**Rapid report**

Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics

---

**Authors for correspondence:**

Guillaume Chomicki  
Tel: +49 89 17861 285  
Email: guillaume.chomicki@gmail.com

Susanne S. Renner  
Tel: +49 89 17861 257  
Email: renner@lmu.de

Received: 28 July 2014  
Accepted: 23 September 2014

---

**Summary**

- Type specimens are permanently preserved biological specimens that fix the usage of species names. This method became widespread from 1935 onwards and is now obligatory.
- We used DNA sequencing of types and more recent collections of wild and cultivated melons to reconstruct the evolutionary history of the genus *Citrullus* and the correct names for its species.
- We discovered that the type specimen of the name *Citrullus lanatus*, prepared by a Linnaean collector in South Africa in 1773, is not the species now thought of as watermelon. Instead, it is a representative of another species that is sister to *C. ecirrhosus*, a tendril-less South African endemic. The closest relative of the watermelon instead is a West African species. Our nuclear and plastid data furthermore reveal that there are seven species of *Citrullus*, not four as assumed.
- Our study implies that sweet watermelon originates from West, not southern Africa as previously believed, and that the South African citron melon has been independently domesticated. These findings affect and explain numerous studies on the origin of these two crops that led to contradictory results because of the erroneous merging of several distinct species.

---

**Introduction**

In Cucurbitaceae, a plant family that contains some of the World’s economically most important crops (Schaefer & Renner, 2011), species identification is difficult because flowers and fruits are often large, fleshy, and difficult to preserve, and leaves can be extremely variable as is true of most plants with a climbing habit. A prime example is the type specimen of the name *Citrullus lanatus* (Fig. 1), the watermelon, collected c. 1773 near Cape Town by Linnaeus’s disciple and collector Carl Peter Thunberg and deposited in the herbarium of Uppsala (Sweden). It consists of a few crumbled leaves that do not allow its secure assignment to any of several *Citrullus* species with the same kind of leaves. The economic importance of Cucurbitaceae crops has fuelled phylogenetic and phylogeographic research to identify their closest relatives, ancestral areas, and divergence times and have so far focused on pumpkin, zucchini, squashes, bottle gourd, melon and cucumber (Sanjur et al., 2002; Clarke et al., 2006; Sebastian et al., 2010; Kistler et al., 2014). Surprisingly, the watermelon, called *C. lanatus* since the 1930s (Bailey, 1930; Mansfeld, 1959; c. 650 scientific papers in Web of Science, accessed 25 May 2014), and its supposed relatives in the genus *Citrullus* have never been analyzed with modern molecular-phylogenetic methods using verifiably identified material (meaning material in permanent collections available for future re-analysis). It is believed that *Citrullus* has four species (e.g. Plant List at http://www.theplantlist.org, accessed 26 May 2014), that the watermelon is of southern African origin (Vavilov, 1987; Wasylikowa & van der Veen, 2004; Hancock, 2012; Meyer et al., 2012), with either *C. colocynthis* as closest relative or progenitor (Zohary, 1983; Aquino et al., 2000; Zamir, 2001) or the preserving
instead unrelated species, namely *C. lanatus* subsp. *lanatus*, *C. lanatus* subsp. *mucosospermus*, and *C. lanatus* subsp. *vulgaris* (e.g. Guo et al., 2013; the re-sequenced germplasm accessions PI482276, PI482303, PI482326, PI296341 are not part of the gene pool of wild or domesticated watermelon); and (3) the failure until now to identify the sister species of watermelon.

**Materials and Methods**

**Sequencing and alignment**

We constructed a matrix with 11 gene regions based on the nuclear ITS region (ITS1, 5.8S rDNA, ITS2), the trnL intron, trnL-trnF spacer, rpl20-tp5 spacer, trnR-tp5 spacer, Ycf9-trnG spacer, Ycf6-PsbM spacer, and the genes ndhF, rbcL and matK. Vouchers, geographic origin and GenBank accession numbers are shown in Supporting Information Table S1. Except for a cultivated sweet watermelon (voucher Renner 2816), one of several accessions of citron melon (voucher Chomiczki 2), and one of two accessions of *C. mucosospermus* (voucher: Vavilov Research Institute of Plant Industry CIT 204) all material was wild-collected.

We sequenced new accessions of *Citrullus* from silica-dried leaves or herbarium material, and included the new sequences in a matrix of previously sequenced Benincaseae. Total genomic DNA was extracted from c. 20 mg of leaf tissues, using a commercial plant DNA extraction kit (NucleoSpin; Macherey–Nagel, Düren, Germany) according to manufacturer protocols. Polymerase chain reaction (PCR) was performed using Taq DNA polymerase (New England Biolabs, Cambridge, MA, USA) and the same plastid and nuclear primers as Dane & Lang (2004) and Sebastian et al. (2010). PCR products were purified using the ExoSap clean-up kit (Fermentas, St Leon-Rot, Germany), and sequencing relied on Big Dye Terminator kits (Applied Biosystems, Foster City, CA, USA) on an ABI 3130 automated sequencer (Applied Biosystems, Perkin-Elmer). Sequences were edited in Sequencher 5.1 (Gene Codes, Ann Arbor, MI, USA). All new sequences were BLAST-searched in GenBank. Sequence alignment was performed in MAFFT v. 7 in the online server (http://mafft.cbrc.jp/alignment/server; Katoh & Standley, 2013) under standard parameters except for the ITS region which was aligned a Q-INS-i, which takes rRNA secondary structure into consideration and is recommended for the ITS region. Minor alignment errors were corrected manually in Mesquite v. 2.75 (Maddison & Maddison, 2011).

**Phylogenetic analyses and molecular clock dating**

In the absence of statistically supported incongruence (i.e. BS > 80%) based on Maximum likelihood (ML) tree inference, the chloroplast and nuclear data were combined, yielding a matrix of 8618 aligned nucleotides. Maximum likelihood tree inference relied on RAxML-HPC v. 8 (Stamatakis et al., 2008), under the General Time-Reversible model of substitution with Unequal Rates across Sites (GTR + G) (Waddell & Steel, 1996). Statistical support relied on bootstrapping, with 100 replicates under the same model. The analysis was partitioned by gene region (10 partitions for the 11 gene regions, trnL intron and trnl-trnF in the

---

**Fig. 1** The type specimen of *Citrullus lanatus*, collected in South Africa by Linnaeus disciple C. P. Thunberg c. 1773. (a) Type specimen. (b) Thunberg description at the back of the specimen. *Citrullus* leaves are highly variable and yield few good taxonomic characters; the flowers generally preserve poorly, and in dried specimens usually are not helpful for species identification either. Only DNA sequencing could resolve the true identity of this type of specimen.
same partition). Molecular dating analyses relied on BEAST v. 1.8 (Drummond & Rambaut, 2007). BEAST analyses were performed under both strict and uncorrelated lognormal relaxed clocks. The UCLD standard deviation of 0.39 in the relaxed clock analysis indicated that a strict clock fit our data best, and we therefore selected it. We used the GTR + G substitution model with four rate categories and a Yule tree prior. Monte Carlo Markov chains (MCMC) were run for 20 million generations, with log parameters sampled every 10,000 generations. We used Tracer v. 1.6 (Rambaut & Drummond, 2007) to check that the effective sample size (ESS) of all parameters was >200 and that runs had converged. Trees were annotated in TreeAnnotator v. 1.8 (part of the BEAST package) after discarding 10% as burn-in and using a target Maximum clade credibility tree with a posterior probability limit of 0.98; the final tree was visualized in FigTree v. 1.4 (Rambaut, 2006–2009). To calibrate our tree, we used a secondary constraint from the most comprehensive dated phylogeny of the family, which used three fossil and one geological secondary constraint from the most comprehensive dated phylogeny of the family, which used three fossil and one geological calibration, an island with an endemic radiation (Schaefer et al., 2009). In that analysis, the crown age of the (((Peponium (Lagenaria))(Citrus)) clade was dated to 16 ± 3 million yr (Schaefer et al., 2009), and we used this age to calibrate our Citrus tree, assigning it a normal distribution with a mean of 16 and a standard deviation of 1.4, representing the 95% HPD of Schaefer et al. (2009).

Ancestral area and character reconstruction

Our sampling includes accessions from all the main geographic regions in which wild species of Citrullus occur, as well as cultivated forms of sweet watermelon and citron melon. Species ranges for the five species and the four outgroups were coded using online floras, gbif (http://www.gbif.org/) and recent publications (De Winter, 1990; Dane & Lang, 2004; Dane & Liu, 2007; Dane et al., 2007). The geographic range categories were: southern Africa, Tropical Africa and North Africa, reflecting the distribution pattern of our species of interest (see map on Fig. 2b). The colocynth, C. colocynthis, which has a large, anthropogenic distribution extending to India and Australia, was coded as northern Africa in light of a recent phylogeographic study (Dane et al., 2007); the sweet watermelon and the preserving melon were coded as unknown, to reflect their anthropogenic range and unknown origin (following Sebastian et al., 2010; Carvalho & Renner, 2012). Ancestral area reconstruction (AAR) relied on the dispersal-extinction-cladogenesis (DEC) model as implemented in Lagrange (Ree & Smith, 2008; version 20130526) and we used the
ultrametric tree generated in BEAST using a strict clock model. Lagrange has two user-defined input matrices, a so-called adjacency matrix and a dispersal matrix. In the former, we allowed combined ancestral areas except for the combination of northern Africa and southern Africa because no extant species of *Citrullus* has a range spanning the entire African continent. In the dispersal matrix, we assigned migration between disjoint areas (northern and southern Africa) a lower probability (0.5) than migration between adjacent areas (northern and central Africa; central and southern Africa), which was assigned a probability of ‘1’. Python scripts were generated using the Lagrange configurator (http://www.reelab.net/ lagrange/configurator/index) and subsequently run in Lagrange version 20130526.

**Results**

**Molecular phylogeny of *Citrullus***

We generated a maximum likelihood phylogeny from plastid and nuclear data totaling 8618 aligned nucleotides. The genus *Citrullus* placed closest to a clade formed by *Pepomoid* and *Lagenaria* (Fig. 2a), the latter including the botanical gourd (*L. siceraria*), confirming previous findings (Schaefer & Renner, 2011). *Citrullus naudinianus*, from the Namib-Kalahari region, is sister to the other six species, followed by the North African/Western Asian *C. colocynthis* and *C. rehmi*, an annual Namib-Kalahari species (Fig. 2a). *Citrullus amarus*, commonly called tsamma, cow, or citron melon, and its cultivated form, the preserving melon, used to make jams since at least the fifteenth century (Bailey, 1930), are sister to *C. ecirrhosus*, both native to the Namib-Kalahari region. The latter is the only species in the genus lacking tendrils (Fig. 3f inset) and also one of only three perennials, the remaining species all being annuals except for *C. colocynthis* and *C. naudinianus*. As currently circumscribed, *C. lanatus* is therefore not a biological species since one of its members, *C. lanatus* subsp. *lanatus*, is more closely to another species, *C. ecirrhosus*, than it is to any of the other supposed subspecies of *C. lanatus* (Fig. 2a). Our nomenclatural research shows that *C. amarus* is the oldest available name for the citrus or cow melon, with the synonyms *C. caffer* and *C. lanatus* var. *citroides* (all author names and relevant references are found in Supporting Information Table S1).

An important discovery of this study is that the type specimen of the name *Citrullus lanatus* (Thunb.) Matsum. & Nakai (Monodica *lanatus* Thunb.) is not the species that Linnaeus described as *C. vulgaris*, based on cultivated Mediterranean specimens, and not the species whose draft genome has been published in 2013 (Guo *et al.*, 2013). Plastid sequences from the 1773 original South African collection (Fig. 1) are embedded within the *C. amarus* clade (Fig. 2a), whose members all share an apomorphic 30 pb deletion in their trnS-trnG gene sequences (Supporting Information Fig. S1), and is highly supported (maximum likelihood bootstrap 94%).

Genetically, the cultivated watermelon is closest to plants from West Africa (such as the accession from Benin in Fig. 2) that represents the gene pool from which watermelon was domesticated. The sister species of watermelon is *Citrullus mucosospermus*, the egusi melon, ranging from Nigeria to Senegal and described by Fursa (1972, 1983). This species usually has large (c. 2 cm long), white, oleaginous seeds with a black margin (Fursa, 1983) (Fig. 3d). Plants from Benin (such as those sequenced; Fig. 2) closely resemble the type collection of the name *C. mucosospermus* (which is from Ghana) and collections grown from seeds at the Komarov Institute in St Petersburg, Russia, where the economic botanist Fursa worked on *Citrullus*. Different from *C. amarus* and *C. ecirrhosus*, which have hard, white and bitter flesh, *C. mucosospermus* has soft and less bitter flesh, which is often pinkish in the centre (Guo *et al.*, 2013).

**Age and geographic origin of the watermelon clade**

We dated *Citrullus* and its sister group using a strict clock model, as best fitting our dataset (see ‘the Materials and Methods section’), using a secondary calibration from a comprehensive Cucurbitaceae molecular clock study (Schaefer & Renner, 2011). The inferred divergence times indicate that *Citrullus* split from its sister clade 11 ± 3 Million yr ago (Ma); the watermelon species from its West African sister species, *C. mucosospermus*, 3 ± 1 Ma; and the South African species *C. amarus* and *C. ecirrhosus* from each other 2.4 ± 1 Ma (Fig. 2b). Our ancestral area reconstruction, which is based on representatives from all major geographic regions in which wild *Citrullus* occurs, recovers southern Africa (the Namib-Kalahari region) for *C. amarus* (ML probability = 1), but West Africa for the watermelon (ML probability = 0.96; Fig. 2a), fitting with the watermelon having high water requirements (Erdem & Nedim Yüksel, 2003; Orta *et al.*, 2003) and *C. amarus* being adapted to sub-desert areas with xerophytic traits, such as thick leaves (Meeuse, 1962).

**Discussion**

Our well-resolved phylogeny of *Citrullus* and discovery that the type specimen of the name *C. lanatus* is not the crop watermelon of which a draft genome has been published (Guo *et al.*, 2013) has the following consequences. First, our data reject *C. colocynthis* as a close relative, much less progenitor (Zohary, 1983; Aquino *et al.*, 2000; Zamir, 2001), of the watermelon as well as a southern African wild origin of this crop (Vavilov, 1987; Wasylikowa & van der Veen, 2004; Hancock, 2012; Meyer *et al.*, 2012). The latter idea was based on the origin of Thunberg’s plant collected in dunes near Cape Town and here shown not to be a watermelon, but instead a distinct species for which there exist the name *C. amarus* (Fig. 2a). Second, the geographic origin of the crop watermelon likely was in West Africa (Fig. 2b), where wild populations of the crop as well as the watermelon sister species, *C. mucosospermum*, are endemic. The watermelon genome analysis and fluorescent in situ hybridization revealed that number and location of 5S and 45S rDNA sites are identical between watermelon and *C. mucosospermus* (erroneously classified as *C. lanatus* subsp. *mucosospermus*; Guo *et al.*, 2013), but differ from those of the South African *C. amarus* (erroneously classified as *C. lanatus* var. *citroides*; Guo *et al.*, 2013), consistent with our findings. Third, *C. amarus*, the tsamma, cow, or citron melon, and its cultivated form, the preserving melon (Bailey,
1930), is not a wild form of the watermelon (Dane & Lang, 2004; Wasylikowa & van der Veen, 2004; Dane & Liu, 2007) or progenitor of the watermelon (Zeven & Zhukovsky, 1975; Navot & Zamir, 1987; Jeffrey, 2001) but instead is a separate crop species, domesticated independently.

That Linnaeus’s student Thunberg described the plant he collected near Cape Town as a new species makes sense for several reasons. Thunberg lived in Cape Town between April 1771 and March 1775, and during this time undertook three collecting trips, north to Saldanha Bay, east along the Breede Valley through the Langkloof as far as the Gamtoos River and into the Little Karoo. He collected the specimen from which our DNA sequences were obtained ‘in dunis prope Cap locis arenosis,’ (in sandy places in dunes near Cape town), and from Thunberg’s journal it is clear that he would have recognized a true watermelon had he seen one. For example, in his journal he remarks that the Cape colonists cultivated many ‘useful products of the vegetable kingdom’, including ‘melons and water-melons’ (Thunberg, 1986 reprint, p. 320). Since Thunberg clearly knew his teacher’s Linnaeus’s (Linnaeus, 1753) 20-yr earlier description of the watermelon (as Cucurbita citrullus L. = Citrullus baccifer Forsk. 1775 = Citrullus vulgaris Schrad. 1836), he focused on traits setting his new species apart from Linnaeus’s watermelon, these being the rougher leaves (foliis scabris) and hairy fruit (fructu lanato) (Thunberg, 1794; see our Fig. 3e inset, compare the hairy young fruits to the smooth young fruits of sweet watermelon, Fig. 3a inset). The erroneous synonymization of Thunberg’s South African species Citrullus lanatus with Linnaeus’s watermelon (Citrullus vulgaris)
occurred in the 1930s (Bailey, 1930; Mansfeld, 1959) and spread during the 1960s, resulting in today’s universal misapplication of *Citrullus lanatus* for the watermelon. By 2000, the name *Citrullus lanatus* has been misapplied in countless publications in the scientific and applied domain. The best solution now is to condone and legalize the erroneous application of the name *Citrullus lanatus* by conserving it with a type that, while discordant with Thunberg’s original intent, sanctions the current all but universal practice (Renner et al., 2014). The other conceivable alternative, reverting to the once popular *Citrullus vulgaris*, is not a realistic option since that name is preceded by an older, until now overlooked name for the sweet watermelon, *Citrullus batich* Forsk., published in 1775.

In studies of the origin of crop species (Vavilov, 1987; Hancock, 2012; Meyer et al., 2012), the watermelon has long been a riddle because of its supposed South African origin (Vavilov, 1987; Dane & Lang, 2004; Wasylikowa & van der Veen, 2004), but completely absent in the wild there, yet present in southwest Libya c. 5000 yr ago (Wasylikowa & van der Veen, 2004). The finding that the watermelon and its sister species are West African plants suggests that the natural range of watermelon may have extended into Libya or into Egypt during more humid periods of the Pleistocene and Holocene (Schulz, 1987; Schulz, 1991). Alternatively, watermelon seeds may have been traded from West Africa to northern Africa. The illustrations found in Egyptian tombs of watermelon served on a tray suggest that these fruits were eaten raw, perhaps as a dessert (Janick et al., 2007). Seeds found in ancient Egyptian tombs, including that of Thutankhamun (Hepper, 1990) should ideally be studied using ancient DNA approaches. The 30-bp deletion in the plastid trnS-trnG intergenic spacer (Supporting Information Fig. S1) is a genetic marker ideal for barcoding since it reliably distinguishes the citron melon, *Citrullus amarus*, from the sweet watermelon, and our successful amplification from a 1773 South African collection suggests that it could be used for seed identification. The history of the domestication of the sweet watermelon now has to be reconsidered in light of a West African origin, and the search for wild progenitor populations should no longer concentrate on South Africa, but instead West Africa.

**Acknowledgements**

For help with literature, type material, and discussion of watermelon evolution the authors thank W. Greuter, M. Hjertson (UPS), M. Appelhans (GOET), K. Pistrick (GAT), V. Dorofeyev (LE), J. Manning, P. Goldblatt, G. Nesom, H. Paris, F. Dane, R. Jarrett (USDA), and L. Ammon. The authors thank F. Bayada and L. Ponsaty for providing them with cultivated preserving melon, E. Achigan-Dako, N. Jürgens, C. McGregor, H. Paris for important photographs, and R. E. Ricklefs for discussion.

**References**


Carvalho F, Renner SS. 2012. A dated phylogeny of the papaya family (Caricaceae) reveals the crop’s closest relatives and the family’s biogeographic history. *Molecular Phylogenetics and Evolution* 65: 46–53.


Rambaut A. 2006–2009. *FigTree*. Tree figure drawing tool, version 1.2.3. URL: http://beast.bio.ed.ac.uk [accessed 1 June 2014].


© 2014 The Authors

*New Phytologist © 2014 New Phytologist Trust*


Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Synapomorphic 30 bp deletion in the trnS-trnG intron of the *Citrullus ecirrhosus* and *C. amarus* clade.

Table S1 Vouchers and Genbank accession numbers for the material of this study

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.