-0003-4875-1447

REFERENCES

Baron, J. H. (2009). Sailors' scurvy before and after James Lind—A reassessment. *Nutrition Reviews*, *67*, 315–332. https://doi.org/10.1111/j.1753-4887.2009.00205.x

Barry, G. H., Caruso, M., & Gmitter, F. G. (2020). Chapter 5 - Commercial scion varieties. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 83–104). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00005-X

Blott, S. J., & Pye, K. (2008). Particle shape: A review and new methods of characterization and classification. *Sedimentology*, *55*, 31–63. https://doi.org/10.1111/j.1365-3091.2007.00892.x

Caruso, M., Smith, M. W., Froelicher, Y., Russo, G., & Gmitter, F. G. (2020). Chapter 7 - Traditional breeding. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 129–148). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00007-3

Clayton, C. R. I., Abbireddy, C. O. R., & Schiebel, R. (2009). A method of estimating the form of coarse particulates. *Géotechnique*, *59*, 493–501. https://doi.org/10.1680/geot.2007.00195

Corey, A. T. (1949). Influence of shape on fall velocity of sandgrains (Master's thesis). Colorado Agricultural; Mechanical College, Fort Collins, CO.

Deng, X., Yang, X., Yamamoto, M., & Biswas, M. K. (2020). Chapter 3 - Domestication and history. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 33–55). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00003-6

Di Marzio, M., Fensore, S., Panzera, A., & Taylor, C. C. (2019). Kernel density classification for spherical data. *Statistics & Probability Letters*, *144*, 23–29. https://doi.org/10.1016/j.spl.2018.07.018

Diaz-Toca, G. M., Marin, L., & Necula, I. (2020). Direct transformation from Cartesian into geodetic coordinates on a triaxial ellipsoid. *Computers & Geosciences*, *142*, 104551. https://doi.org/10.1016/j.cageo.2020.104551

FAO. (2021). *Citrus fruit. Fresh and processed. Statistical bulletin 2020*. Food; Agriculture Organization of the United Nations.

Frank, M. H., & Chitwood, D. H. (2016). Plant chimeras: The good, the bad, and the'Bizzaria'. *Developmental Biology*, *419*, 41–53. https://doi.org/10.1016/j.ydbio.2016.07.003

García-Portugués, E., Navarro-Esteban, P., & Cuesta-Albertos, J. A. (2020). On a projection-based class of uniformity tests on the hypersphere. arXiv:2008.09897

García-Portugués, E., Paindaveine, D., & Verdebout, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *Journal of the American Statistical Association, 115*, 1873–1887. https://doi.org/10.1080/01621459.2019.1665527

García-Portugués, E., Paindaveine, D., & Verdebout, T. (2021). rotasym: Tests for rotational symmetry on the hypersphere. R package version 1.1.3. URL: https://CRAN.R-project.org/package=rotasym

García-Portugués, E., & Verdebout, T. (2021). sphunif: Uniformity tests on the circle, sphere, and hypersphere. R package version 1.0.1. https://CRAN.R-project.org/package=sphunif

Gmitter, F. G., Wu, G. A., Rokhsar, D. S., & Talon, M. (2020). Chapter 1 - The citrus genome. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 1–8). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00001-2

Harris, J. W., & Stöcker, H. (1998). *The handbook of mathematics and computational science* (1st ed.). Springer-Verlag. https://doi.org/10.1007/978-1-4612-5317-4

Hawkes, D. (Ed.). (1985). *The songs of the south: An ancient Chinese anthology of poems by Qu yuan and other poets*. Penguin Books.

Isaac, E. (1959). Influence of religion on the spread of citrus. *Science*, *129*, 179–186. https://doi.org/10.1126/science.129.3343.179

Iwata, H., Nesumi, H., Ninomiya, S., Takano, Y., & Ukai, Y. (2002). The evaluation of genotype × environment interactions of citrus leaf morphology using image analysis and elliptic Fourier descriptors. *Breeding Science*, *52*, 243–251. https://doi.org/10.1270/jsbbs.52.243

Janke, N. C. (1966). Effect of shape upon the settling vellocity of regular convex geometric particles. *Journal of Sedimentary Research*, *36*, 370–376. https://doi.org/10.1306/74D714C4-2B21-11D7-8648000102C1865D

Knight, T. G., Klieber, A., & Sedgley, M. (2001). The relationship between oil gland and fruit development in Washington navel Orange (*Citrus sinensis* L. Osbeck). *Annals of Botany*, *88*, 1039–1047. https://doi.org/10.1006/anbo.2001.1546

Krumbein, W. C. (1941). Measurement and geological significance of shape and roundness of sedimentary particles. *Journal of Sedimentary Research*, *11*, 64–72. https://doi.org/10.1306/D42690F3-2B26-11D7-8648000102C1865D

Langgut, D. (2017). The citrus route revealed: From Southeast Asia into the Mediterranean. *HortScience Horts*, *52*, 814–822. https://doi.org/10.21273/HORTSCI11023-16

Ley, C., & Verdebout, T. (2017). *Modern directional statistics* (1st ed.). Interdisciplinary statistics. Chapman. https://doi.org/10.1201/9781315119472

Li, Q., & Griffiths, J. (2004). Least squares ellipsoid specific fitting. In *Geometric modeling and processing. Proceedings* (pp. 335–340). https://doi.org/10.1109/GMAP.2004.1290055

Luro, F., Curk, F., Froelicher, Y., & Ollitrault, P. (2017). Recent insights on Citrus diversity and phylogeny. In V. Zech-Matterne & G. Fiorentino (Eds.), *AGRUMED: Archaeology and history of citrus fruit in the Mediterranean: Acclimatization, diversification, uses*. Publications du Centre Jean Bérard. https://doi.org/10.4000/books.pcjb.2169

Magiorkinis, E., Beloukas, A., & Diamantis, A. (2011). Scurvy: Past, present and future. *European Journal of Internal Medicine*, *22*, 147–152. https://doi.org/10.1016/j.ejim.2010.10.006

Mahato, N., Sharma, K., Koteswararao, R., Sinha, M., Baral, E., & Cho, M. H. (2019). Citrus essential oils: Extraction, authentication and application in food preservation. *Critical Reviews in Food Science and Nutrition*, *59*, 611–625. https://doi.org/10.1080/10408398.2017.1384716

Mahomoodally, M. F., & Mooroteea, K. (2021). A comparative ethno-religious study of traditionally used medicinal plants employed in the management of cardiovascular diseases. *Journal of Herbal Medicine*, *25*, 100417. https://doi.org/10.1016/j.hermed.2020.100417

Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics* (1st ed.). Probability and statistics. John Wiley & Sons, Ltd., Chichester. https://doi.org/10.1002/9780470316979

Mei, L., Sheng, O., Peng, S., Zhou, G., Wei, Q.-j., & Li, Q. (2011). Growth, root morphology and boron uptake by citrus rootstock seedlings differing in boron-deficiency responses. *Scientia Horticulturae*, *129*, 426–432. https://doi.org/10.1016/j.scienta.2011.04.012

NASS. (2021). *Citrus fruits. 2021 summary*. National Agricultural Statistics Service. United States Department of Agriculture.

Nati, P. (1674). *Florentina phytologica observatio de malo Limonia citrata-aurantia, Florentiae vulgo la bizzaria*. Hippolyti de Naue.

Niklas, K. J. (2004). Plant allometry: Is there a grand unifying theory? *Biological Reviews*, *79*, 871–889. https://doi.org/10.1017/S1464793104006499

Ollitrault, P., Curk, F., & Krueger, R. (2020). Chapter 4 - Citrus taxonomy. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 57–81). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00004-8

Pal, D., Malik, S. K., Kumar, S., Choudhary, R., Sharma, K. C., & Chaudhury, R. (2013). Genetic variability and relationship studies of mandarin (*Citrus reticulata* Blanco) using morphological and molecular markers. *Agricultural Research*, *2*, 236–245. https://doi.org/10.1007/s40003-013-0072-8

Panou, G., Korakitis, R., & Pantazis, G. (2020). Fitting a triaxial ellipsoid to a geoid model. *Journal of Geodetic Science*, *10*, 69–82. https://doi.org/10.1515/jogs-2020-0105

Pewsey, A., & García-Portugués, E. (2021). Recent advances in directional statistics. *Test*, *30*, 1–58. https://doi.org/10.1007/s11749-021-00759-x

Sharma, P. C., Bhatia, V., Bansal, N., & Sharma, A. (2007). A review on Bael tree. *Natural Product Radiance*, *6*, 171–178.

Shrestha, K. B., & Dangol, D. R. (2019). Crops with medicinal, religious and cultural values. In B. K. Joshi & R. Shrestha (Eds.), *Working groups of agricultural plant genetic resources (Apgrs) in Nepal* (pp. 198–204). NAGRC.

Smith, N. M., Ebrahimi, H., Ghosh, R., & Dickerson, A. K. (2018). High-speed microjets issue from bursting oil gland reservoirs of citrus fruit. *Proceedings of the National Academy of Sciences*, *115*, E5887–E5895. https://doi.org/10.1073/pnas.1720809115

Smith, R. J. (2009). Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, *140*, 476–486. https://doi.org/10.1002/ajpa.21090

Sneed, E. D., & Folk, R. L. (1958). Pebbles in the lower Colorado River, Texas a study in particle morphogenesis. *The Journal of Geology*, *66*, 114–150. https://doi.org/10.1086/626490

Spreen, T. H., Gao, Z., Fernandes, W., & Zansler, M. L. (2020). Chapter 23 - Global economics and marketing of citrus products. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 471–493).

Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00023-1

Talon, M., Wu, G. A., Gmitter, F. G., & Rokhsar, D. S. (2020). Chapter 2 - The origin of citrus. In M. Talon, M. Caruso, & F. G. Gmitter (Eds.), *The genus citrus* (pp. 9–31). Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812163-4.00002-4

Thompson, D. W. (1942). *On growth and form* (2nd ed.). Cambridge University Press.

Todd, M. J., & Yıldırım, E. A. (2007). On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, *155*, 1731–1744. https://doi.org/10.1016/j.dam.2007.02.013

Tsagris, M., Athineou, G., Adam, C., Sajib, A., Amson, E., & Waldstein, M. J. (2022). Directional: A collection of functions for directional data analysis. R package version 5.5. URL: https://cran.r-project.org/web/packages/Directional

Tseng, C.-C. (1999). An allegory on allegory: Reading "Ju song" as Qu Yuan's *Ars Poetica*. *Dong Hwa Journal of Humanistic Studies*, *1*, 69–101. https://doi.org/10.6420/DHJHS.199907.0069

Vlahos, L., Isliker, H., Kominis, Y., & Hizanidis, K. (2008). Normal and anomalous diffusion: A tutorial. arXiv:0805.0419.

Voo, S. S., Grimes, H. D., & Lange, B. M. (2012). Assessing the biosynthetic capabilities of secretory glands in *citrus* peel. *Plant Physiology*, *159*, 81–94. https://doi.org/10.1104/pp.112.194233

Vovin, A. (Ed.). (2016). *Man'yōshū (book 18): A new English translation containing the original text*. Kana Transliteration, Romanization, Glossing and Commentary. https://doi.org/10.1163/9789004315600

Vuollo, V., & Holmström, L. (2018). A scale space approach for exploring structure in spherical data. *Computational Statistics & Data Analysis*, *125*, 57–69. https://doi.org/10.1016/j.csda.2018.03.014

Wadell, H. (1932). Volume, shape, and roundness of rock particles. *The Journal of Geology*, *40*, 443–451. https://doi.org/10.1086/623964

Wu, G. A., Sugimoto, C., Kinjo, H., Azama, C., Mitsube, F., Talon, M., Gmitter, F. G., & Rokhsar, D. S. (2021). Diversification of mandarin citrus by hybrid speciation and apomixis. *Nature Communications*, *12*, 4377. https://doi.org/10.1038/s41467-021-24653-0

Wu, G. A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F. R., Carbonell-Caballero, J., Alonso, R., Curk, F., Du, D., Ollitrault, P., Roose, M. L., Dopazo, J., Gmitter, F. G., Rokhsar, D. S., & Talon, M. (2018). Genomics of the origin and evolution of citrus. *Nature*, *554*, 311–316. https://doi.org/10.1038/nature25447

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.