Shedding new light on the origin and spread of the brinjal eggplant (Solanum melongena L.; Solanaceae) and its wild relatives

Aubriot, Xavier

2018-07

Aubriot , X , Knapp , S , Syfert , M , Poczai , P & Buerki , S 2018 , ' Shedding new light on the origin and spread of the brinjal eggplant (Solanum melongena L.; Solanaceae) and its wild relatives ' , American Journal of Botany , vol. 105 , no. 7 , pp. 1175-1187 . https://doi.org/10.1002/ajb2.1133

http://hdl.handle.net/10138/241064
https://doi.org/10.1002/ajb2.1133

unspecified
acceptedVersion

Aubriot, X., S. Knapp, M. M. Syfert, P. Poczai, and S. Buerki. 2018. Shedding new light on the origin and spread of the brinjal egg-plant (*Solanum melongena* L.) and its wild relatives. *American Journal of Botany* 105(7): 1175–1187. doi:10.1002/ajb2.1133

Shedding new light on the origin and spread of the brinjal eggplant (*Solanum melongena* L.) and its wild relatives

Xavier Aubriot[1,2,5] , Sandra Knapp[1] , Mindy M. Syfert[1], Péter Poczai[3], and Sven Buerki[1,4,5]

Manuscript received 16 February 2018; revision accepted 4 May 2018.

[1] Department of Life Sciences,Natural History Museum, Cromwell Road, London SW7 5BD, England, UK

[2] Unité Mixte de Recherche 6553 Écosystèmes, Biodiversité, Évolution (ECOBIO), Observatoire des Sciences de l'Univers de Rennes, Centre National de la Recherche Scientifique, Université de Rennes 1, Rennes CEDEX, France

[3] Botany Unit, Finnish Museum of Natural History, University of Helsinki, PO Box 7, Helsinki FI-00014, Finland

[4] Department of Biological Sciences,Boise State University, 1910 University Drive, Boise, Idaho 83725, U.S.A.

[5] Authors for correspondence (e-mail: x.aubriot@gmail.com; svenbuerki@boisestate.edu)

PREMISE OF THE STUDY: While brinjal eggplant (*Solanum melongena* L.) is the second most important solanaceous fruit crop, we lack firm knowledge of its evolutionary relationships. This in turn limits efficient use of crop wild relatives in eggplant improvement. Here, we examine the hypothesis of linear step-wise expansion of the eggplant group from Africa to Asia.

METHODS: We use museum collections to generate nuclear and full-plastome data for all species of the Eggplant clade. We combine a phylogenomic approach with distribution data to infer a biogeographic scenario for the clade.

KEY RESULTS: The Eggplant clade has Pleistocene origins in northern Africa. Dispersals to tropical Asia gave rise to *Solanum insanum*, the wild progenitor of the eggplant, and to African distinct lineages of widespread and southern African species. Results suggest that spread of the species to southern Africa has been recent and likely facilitated by large mammalian herbivores, such as the African elephant and impala feeding on *Solanum* fruit.

CONCLUSIONS: Rather than a linear 'Out Of Africa' sequence, our results are more consistent with an initial dispersal event into Asia, and subsequent wide dispersal and differentiation across Africa driven by large mammalian herbivores. Our evolutionary results will affect future work on eggplant domestication and affect the use of wild relatives in breeding of this increasingly important solanaceous crop.

KEY WORDS biogeography; chloroplast genome; crop wild relatives (CWRs); dispersal; eggplant (aubergine); mammalian herbivores; molecular dating; phylogenomics; Solanaceae.

Crop wild relatives (CWRs) are likely to play a significant role in securing 21st century food security (Dempewolf et al., 2014), because of their potential use in breeding to withstand adverse effects of climate change, increasing scarcity of nutrients, water and other inputs, and new pests and diseases (Hopkins and Maxted, 2010). Understanding the phylogenetic relationships, geographic origins, and spread of CWRs is critical for improving their use in future plant breeding in the face of environmental and climate change.

The brinjal eggplant, *Solanum melongena* L. (also known as the aubergine; Solanaceae) is the second most economically important solanaceous fruit crop after the tomato, with a total production of approximately 51 Mt in 2016 (FAO, 2017). The brinjal eggplant has a long history of cultivation (see Daunay and Janick, 2007; Wang et al., 2008), and is thought to have been domesticated in Asia sev-eral times (Daunay and Hazra, 2012; Meyer et al., 2012). Increasing genetic diversity in the cultivated crop is a major target in brinjal eggplant breeding (Muñoz-Falcón et al., 2009). Improvements in yield, fruit and postharvest quality (Gramazio et al., 2014), resist-ance to plant pathogens (Pautasso et al., 2012; Bebber et al., 2014), and introduction of abiotic stress traits such as tolerance to in-creased drought, temperatures, and extreme precipitation are all important, especially in the face of climate change model predic-tions of higher temperatures and extreme weather patterns world-wide (Cattivelli et al., 2008). Crop wild relatives are an increasingly important source for traits and genes for brinjal eggplant improve-ment (Daunay et al., 1991; Muñoz-Falcón et al., 2009; Hurtado et al., 2012; Kaushik et al., 2016), and knowledge about their distri-bution, traits, and threat status has been the focus of recent studies (Syfert et al., 2016; S. Kinsell, X. Aubriot, N.P. Castañeda-Álvarez, S. Knapp, J. Prohens, T. Särkinen, unpublished data).

Solanum L., with ca. 1400 species, is one of the largest genera of flowering plants (Frodin, 2004) and has been the subject of much recent taxonomic and phylogenetic work (e.g., see references in Särkinen et al., 2013; Vorontsova and Knapp, 2016). The species-level taxonomy of *S. melongena* and its wild relatives has long been recognized as complex (Lester and Hasan, 1991; Daunay and Hazra, 2012), but recent revision and clarification of taxonomic limits in the group (Knapp et al., 2013; Vorontsova and Knapp, 2016) has provided a set of species identities for use in further analyses. Previous analyses recognized only two species in the Eggplant clade (e.g., Lester and Hasan, 1991; Weese and Bohs, 2010), but further work with African solanums suggested the clade comprised 10 species characterized in part by their andromonoecious breeding system (see Table 1 in Knapp et al., 2013). Even though the number of taxa recognized increased, several extremely widespread species (e.g., *S. campylacanthum* A.Rich. across eastern Africa) include much geographical variation, suggesting a role for ecological structuring in these taxa (Ranil et al., 2017; Vorontsova and Knapp, 2016).

Solanum melongena and its wild relatives are members of the largest clade of the genus, the Leptostemonum clade (the 'spiny' solanums; Bohs, 2005; Weese and Bohs, 2010; Särkinen et al., 2013), and within that, a monophyletic group comprising almost all Old World spiny solanums (Stern et al., 2011; Vorontsova et al., 2013; Aubriot etal., 2016). The Eggplant clade has been recognized as monophyletic in previous analyses (e.g., Stern et al., 2011; Weese and Bohs, 2010), but sampling of Old World taxa was limited. Using the same set of germplasm accessions Weese and Bohs (2010) tested Lester's (Lester and Hasan, 1991) evolutionary scenario of brinjal eggplant evolution, but showed that the weedy species *S. linnaeanum* Hepper & P.M.-L.Jaeger also belonged to the group. Their analysis (Weese and Bohs, 2010) supported the scenario of an African origin for the group with step-wise expansion to tropical Asia and migration back to the Middle East as feral forms. Vorontsova et al. (2013) expanded the data set by including many African taxa and still recovered a monophyletic Eggplant clade, but with little internal resolution. Their results placed the narrow Kenyan endemic *S. agnewiorum* Voronts. within the monophyletic Eggplant clade—a surprising re-sult given its small fruit and hermaphroditic flowers. Inclusion of more species from Africa and southeast Asia (Aubriot et al., 2016) revealed that two additional African species with hermaphroditic flowers were members of the monophyletic group (*S. lanzae* J.-P. Lebrun & Stork and *S. usambarense* Bitter & Dammer). These stud-ies used both plastid and nuclear molecular markers, and although the circumscription of the Eggplant clade improved, resolution within it was still poor. Although the brinjal eggplant and its wild relatives have clear morphological, molecular, and ecogeographical differences (see Meyer et al., 2012; Knapp et al., 2013; Ranil et al., 2017), they have been difficult to resolve using traditional molecu-lar markers (Vorontsova et al., 2013; Aubriot et al., 2016).

High-throughput sequencing and DNA library preparation improvements have made obtaining whole plastid genomes from total DNA extraction affordable for many nonmodel organisms driving a rapid increase in the number of genomes in public sequence repositories (see Tonti-Filippini et al., 2017). It has been suggested that complete plastid genomes can be used as 'superbarcodes' thus providing an abundance of characters for construction of better-resolved maternal phylogenies (Williams et al., 2016).

Here we use sequences of whole chloroplast genomes to improve circumscription of species and resolution of relationships in the Eggplant clade, and to test hypotheses concerning the origin of the close relatives of the cultivated eggplant, *Solanum melongena*. To account for morphological and geographical variability of the species, we first used Sanger data (both from nuclear and chloro-plast genomes) to test species circumscriptions for four widespread species of the Eggplant clade; then, based on this first analysis, we generated a plastome phylogeny of the clade. This phylogeny was used to check for incongruence between nuclear and plastid data, and to reassess Lester and Hasan's (1991) hypothesis of linear step-wise migration from Africa to Asia. This was done using ancestral areas reconstruction to investigate biogeography and dispersal of the Eggplant clade as currently circumscribed. We relate dispersal and spread in widespread eastern African species to large mam-malian herbivores that feed on *Solanum* fruits (Pringle et al., 2014; Kartzinel et al., 2015). Our results have implications for future re-search on wild sources of genetic variability for eggplant improve-ment, particularly in the context of global climate change and efforts to develop the genetic basis for sustainable pest control.

MATERIALS AND METHODS

Taxon sampling

Sampling for preliminary Sanger analysis —For Sanger analysis, we sampled across the distribution range of four widespread species (*Solanum campylacanthum*, *S. cerasiferum* Dunal, *S. incanum* L., and *S. insanum* L.) to test species monophyly and reveal any geographic patterns. For the widespread and weedy eastern African species *S. campylacanthum*, we sampled three specimens from the northernmost part of its distribution area (Ethiopia), one from Tanzania, and two from the southernmost part (Namibia and South Africa). Both *S. cerasiferum* and *S. incanum* are dis-tributed across northern sub-Saharan Africa, with *S. incanum* also extending into regions of the Middle East, and central Asia as far as Pakistan (Vorontsova and Knapp, 2016). For *S. cerasiferum*, we used specimens from Ethiopia and the Central African Republic, and for *S. incanum*, we used one specimen from the westernmost (Burkina Faso), two from the center (Kenya and Oman), as well as one from the easternmost (Pakistan) part of the distribution range. *Solanum insanum*, the wild progenitor of the cultivated eggplant (Meyer et al., 2012; Ranil et al., 2017), is distributed across tropical Asia, but also occurs in Madagascar; it does not occur in continen-tal Africa or the Middle East (Ranil et al., 2017). We sampled two specimens from tropical Asia (China and India) and one individ-ual from Madagascar. Our sampling is not designed to explore the structure of relationships between the populations of *S. insanum* and *S. melongena*; this will require population-level sampling across the range of both taxa (e.g., Meyer et al., 2012; J.-X. Wang, personal communication).

To place our study within the wider framework of Old World spiny solanums, we also sampled closely related species (sensu Vorontsova et al., 2013, Aubriot et al., 2016). A clade of four tropical Asian species (" *Solanum violaceum* and relatives" sensu Aubriot et al., 2016), and three African species (*S. nigriviolaceum* Bitter, *S. polhillii* Voronts. and *S. supinum* Dunal) were found to be closely related to the Eggplant clade (Vorontsova et al., 2013; Aubriot et al., 2016); therefore we included in our sampling one accession of *S. violaceum* Ortega and one accession for each of these three African species. Two informal groupings of a large number of African and Asian species, the 'Anguivi grade' and the 'Climbing clade' (Vorontsova et al., 2013; Aubriot et al., 2016) were also sampled in order to bet-ter test the circumscription of the Eggplant clade. We selected six

TABLE 1. Summary of GenBank accession numbers, species name, collection information and plastome composition and coverage for taxa used in the NextGen analyses. Herbarium codes in column 4 follow Index Herbariorum (Thiers, 2017). Abbreviation in columns 6 to 10 are as follows: herb. = herbarium material; sil. = silica gel-dried material; LSC = large single copy region; SSC = Small single copy region; IRs = inverted repeat regions.

GenBank accessions	Name in topologies	Species name	Voucher	Collection location	Material	Total length	LSC	SSC	IRs	Mean coverage
MH283717	Solanum aethiopicum	*Solanum aethiopicum* L.	Vorontsova et al. 156 (BM)	Tanzania, Morogoro, Tegetero	sil.	155,608	86,219	18,563	25,413	152.5
MH283720	Solanum agnewiorum	*Solanum agnewiorum* Voronts.	Vorontsova et al. 195 (BM)	Kenya, Eastern, Nyambene	herb.	155,586	86,163	18,587	25,418	19.5
MH283724	Solanum anguivi	*Solanum anguivi* Lam.	Vorontsova et al. 1302 (K)	Madagascar, Fianarantsoa, Itremo Massif	sil.	155,577	86,188	18,559	25,415	124.6
MH283701	Solanum aureitomentosum	*Solanum aureitomentosum* Bitter	Chase 5271 (BM)	Zimbabwe, Southern Rhodesia, Odzi	herb.	155,303	85,950	18,497	25,428	38
MH283718	Solanum campylacanthum	*Solanum campylacanthum* Hochst. ex A.Rich.	Vorontsova et al. 158 (BM)	Tanzania, Tanga, Mkata Kideleko	sil.	155,017	85,667	18,506	25,422	54.5
MH283703	Solanum cerasiferum	*Solanum cerasiferum* Dunal	Fay 2715 (K)	Central African Republic, Bamingui-Bangoran, Bozobo	herb.	155,413	86,129	18,516	25,384	17.2
MH283716	Solanum dasyphyllum	*Solanum dasyphyllum* Schumach. & Thonn.	Vorontsova et al. 151 (BM)	Tanzania, Morogoro, Tegetero	sil.	155,716	86,314	18,536	25,433	62.8
MH283709	Solanum glabratum	*Solanum glabratum* Dunal	Miyazaki 606I28 (A)	Saudi Arabia, Asir, Abah	herb.	155,368	86,081	18,517	25,385	20.2
MH283721	Solanum incanum 1	*Solanum incanum* L.	Vorontsova et al. 203 (BM)	Kenya, Eastern, Laisamis	herb.	155,657	86,276	18,539	25,421	105.3
MH283713	Solanum incanum 2	*Solanum incanum* L.	Sanou & Traore BUR48 (K)	Burkina Faso, Houet, Saguere	herb.	155,617	86,269	18,508	25,420	92.5
MH283711	Solanum insanum	*Solanum insanum* L.	Sampath Kumar et al. 941 (MH)	India, Tamil Nadu, Thamaraikulam	sil.	155,570	86,240	18,498	25,416	254.8
MH283723	Solanum lanzae	*Solanum lanzae* J.-P.Lebrun & Stork	Vorontsova et al. 783 (K)	Kenya, Rift Valley, Kifuko Ranch	sil.	155,269	86,253	18,198	25,409	175
MH283702	Solanum lichtensteinii	*Solanum lichtensteinii* Willd.	Crawford et al. FC139 (K)	Namibia, Kunene, Outjo	herb.	155,574	86,217	18,515	25,421	30.3
MH283704	Solanum linnaeanum	*Solanum linnaeanum* Hepper & P.-M.L. Jaeger	Jury et al. 13209 (BM)	Morocco, Rabat-Salé-Kénitra, Moulay Bousselham	herb.	155,574	86,219	18,509	25,423	73.5
MH283714	Solanum macrocarpon	*Solanum macrocarpon* L.	Tepe et al. 2770 (BM)	Tanzania, Morogoro, Ruvu	sil.	155,820	86,398	18,554	25,434	55.5
MH283708	Solanum melongena 1	*Solanum melongena* L.	Meeboonya et al. RM294 (BM)	Thailand, Chumphon, Pang Wan	sil.	155,583	86,250	18,501	25,416	44.6
MF818319	Solanum melongena 2	*Solanum melongena* L.	Poczai H07653 (H)	Finland, Helsinki, Kumpula Botanical Garden	sil.	155,569	86,236	18,501	25,416	800
MH283715	Solanum polhillii	*Solanum polhillii* Voronts.	Vorontsova et al. 15 (BM)	Kenya, Rift Valley Mt. Suswa	herb.	155,422	86,044	18,544	25,417	196.9
MH283722	Solanum richardii	*Solanum richardii* Dunal	Vorontsova et al. 633 (K)	Madagascar, Fianarantsoa, Vangaindrano	sil.	155,187	86,113	18,244	25,415	118
MH283706	Solanum rigidum	*Solanum rigidum* Lam.	Malato-Beliz 141 (LISC)	Cape Verde, Maio	herb.	155,594	86,244	18,510	25,420	10.2
MH283705	Solanum supinum	*Solanum supinum* Dunal	Kabelo 170 (K)	Botswana, Ghanzi, Okwa Valley	herb.	155,769	86,362	18,549	25,429	111.9
MH283707	Solanum trilobatum	*Solanum trilobatum* L.	Meeboonya et al. RM243 (BM)	Thailand, Samut Songkhram, Bueng Yi San	sil.	155,289	86,218	18,245	25,413	95.8
MH283710	Solanum umtuma	*Solanum umtuma* Voronts. & S.Knapp	Nevhutalu 921 (K)	South Africa, KwaZulu-Natal, Swart Umfolozi	herb.	155,541	86,173	18,514	25,427	29.5
MH283719	Solanum usambarense	*Solanum usambarense* Bitter & Dammer	Vorontsova et al. 166 (BM)	Tanzania, Tanga, Magamba Forest Reserve	sil.	155,411	86,028	18,551	25,416	42.3
MH283712	Solanum violaceum	*Solanum violaceum* Ortega	Sampath Kumar et al. 945e (MH)	India, Tamil Nadu, Palar Dam	sil.	155,748	86,309	18,609	25,415	147.2

species of the ‘Anguivi grade,’ including the two cultivated African solanums, the Gboma eggplant (*S. macrocarpon* L.), and the Scarlet eggplant (*S. aethiopicum* L.), as well as their wild progenitors (*S. dasyphyllum* Schumach. & Thonn. and *S. anguivi* Lam., respectively). Most species of the ‘Anguivi grade’ have geographic distributions centered in continental eastern Africa, but several species occupy disjunct regions such as the Seychelles and India (see Vorontsova et al., 2013; Vorontsova and Knapp, 2016; Aubriot et al., 2016 for details on geographic distributions). To identify potential long-distance dispersal events, we sampled one species with a tropical Asian distribution (*S. trilobatum* L.) and another that occurs within the Middle Eastern region (*S. glabratum* Dunal). We used *S. richardii* Dunal, an eastern African scandent shrub from the ‘Climbing clade’ (sensu Vorontsova et al., 2013), to root all the trees. For this preliminary Sanger analysis, our sampling included 36 accessions, representing all 13 species of the Eggplant clade (25 accessions), and 11 additional spiny solanum species from Africa and Asia.

Sampling for whole chloroplast genome analysis —After using Sanger sequencing for initial assessment of species delimitation in widespread species, we generated full plastome data for all species of the Eggplant clade, with the exception of the cultivated brinjal eggplant for which we had two sets of Illumina reads at our dis-posal. If a species was not shown to be monophyletic based on pre-liminary Sanger analysis, one accession per lineage was included. All outgroup species used in the preliminary analysis were included in the full-plastome analysis, with the exception of *Solanum nigrivi-olaceum* for which we had insufficient DNA extract available.

DNA extraction, amplification, sequencing, and data assembly

Silica gel dried or herbarium material was ground using a Mixer Mill MM 300 with glass beads (Qiagen Inc., Valencia, CA, USA) and with molecular grade sand. Total genomic DNA was extracted using the same two-step protocol as described by Aubriot et al. (2016). We also extracted chloroplast DNA from fresh leaves of *Solanum melongena* ‘Black Beauty’ following Shi et al. (2012).

Amplified DNA regions used in the preliminary Sanger analysis were selected based on previous broad-scope studies on *Solanum* phylogenetics (Särkinen et al., 2013; Aubriot et al., 2016). These latter have shown that the plastid intergenic spacer *ndhF-rpL32*, the internal transcribed spacer (ITS) and the granule-bound starch synthase I (GBSSI) or *waxy* gene are particularly well-fitted for resolving phylogenetic relationships at the species level. Considering the abundant literature on DNA barcoding of land plants and on phylogenetics of *Solanum*, the use of these three rapidly evolving and a priori unlinked DNA regions should be reliable for testing current species boundaries. We downloaded 62 sequences from GenBank, and generated 61 for this study (see Appendix 1). For each additional accession, the three DNA regions were amplified and sequenced following the same procedures and techniques as those described in Aubriot et al. (2016). Sanger reads were assembled and edited in Geneious v.7.1.9 (Biomatters Ltd., Auckland, New Zealand). All sequences were automatically aligned with MAFFT v.7.305 (Katoh and Standley, 2013), using the L-INS-i algorithm. Newly generated sequences are deposited in GenBank and listed in Appendix 1.

Starting quantities of total genomic DNA for the 25 samples used in the phylogenomic study were determined by measurement with a Qubit 2.0 Fluorometer (Life Technologies Corp., Carlsbad, California, USA). The DNA samples were diluted to approximately 3.8 ng/µl and sheared by sonication using a M220 platform (Covaris, Woburn, MA, USA); DNA extracted for *Solanum cerasiferum* was already fragmented and was not subjected to further sonication. Libraries were prepared using the TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA) following manufacturer instructions. Sequencing was performed at the Natural History Museum (London) sequencing facility using the Illumina MiSeq platform (2 × 250 bp with library insert size of approximately 500–700 pb) and loading concentration of 6 pM. Illumina adapters were removed, and all reads were quality trimmed and quality filtered using Trimmomatic 0.35 (Bolger et al., 2014) with default settings.

Plastome assembly, annotation, and alignment

All paired reads were *de novo* assembled with IDBA-UD 1.1.1 (Peng et al., 2012), using maximum *k*-value = 250 bp and minimum *k*-value = 50 bp. Resulting contigs were uploaded into Geneious v.7.1.9 (Kearse et al., 2012) where they were quality-checked and length-filtered. Segments of the *de novo* plastomes were then circularized by matching end points. A high coverage plastid genome (800×) was also assembled *de novo* for *Solanum melongena* ‘Black Beauty’ from the collected reads using the built-in Geneious assembler platform with zero mismatches or gaps allowed among the reads. A similar procedure was conducted using Velvet v1.2.10 (Zerbino and Birney, 2008) with *k*-mer length 37, minimum contig length 74, and default settings by applying 800× upper coverage limit. The results of the two *de novo* methods were compared and inspected. Sanger-based gap closure and inverted repeat (IR) junction verification was car-ried out following Moore et al. (2007).

Sequence inconsistencies were checked by mapping each set of reads to their respective plastome as described in Wysocki et al. (2015). Assembled plastomes were aligned to the completely annotated bittersweet (*Solanum dulcamara* L.) plastome available from GenBank (NC035724; Amiryousefi et al., 2018). Annotations were transferred from the references to the assembled plastomes using Geneious v.7.1.9 with a similarity threshold of 80%, and were then checked and refined manually. All 25 plastomes were aligned with MAFFT v.7.305 using the FFT-NS-2 algorithm, and the alignment was inspected for structural mutations. The pairwise distances between each pair of plastomes were calculated and analyzed using the Kimura Maximum Composite Likelihood model implemented in MEGA 7.0 (Kumar et al., 2016).

Phylogenetic analyses and dating

Sanger and genomic alignments were subjected to both maximum likelihood (ML) and Bayesian inference (BI) analyses. All analyses were run via the CIPRES Science Gateway (Miller et al., 2010) using RAxML-HPC v.8.1.24 (Stamatakis, 2014) for ML and MrBayes 3.2.6 (Huelsenbeck and Ronquist, 2001) for BI. In all analyses, gaps were treated as missing data (N) and *Solanum richardii* served as the outgroup. Sanger alignments were analyzed both separately and as combined data sets. RAxML analyses were run by using the rapid bootstrap algorithm with 1000 bootstrap iterations. The GTR + G rate substitution model was applied to each matrix to fit the sub-stitution models implemented in BI (see below); the proportion of invariable sites was not estimated because it is not recommended by the RAxML developer (Stamatakis, 2016). Bootstrap replicate trees were used to draw bipartitions on the optimal ML tree; nodes with bootstrap support values (BS) below 50% were collapsed to obtain the ML 50% majority-rule tree. For BI analyses, the best-fitting nucleotide substitution model inferred with MrModeltest v.2.3 (Nylander, 2004) via estimation of the Akaike information criterion was assigned to each of DNA region. The regions *ndhF-rpL32* and ITS followed a GTR + I + G substitution model, whereas the GTR + I model was better fitted for the *waxy* region. MrBayes analyses constituted of two independent parallel runs of four Markov chains each, run for 10 million generations, and trees were sampled every 1000 generations. Adequate mixing of Markov chains and convergence of the two runs were confirmed with Tracer v1.6 (Rambaut et al., 2014). After a 10% burn-in, remaining trees were used to generate a 50% Bayesian majority-rule consensus tree, in which nodes with posterior probability values below 50% were collapsed. Congruence between the topologies obtained for each of the three markers was visually inspected before concatenating them into a combined matrix that was divided into three partitions, to which the best-fitted substitution models were applied. The combined ma-trix was run under the same conditions detailed above for the single marker matrices.

We used the same BI and ML approaches to analyze the phylogenomic data set. The plastome matrix was analyzed with MrModeltest, and the GTR + I + G nucleotide substitution model was selected as best fitting the data. For the same reasons as stated above, this model was implemented in MrBayes whereas the GTR + G rate substitution model was preferred for RAxML analyses. All conditions of MrBayes and RAxML runs were identical to those im-plemented for analyses of Sanger alignments.

Relative divergence times, together with the topology, were estimated using BEAST v.1.8.4 (Drummond et al., 2012), with a GTR + I + G rate substitution model applied to the plastome matrix. A Yule speciation tree prior and a relaxed uncorrelated clock-model that allows rates to vary independently along branches (Drummond et al., 2006) were used with all other parameters set to default. We used two temporal constraints to calibrate the BEAST analyses: (1) age of the Eggplant clade (mean = 1.54 Myr; standard deviation = 0.5), (2) age of the root (mean = 3.38 Myr; standard deviation = 0.5). These were derived from the Solanaceae-wide phylogeny of Särkinen et al. (2013), with uncertainty regarding these dates incorporated by assigning normal prior distributions to the two calibration points (Couvreur et al., 2008; Evans et al., 2014). Four independent BEAST analyses were run, each 100 million generations, sampling every 10,000 generations. Convergence of all parameters was assessed using Tracer 1.5 (Rambaut et al., 2014), the first 10% of each of the four Markov chains was removed as burn-in, and chains were combined with LogCombiner 1.7.2. (Drummond et al., 2012) to calculate the maximum clade credibility tree.

Biogeographic inference and dispersal analysis

Species distributions were determined from collection data stored in Solanaceae Source Database (2017, http://www.solanaceae-source.org/); cultivated species (*Solanum melongena*, *S. aethiopicum*, *S. macrocarpon*) were excluded from the biogeographic analysis. Because several species are widespread, with distribution areas that encompass all of eastern Africa (*S. campylacanthum*) or that extend from western Africa to Pakistan (*S. incanum*), circumscription of biogeographic areas that are meaningful in terms of patterns of species distribution and that are coherent climatically and biogeographically speaking is challenging. We defined seven biogeographic areas based on climatic and biogeographic evidence: (A) Tropical Asian region; (B) Somalian and Middle East region; (C) Sudanian region; (D) Congolian region; (E) Zambesian region; (F) Southern African region; and (G) Malagasy region (see Fig. 1A). Four of these (C, D, E, and F) directly correspond to those of Linder et al. (2012). Linder's regions are statistically defined using patterns from groups of vascular plants and vertebrates, and were shown to be robust between groups. This biogeographic framework is well suited for these solanums, because 10 of the 23 taxa included in our analysis are restricted to one region, and several clades are char-acteristic of a set of connected biogeographic regions (Southern African clade; Fig. 1B).

Areas around the Tanzania/Kenya border in eastern Africa encompass the Albertine Rift and Kilimanjaro Massif and are a mountainous zone of great geological complexity. These areas have been previously assigned to either the Somalian or the Zambesian re-gions, depending on the type of data used (White, 1983, 1993; Linder etal., 2012). This region has also been identified as a center of diver-sity for African spiny solanums (Syfert et al., 2016; Vorontsova and Knapp, 2016). Two of our sampled species, *Solanum polhillii* and *S. usambarense*, are restricted to this transitional region (Vorontsova and Knapp, 2016) and assigning these to a particular biogeographic region was challenging; therefore we accounted for this complex-ity by assigning these two species to both the Somalian and the Zambesian regions.

Because the Middle East acts as a biogeographic crossroads between Asia and Africa, its biota is of composite origin (Buerki et al., 2012), and precise delimitation of its boundary is challenging. Of the species included in our analysis, only *Solanum glabratum* and *S. incanum* are present in the Middle East; both have distribution ranges centered on northeastern Africa (i.e., the Sudanian region), with few scattered populations in the Middle East, generally along the coast. Among all spiny solanums native to the Middle East re-gion, two are distributed from eastern tropical India to the Arabian Peninsula (*S. pubescens* Willd., *S. virginianum* L.) while others are restricted to dry regions between northeastern Africa and north-western India (*S. cordatum* Forssk., *S. forskalii* Dunal, *S. glabra-tum*, *S. incanum*, and *S. platacanthum* Dunal). Notwithstanding their current distribution, all these species have African rather than Asian affinities (Aubriot et al., 2016). In addition, the Red Sea is un-likely to act as strong barrier to dispersal; it is only about 25 km wide at its southernmost and narrowest (natural) point. Based on these observations, and using a conservative approach for our bio-geographical analysis, we decided to combine the Middle East with the Sudanian region, hence delimiting our region B. We used the eastern edge of the Thar Desert of India as the eastern boundary of zone B (Sudanian region and Middle East). This delimitation corre-sponds to the western limit of the Asian monsoon climatic regime, which therefore acts as a proxy for the limit between the tropical areas of region A (Tropical Asia) and the mostly dry areas of region B (Holt et al., 2013).

The history of Madagascar and of its neighboring archipelagos (Mascarenes, Granitic Seychelles, and Comoros) is characterized by complex geological and climatic events that accompanied their iso-lation from Africa and the Indian subcontinent. The Malagasy en-demic flora has a composite origin, with elements from Africa, the Indian subcontinent, and South-East Asia (Buerki et al., 2013). The spiny solanums endemic to Madagascar (not sampled here) form a monophyletic group within the larger Old World lineage of spiny solanums, but are not closely related to the Eggplant clade (Vorontsova et al., 2013; Aubriot et al., 2016). We sampled four widespread species

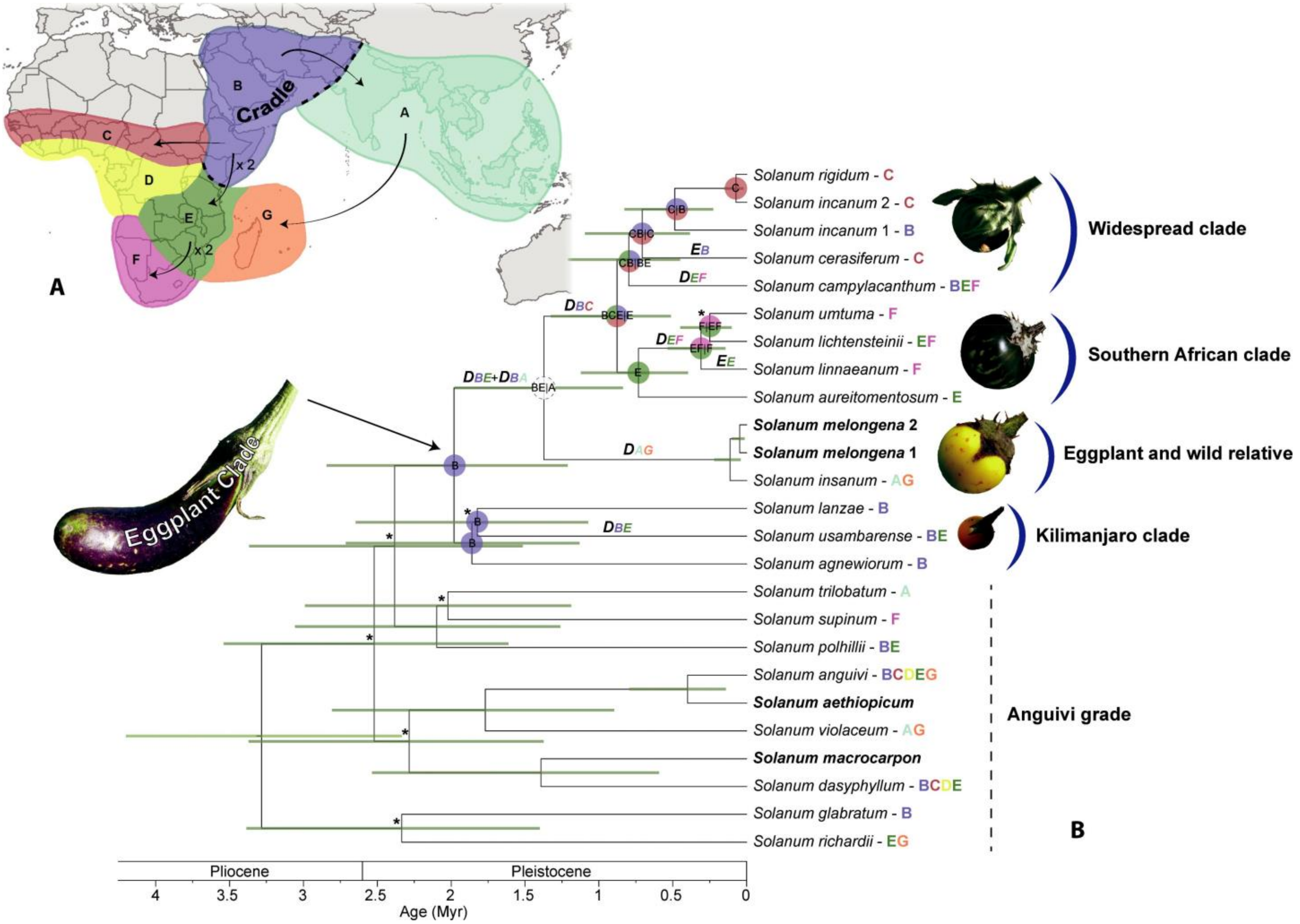

FIGURE 1. Phylogeny and biogeography of the Eggplant clade based on whole chloroplast genome sequences. (A) Map showing the seven biogeographic areas used to infer the biogeographic history of the Eggplant clade. Dotted lines indicate instances of complex biome transitions. Arrows illustrate the dispersal events inferred from the biogeographic analysis. (B) Full-plastome dated phylogeny of the Eggplant clade (consensus of 4 BEAST analyses; 159,227 bp matrix). All nodes have PP ≥ 0.97 and BS ≥ 87% except for the nodes designated with *. Green bars correspond to 95% HDP intervals for the dates of the nodes (Myr). Geographical distribution for each terminal is indicated using the biogeographic regions subdivision. The most probable ancestral area is figured at each node of the Eggplant clade; high levels of biogeographic uncertainty are indicated with dotted lines. Pie-charts with relative probabilities of ancestral states at each node are provided in Appendix S4a,b. Hypothesised dispersal (*D*) and extinction (*E*) events are figured above the branches of the tree. Within the Eggplant clade, clades are given provisional names based on their distributions. Names of cultivated species are in bold. Pictures from left to right and top to bottom are fruits of the cultivated eggplant (*Solanum melongena*), *S. incanum*, *S. linnaeanum*, *S. insanum* (eggplant wild relative) and *S. agnewiorum*. Photograph credits: (*S. melongena*, *S. insanum*) X. Aubriot; (*S. incanum*, *S. agnew-iorum*) M. S. Vorontsova; (*S. linnaeanum*) S. Knapp.

that are native to Madagascar and the Mascarenes, but are not part of the monophyletic endemic Malagasy group; two are present in Africa and in Madagascar (*Solanum anguivi* and *S. richardii*), while the two others have a distribution range centered on tropical Asia, with *S. insanum* (the wild progenitor of the eggplant) present in Madagascar (Knapp et al., 2013; Aubriot et al., 2016; Ranil et al., 2017) and *S. violaceum* recorded from Mauritius (Vorontsova and Knapp, 2016). Hence, we treated Madagascar and the neighboring archipelagos as a distinct biogeographic area; the Malagasy region (region G).

The dispersal-extinction-cladogenesis (DEC) likelihood model implemented in Lagrange v. 20130526 (Ree and Smith, 2008) was used to investigate the biogeographic history of the Eggplant clade. The analysis was performed on the BEAST maximum clade credibility tree using the following set of parameters: (1) species ranges were coded using presence in the seven biogeographic regions; (2) maximum range size at nodes was constrained to 3; (3) areas not formed of adjacent regions were excluded from the analy-sis (adjacency matrix); (4) the dispersal constraints matrix was used to set the dispersal rates between the seven biogeographic regions with a value of 1 for adjacent regions (e.g., E and F), 0.5 between sea-separated regions (e.g., G) and 0.01 for nonadjacent regions (e.g., A and F). Ancestral area reconstruction for each node was plotted on the BEAST maximum clade credibility tree using pie charts (see Buerki et al., 2011 for details on the R script).

To assess the potential influence of fruit-eating migratory mammals in the dispersal of the African widespread Eggplant clade species (*Solanum campylacanthum*, *S. cerasiferum*, and *S. incanum*), a distribution map of these species together with two of their known large mammalian seed dispersers (African ele-phant, *Loxodonta africana* (Blumenbach) and impala, *Aepyceros melampus* (Lichentenstein); see Pringle et al., 2014; Kartzinel et al., 2015) was reconstructed using the R packages “maps” and “dismo”. Shapefiles of the current distributions of the African ele-phant and impala were downloaded from the International Union for Conservation of Nature Red List of Threatened Species website (IUCN SSC Antelope Specialist Group, 2016; Blanc, 2008). Using georeferenced herbarium specimens, we gathered data on 1150 populations of *S. campylacanthum*, *S. cerasiferum*, and *S. incanum*. A Venn diagram showing the overlap between the distributions of the widespread Eggplant clade species and the African elephant and impala was inferred using the R package “VennDiagram”.

RESULTS

Sanger reconstructions and species delimitation

Visual comparison of 50% majority-rule trees obtained from sin-gle marker matrices with BI and ML showed no strongly supported topological conflict (BS > 75% and PP > 0.9) (only the BI consen-sus trees shown; see Appendix S1a-c in the Supplemental Data with this article). All topologies are overall poorly resolved with the ex-ception of few strongly supported nodes; they only differed in the degree of resolution, with the ITS consensus trees exhibiting higher proportions of supported nodes than those inferred from *waxy* and *ndhF-rpL32*.

The combined Sanger data set yielded a total of 3121 base pairs (bp) of which 256 bp were variable. There were no supported con-flicts between the BI and ML 50% majority-rule trees, with the BI tree showing a much greater resolution than the ML one, al-beit often with poorly supported nodes. The combined topology (Appendix S1d) shows that the Eggplant clade, as previously de-fined (Vorontsova et al., 2013; Aubriot et al., 2016), is monophyl-etic (PP = 0.99) and it includes the Cape Verde endemic *Solanum rigidum* Lam. Within the Eggplant clade, two of the widespread species are resolved as monophyletic (*S. campylacanthum* and *S. cerasiferum*) and one as paraphyletic (*S. incanum*). The three acces-sions of *S. incanum* from the eastern part of the species distribution range (Kenya, Oman, and Pakistan) form a strongly supported lin-eage (BS = 99%; PP = 1) that excludes the accession from Burkina Faso; the latter is sister to the clade formed by the eggplant (*S. mel-ongena*) and its wild progenitor (*S. insanum*). *Solanum melongena* is part of a polytomy that includes all three accessions of *S. insanum*; this confirms previous phylogenetic results (Vorontsova et al., 2013; Aubriot et al., 2016). Affinities of the species of the Eggplant clade remain largely unresolved using these markers, with the exception of a group of species from southern Africa (*S. aureitomentosum* Bitter, *S. lichtensteinii* Willd., *S. linnaeanum*, *S. umtuma* Voronts. & S.Knapp) forming a strongly supported group (BS = 90%; PP = 1) that had been recognized in previous analyses (Vorontsova et al., 2013; Aubriot etal., 2016).

Plastome phylogenomic analyses

Plastome lengths ranged from 155,017 to 155,820 bp, with a mean coverage between 10.2× and 254.8×; see Table 1 and Appendix S2 for details on the plastomes. The plastome alignment was 159,227 bp in length, of which 3890 bp were variable. There were no supported conflicts between the BI topologies obtained with BEAST and MrBayes, and the RAxML 50% majority-rule tree (see Appendix S3). Phylogenetic trees are all fully resolved with the ex-ception of a polytomy in the MrBayes and RAxML 50% majority-rule trees. For the MrBayes topology, all other nodes have PP ≥ 0.99 (Appendix S3b). For the BEAST and RAxML trees, 13 out of 15 nodes within the Eggplant clade are strongly supported (i.e., BS ≥ 87% and PP ≥ 0.97; Fig. 1B and Appendix S3a,c). Plastome trees confirm results from the Sanger analyses, but provide much better resolution of relationships in the clade (Fig. 1B). The Eggplant clade (sensu Vorontsova et al., 2013; Aubriot et al., 2016) is recovered as monophyletic; it is sister (BS = 82%; PP = 0.94) to a lineage formed by two African (*Solanum polhillii* and *S. supinum*) and one tropical Asian species (*S. trilobatum*), partly confirming previous phyloge-netic results (Vorontsova et al., 2013; but see Aubriot et al., 2016 for an alternative position of *S. trilobatum*). The first-branching line-age, here called ‘Kilimanjaro clade’ (Fig. 1B), groups three species from the eastern African region (*S. agnewiorum*, *S. lanzae*, and *S. usambarense*). Affinities within this clade are still unresolved; in the BEAST analysis *S. agnewiorum* branches first, but this result is not well-supported (PP = 0.41). The next branching lineage is formed by the eggplant and its wild progenitor, *S. insanum*, the two accessions of the cultivated eggplant being grouped together (Fig. 1B). This lineage is sister to a monophyletic group that includes all remaining species of the Eggplant clade and is composed of two sister clades:
(1) the Southern African clade (Fig. 1B) with four species from southern Africa (*S. aureitomentosum* Bitter, *S. lichtensteinii*, *S. linnaeanum*, and *S. umtuma*), and (2) the Widespread clade (Fig. 1B), which includes three species with large distribution ranges (*S. campylacanthum*, *S. cerasiferum*, and *S. incanum*) and the Cape Verde islands endemic *S. rigidum*. Within the Southern African clade, *S. aureitomentosum* branches first, followed by S. *linnaeanum* that is sister to a moderately supported grouping formed by *S. lichtenstei-nii* and *S. umtuma* (BS = 83%; PP = 1). *Solanum campylacanthum* is the earliest diverging species of the Widespread clade; it is sister to a lineage in which *S. cerasiferum* branches first, followed by one of the two accessions of *S. incanum* (from Kenya) that is in turn sister to a grouping formed by the Cape Verdean *S. rigidum* and the second accession of *S. incanum* (from Burkina Faso).

Historical biogeography of the Eggplant clade and dispersal analysis

The most likely biogeographic scenario for the Eggplant clade is dis-played in Fig. 1B. The Eggplant clade originated in area B (Somalian and Middle East region) sometime during the early Pleistocene. The Kilimanjaro clade maintained this distribution range with the ex-ception of *Solanum usambarense*, which expanded its distribution range into area E (Zambesian region). There is a high level of bio-geographic uncertainty on the node separating the cultivated egg-plant plus its progenitor (*S. insanum*) from the rest of the species in the clade. This is caused by a disjunct distribution between these two clades (Fig. 1B). The most parsimonious scenario invokes two dispersal events leading to the most common ancestor of the whole clade therefore achieving a widespread distribution across regions BE and A (Tropical Asian region). The disjunct distribution of *S. insanum* in A and G (Malagasy region) most likely results from a long-distance dispersal event from A to G during the last million years. The majority of inferred dispersal events took place between the Southern African and Widespread clades. Southward and west-ward dispersal from B has been active in these two clades over the last 2 Myr, and F (Southern African region) was colonized during the past 1 Myr. In addition to having acted as a cradle, area B has been a pivotal region in shaping the current distribution of the Eggplant clade. The inferred dispersal scenarios are illustrated in Fig.1A.

The distribution of *Solanum campylacanthum*, *S. cerasiferum*, and *S. incanum* together with their large mammalian seed dispers-ers is displayed in Fig. 2A. The distribution of African elephants and impalas covers 53.7% (617 in total) of the populations of these three widespread species (Fig. 2B). The elephant range currently only covers 24.3% (150) of the three *Solanum* species populations, while the impala distribution covers 94.4% (583); most *Solanum* populations only overlap the impala range (467 populations) (Fig. 2B).

DISCUSSION

While studies focusing on the morpho-ecological and genetic diversity of the eggplant and its wild relatives are increasing (e.g., Gramazio et al., 2016; Kaushik et al., 2016; Ranil et al., 2017; Syfert et al., 2016; Acquadro et al., 2017), our study is the first designed to test phylogenetic and biogeographic scenarios in the Eggplant clade (sensu Aubriot et al., 2016). Our sampling for Sanger and genomic studies allowed us to test recent species delimitations and build a phylogeny for the Eggplant clade that is well resolved

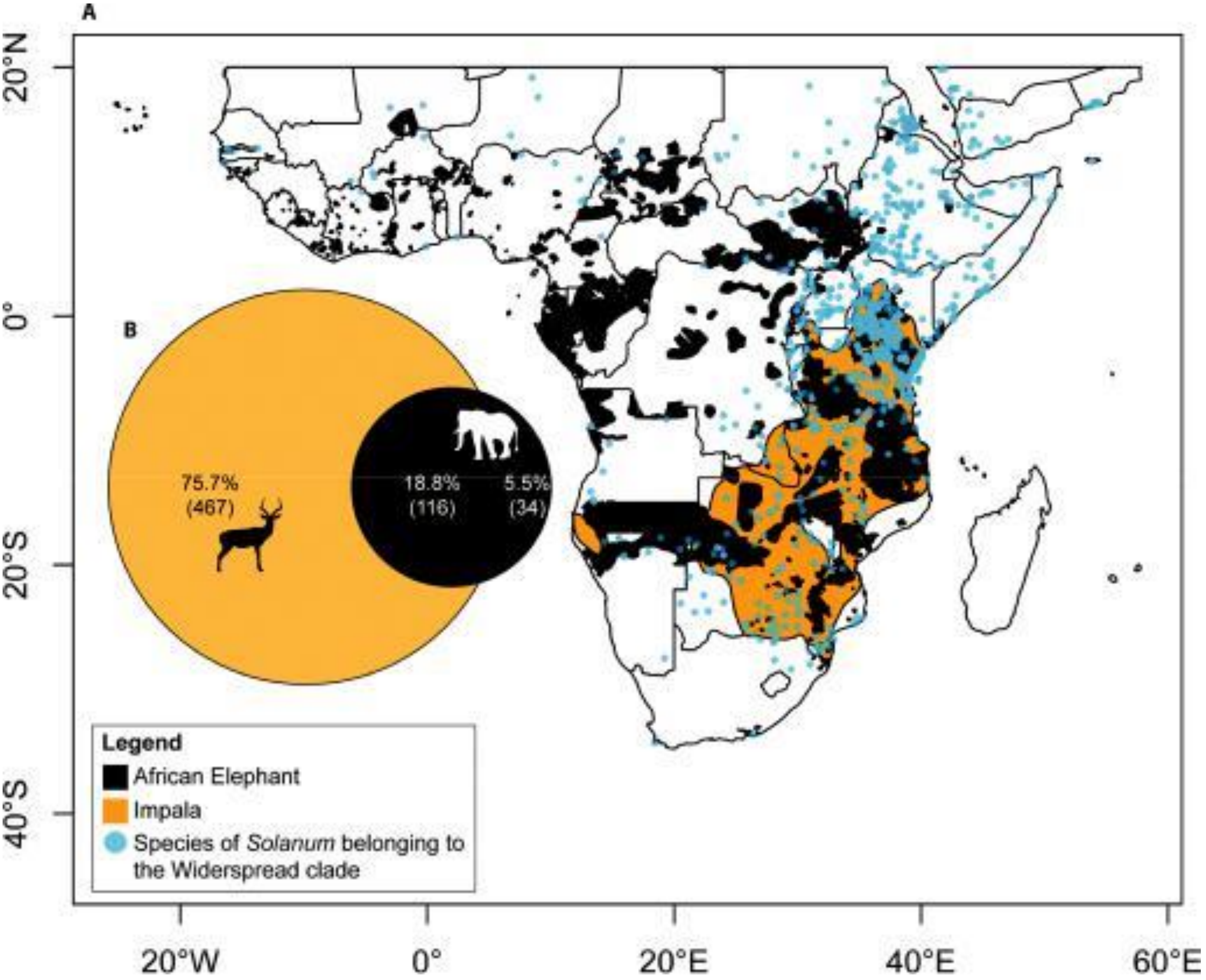

FIGURE 2. Overlap between the distribution of the widespread Eggplant clade species in Africa and two of their recorded seed dispersers, the African elephant and impala. (A) Distribution of species of *Solanum* belonging to the Widespread clade (with the exclusion of the Cape Verdean *S. rigidum*; see Fig. 1B for more details on this clade); current distributions of the African elephant and impala are also displayed on this map. (B) Venn diagram displaying the current overlap of distributions between populations of *S. campylacanthum* + *S. cerasiferum* + *S. incanum* and the African elephant and impala. Overall, 46.3% of the distribution of the three *Solanum* species is not currently overlapping with either the African elephant or the impala; only the overlapping populations are included in the Venn diagram.

internally. This study strongly supports *Solanum insanum*, the only wild Eggplant clade species currently distributed within the original center of domestication of *S. melongena* (Meyer et al., 2012), as its most likely progenitor (Fig. 1A, B). The Somalian and Middle East region is identified as the cradle of the Eggplant clade with recent dispersals to southern/western Africa and tropical Asia; distribu-tion data suggest that large savannah herbivores contributed to the geographical expansion of several of the species (Fig. 2A, B). The results do not, however, provide a definitive biogeographic scenario for the disjunct distribution between *S. insanum* + *S. melongena* and the rest of the species of the Eggplant clade that are all restricted to Africa and western Asia (see node 10 in Appendix S4a).

Phylogenetic relationships and species concepts in the Eggplant clade

Within the Eggplant clade, four species circumscriptions are par-ticularly challenging: *Solanum campylacanthum*, *S. cerasiferum*, *S. incanum*, and *S. insanum* (the three first species belong to the Widespread clade, the last is sister to *S. melongena*; Fig. 1B). Widespread across Africa and Asia, these species are also extremely polymorphic, with morphological characters that vary within and between populations (Knapp et al., 2013; Ranil et al., 2017). Olet and Bukenya-Ziraba (2001) suggested that *S. campylacanthum* (as "*S. incanum*") and *S. cerasiferum* belong to the same biological spe-cies; our phylogenetic results support the monophyly of these taxa and confirm they represent two distinct lineages. Sanger and full-plastome data unambiguously confirm *S. insanum* as the wild pro-genitor of the cultivated brinjal eggplant. While the two eggplant accessions are sister in the plastome topology, pairwise distances between the two plastomes of the eggplant and plastome of *S. insa-num* are identical (1×10^{-4}) and ~3–50 times smaller than the inter-specific pairwise distances of the data set (2.6×10^{-4} to 5.4×10^{-3}). Strong plastome identity between the brinjal eggplant and its wild progenitor is not surprising because they are known to interbreed relatively freely when in sympatry (e.g., Davidar et al., 2015; Mutegi et al., 2015). Our results also illustrate the fact that, while the cul-tivated eggplant and *S. insanum* present distinctive morphological features (Ranil et al., 2017), diagnostic molecular data are still lack-ing. *Solanum incanum* is recovered as paraphyletic, with the sample from the westernmost part of the distribution range (Burkina Faso) that does not group unambiguously with the other *S. incanum* ac-cessions. Genomic data resolve the Cape Verde endemic *S. rigidum* nested within *S. incanum*, strongly supported as sister to the sample from Burkina Faso. The plastid genome of *S. rigidum* is very similar to that of the two accessions of *S. incanum*; mean pairwise distance between these three plastomes is 3.5×10^{-4}. The phylogeny does not reflect morphological characters; *S. rigidum* is morphologically sim-ilar to *S. cerasiferum* and readily distinguishable from *S. incanum* by its deeply lobed leaves and sparser indumentum. This result can be interpreted in two ways. *Solanum rigidum* may be a recent morpho-geographical segregate of the continental *S. incanum*, both taxa having conserved their ancestral chloroplast genome. Internal tran-scribed spacer (ITS) data (Appendix S1a) resolve *S. rigidum* as sister to *S. macrocarpon* (gboma eggplant) and *S. dasyphyllum*, although this result is poorly supported (PP = 0.82); this would suggest a hy-brid origin with contributions from *S. incanum* (as the maternal parent of the cross) and arguably from *S. macrocarpon* or its wild progenitor, *S. dasyphyllum* (see also Knapp and Vorontsova, 2013). Obtaining broader geographic sampling of Eggplant clade species, particularly *S. rigidum* and accessions from western Africa, using data suited to this taxonomic depth, such as nuclear SSRs or SNPs (e.g., van Loo et al., 2015; Acquadro et al., 2017) may help better resolve circumscription of *S. incanum* and *S. rigidum*.

Our inferred plastome phylogeny provides much better resolu-tion at the interspecific level within the Eggplant clade than pre-vious phylogenetic investigations, which either lacked taxonomic representativeness (Meyer et al., 2012; Acquadro et al., 2017) or phy-logenetic resolution (Vorontsova et al., 2013; Aubriot et al., 2016) or both (Weese and Bohs, 2010). The topology we obtained, with the exception of two internal nodes (one node within the Kilimanjaro clade and one in the Southern African clade; Fig. 1B), is fully re-solved with strongly supported nodes. This plastome phylogeny con-firms several phylogenetic hypotheses while it contradicts others. Taxonomic composition and phylogenetic structure of the Eggplant clade as found by Aubriot et al. (2016) is confirmed, but with the addition of *Solanum rigidum* and much improved resolution and support. Three species that were found in polytomy at the base of the Eggplant clade (*S. agnewiorum*, *S. lanzae*, and *S. usambarense*) here are resolved as a basal Kilimanjaro clade; the Southern African clade and Widespread clade, whose compositions globally overlap clades that were previously poorly supported with Sanger data (see Aubriot et al., 2016; Vorontsova et al., 2013) are here resolved with highly supported nodes. Phenetic results obtained by Acquadro et al. (2017) with a data set of 12,859 nuclear SNPs, but with limited taxonomic sampling (six Eggplant clade spp.) are not fully congru-ent with our plastome topology; comparisons are difficult because of differences in sampling and methodological approaches.

Effect of savannahs and mammalian dispersers on biogeography of the Eggplant clade

Our biogeographic results suggest that the Eggplant clade origi-nated recently, during the last ~3 Myr, in the Somalian and Middle East region (Fig. 1A, B). Further dispersal and differentiation then occurred mostly within Africa; the progenitor of the eggplant, *Solanum insanum*, is the only wild species that is not present in con-tinental Africa, but that has a Malagasy-Southeast Asian distribu-tion. Dispersal to tropical Asia which gave rise to *S. insanum* is likely to have happened during the Early Pleistocene via the Middle East region; we suggest dispersal to Madagascar was a later event, pos-sibly linked to recent human migration (see Crowther et al., 2016). Our results challenge Lester and Hasan's (1991) hypothesis of an eastern African origin for the 'ancestors of *S. melongena*' with subse-quent spread to tropical Asia via a step-wise expansion through the Middle East. Instead, our biogeographic analysis shows that the ori-gin of the Eggplant clade lies in northeastern Africa and the Middle East, with spread both south (Widespread and Southern African clades) and east (*S. insanum* + *S. melongena*). Importantly, it shows that the tropical Asian lineage of *S. insanum* did not proceed from a step-wise expansion through the Middle East, but instead from an early dispersal from Africa, unrelated to the southwards spread of the Widespread and Southern African clades. This first attempt to provide a testable biogeographic framework for the Eggplant clade shows that its history consists in a wide dispersal across the African continent rather than an 'Out of Africa' phenomenon. Nesting of Asian species such as *S. insanum* in otherwise African lineages, sug-gests dispersal events in African lineages are not rare in Old World spiny solanums; *S. trilobatum* and *S. violaceum* also show this pat-tern in our plastome phylogeny (see also Aubriot et al., 2016).

Species of the Eggplant clade have a recent biogeographic his-tory and several species have large and complex distribution ranges, which makes it difficult to untangle biotic and abiotic factors that could explain evolutionary dynamics in the clade. Many spiny sola-nums are plants of disturbed areas, and most species of the Eggplant clade are weedy and occur in nitrogen-rich habitats of arid-areas or gaps in tropical forests, at a wide range of altitudes (Vorontsova and Knapp, 2016; Knapp and Vorontsova, 2013). Only the three spe-cies of the Kilimanjaro clade are habitat restricted; they are found in higher elevation (>1200 m) montane woodlands (Vorontsova and Knapp, 2016). These are also either hermaphroditic (*Solanum lan-zae*, *S. usambarense*) or only weakly andromonoecious (*S. agnewio-rum*) with small fruits, while strong andromonecy and larger fruits characterize all other species of the Eggplant clade (Vorontsova and Knapp, 2016; Vorontsova et al., 2013). Our data suggest that the combination of breeding system and ecology (see below) in this lin-eage have been important factors in the dramatic expansion of its distribution range over the last 2 Myr.

The increased geographical spread of members of the Eggplant clade coincides with profound habitat changes in Africa. Decrease in atmospheric CO_2 concentrations and aridification beginning in the late Miocene led to global increase in the biomass of plants using C_4 photosynthesis. This resulted in the extension of savannahs and fire-prone environments, first in northeastern Africa and then in South Africa (Cerling et al., 1997; Uno et al., 2016). Recent studies on hominin environments (Cerling et al., 2011; Uno et al., 2016) found that woody grassland savannahs became persistent in north-eastern Africa at least 6 Myr BP, with major increases in the abun-dance of savannah grazing mammals over this period (Uno et al., 2011). Open grassland and wooded grassland savannahs are typical habitats for members of both the Southern African and Widespread clades, and large savannah mammals such as African elephants and impalas have been recorded as important herbivores and seed dis-persers of Eggplant clade species (Pringle et al., 2014; Kartzinel et al., 2015). African elephants and impalas have native distributions that overlap those of members of the Eggplant clade, and particularly species of the Widespread clade (see Fig. 2A, B). Although African elephants were previously distributed across tropical Africa with their historical range fully encompassing the current distribution of the Eggplant clade species, they now only occupy ca. 20% of their historical range (see Ripple et al., 2015). Current impala distribution remains largely unchanged from their historical range (IUCN SSC Antelope Specialist Group, 2016); it appears that today the impala is likely to be more important for the seed dispersal of these wide-spread species. We suggest that the extensive geographical expan-sion over the last 2–3 Myr and the impressive ranges of some of the extant species (e.g., *Solanum campylacanthum*, *S. cerasiferum*, *S. incanum*) are related to the dispersal of seeds by these large, mi-gratory mammalian herbivores. Recent and ongoing contraction of ranges of these large mammalian seed dispersers could ultimately contribute to isolation of populations previously more intermixed, leading to differentiation and speciation. Human use of these spe-cies could have also contributed to their recent wide dispersal and spread; many species of the clade are used medicinally and for tan-ning hides (Bitter, 1923; Ranil et al., 2017; Vorontsova and Knapp, 2016). Fragments of *Solanum* seeds have been recorded from a 780,000-year-old hominin site in Israel (Melamed et al., 2016), and archaeological remains may provide additional information about early interactions between hominins and Eggplant clade species.

Disentangling the relationship between complex changes in ecology and differentiation in the Eggplant clade, and in spiny sola-nums more broadly, is challenging, especially in the light of recent profound environmental change; new types of data will be required to understand the evolutionary success of this widely distributed and very diverse group.

Eggplant domestication, still a biogeographic puzzle

Our understanding of the evolutionary history and biogeography of the eggplant and its wild relatives is now better understood fol-lowing the analyses conducted here using whole plastome and nu-clear sequence data. The two nuclear regions we analyzed here do not contradict plastome data, but nuclear genomic data for all the species of the Eggplant clade will be necessary to assess phyloge-netic discordance and detect historical gene flow. The combination of plastome and denser nuclear genomic data would also open the way to in-depth investigation of the genomic basis of environment-specific and lineage-specific adaptations in eggplant CWR (e.g., Pease et al., 2016). This will ultimately provide plant breeders and physiologists with an estimate of the potential for selection from standing variation across eggplant CWR, potentially revolutioniz-ing genetic improvement of eggplants.

Our study also shows that even with sampling across the full diversity of the Eggplant clade and a robust phylogenomic framework, our understanding of the biogeographic history of the Eggplant clade can still be improved; in particular the dispersal of *Solanum insanum* lineages to tropical Asia and emergence and domestication of the brinjal eggplant. Meyer et al. (2012) suggested that *S. melon-gena* was domesticated several times from *S. insanum* (as *S. inca-num* in Meyer et al., 2012); use of *S. insanum* as a medicine (Ranil et al., 2017) and extensive movement of peoples across southeastern Asia support this idea. Combining sampling for *S. insanum* through its whole geographical range with the use of recently developed targeted high-throughput sequencing (Kadlec et al., 2017) will provide a promising approach for reconstructing reliable scenarios of evolu-tion of brinjal eggplant from its wild progenitor, *S. insanum*.

ACKNOWLEDGEMENTS

We thank the curators of A, BM, K, LISC, and MH for permission to sample DNA from specimens held in their collections; Pr. H. Balslev (Aarhus University), the Bangkok Forest Herbarium (BKF), Pr. P. Chantaranothai, and the Botanical team of the Applied Taxonomic Research Center of Khon Kaen University for field assistance in Thailand; Dr. D. Narasimhan (Madras Christian College), Dr. V. Sampath Kumar, Dr. G. V. S Murthy, Dr. P. Lakshminarasimhan, Dr. A. Benniamin, Dr. G. Gnanasekaran (Botanical Survey of India), P. Kale, and R. Singh (BLAT) for their invaluable assistance during visits of Indian herbaria. We thank M. S. Vorontsova, M.-C. Daunay, M. C. Duarte, and J. Prohens for their comments that significantly improved the manuscript; we thank the NHM sequencing facility team, particularly A. Briscoe, P. Dyal, and K. Hopkins, for assistance with molecular work. This work was funded in part by the Natural History Museum's Natural Resources and Hazards Initiative and by the Natural History Museum's Departmental Investment Fund. We thank the three anonymous referees and the associate editor for their comments that significantly improved the manuscript.

DATA ACCESSIBILITY

The GenBank accession numbers for all the sequences used in this study are provided in Table 1 and Appendix 1. All original data and scripts used are available on the Natural History Museum Data Portal as a single data set composed of four files (https://doi.org/10.5519/0068715). File 1 is the static data set of geographical information for the collections of the three African widespread Eggplant clade species (*Solanum campylacanthum*, *S. cerasiferum*, and *S. incanum*) used for the range overlap analysis presented in Fig. 2. Files 2 and 3 are the alignments used for the 3-locus and full-plastome phylogenetic analyses, and File 4 includes the Python script used in Lagrange for the biogeographic analysis.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

Acquadro, A., L. Barchi, P. Gramazio, E. Portis, S. Vilanova, C. Comino, M. Plazas, et al. 2017. Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS One* 7: e0180774.

Amiryousefi, A., J. Hyvönen, and P. Poczai. 2018. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): plastid genome structure evolution in Solanaceae. *PLoS One* 13: e0196069.

Aubriot, X., P. Singh, and S. Knapp. 2016. Tropical Asian species show the Old World clade of “spiny solanums” (the Leptostemonum Clade: Solanaceae) is not monophyletic. *Botanical Journal of the Linnean Society* 180: 1–27.

Bebber, D. P., T. Holmes, and S. J. Gurr. 2014. The global spread of crop pests and pathogens. *Global Ecology and Biogeography* 23: 1398–1407.

Bitter, G. 1923. Solana Africana Part IV. *Repertorium specierum novarum regni vegetabilis* 16: 1–320. Dahlem bei Berlin, Berlin, Germany.

Blanc, J. 2008. *Loxodonta africana*. The IUCN Red List of Threatened Species 2008: e.T12392A3339343. [online] Website https://doi.org/10.2305/iucn.uk.2008.rlts.t12392a3339343.en. [accessed 9 February 2018].

Bohs, L. 2005. Major clades in *Solanum* based on *ndhF* sequences. *In* R. C. Keating, V. C. Hollowell, and T. B. Croat [eds], A festschrift for William G. D’Arcy: the legacy of a taxonomist, 27–49. Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Buerki, S., D. S. Devey, M. W. Callmander, P. B. Phillipson, and F. Forest. 2013. Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society* 171: 304–329.

Buerki, S., F. Forest, N. Alvarez, J. A. A. Nylander, N. Arrigo, and I. Sanmartín. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography* 38: 531–550.

Buerki, S., S. Jose, S. R. Yadav, P. Goldblatt, J. C. Manning, and F. Forest. 2012. Contrasting biogeographic and diversification patterns in two Mediterranean-type ecosystems. *PLoS One* 7: e39377.

Cattivelli, L., F. Rizza, F. W. Badeck, E. Mazzucotelli, A. M. Mastrangelo, E. Francia, C. Marè, et al. 2008. Drought tolerance improvement in crop plants: an inte-grated view from breeding to genomics. *Field Crops Research* 105: 1–14.

Cerling, T. E., J. M. Harris, B. J. MacFadden, M. G. Leakey, J. Quade, V. Eisenmann, and J. R. Ehleringer. 1997. Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389: 153–158.

Cerling, T. E., J. G. Wynn, S. A. Andanje, M. I. Bird, D. K. Korir, N. E. Levin, W. Mace, et al. 2011. Woody cover and hominin environments in the past 6-million years. *Nature* 476: 51–56.

Couvreur, T. L. P., L. W. Chatrou, M. S. M. Sosef, and J. E. Richardson. 2008. Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology* 6: 54.

Crowther, A., L. Lucas, R. Helm, M. Horton, C. Shipton, H. T. Wright, S. Walshaw, et al. 2016. Ancient crops provide first archaeological signature of the west-ward Austronesian expansion. *Proceedings of the National Academy of Sciences* 113: 6635–6640.

Daunay, M.-C., and P. Hazra. 2012. Eggplant. *In* P. Hazra, and K. V. Peter [eds], Handbook of vegetables, 257–322. Studium Press, Houston, Texas, USA. Daunay, M.-C., and J. Janick. 2007. History and iconography of eggplant. *Chronica Horticulturae* 47: 16–22.

Daunay, M.-C., R. N. Lester, and H. Laterrot. 1991. The use of wild species for the genetic improvement of brinjal (*Solanum melongena* L.) and tomato (*Lycopersicon esculentum* Miller). *In* J. G. Hawkes, R. N. Lester, M. Nee, and N. Estrada [eds], Solanaceae III: taxonomy, chemistry, evolution, 389–412. Royal Botanic Gardens Kew, Kew, UK.

Davidar, P., A. A. Snow, M. Rajkumar, R. Pasquet, M.-C. Daunay, and E. Mutegi. 2015. The potential for crop to wild hybridization in eggplant (*Solanum melon-gena;* Solanaceae) in southern India. *American Journal of Botany* 102: 128–138.

Dempewolf, H., R. J. Eastwood, L. Guarino, C. K. Khoury, J. V. Müller, and J. Toll. 2014. Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. *Agroecology and Sustainable Food Systems* 38: 369–377.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phy-logenetics and dating with confidence. *PLoS Biology* 4: e88.

Drummond, A. J., M. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phyloge-netics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Evans, M., X. Aubriot, D. Hearn, M. Lanciaux, S. Lavergne, C. Cruaud, P. P. Lowry II PP, and T. Haevermans. 2014. The evolution of succulence: insights from a remarkable radiation in Madagascar. *Systematic Biology* 63: 698–711.

FAO. 2017. The statistical database (FAOSTAT). [online] Website http://www.fao.org/faostat/en/. [accessed 9 February 2018].

Frodin, D. G. 2004. History and concepts of big plant genera. *Taxon* 53: 753–776.

Gramazio, P., J. Blanca, P. Ziarsolo, F. J. Herraiz, M. Plazas, J. Prohens, and S. Vilanova. 2016. Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common egg-plant (*Solanum melongena*) with interest for breeding. *BMC Genomics* 17: 300.

Gramazio, P., J. Prohens, M. Plazas, I. Andjar, F. J. Herraiz, E. Castillo, S. Knapp, et al. 2014. Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. *BMC Plant Biology* 14: 350.

Holt, B. G., J.-P. Lessard, M. K. Borregaard, S. A. Fritz, M. B. Araujo, D. Dimitrov, P.-H. Fabre, et al. 2013. An update of Wallace’s zoogeographic regions of the world. *Science* 341: 343–343.

Hopkins, J., and N. Maxted. 2010. Crop wild relatives: plant conservation for food security. Natural England, Peterborough, UK.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Hurtado, M., S. Vilanova, M. Plazas, P. Gramazio, H. H. Fonseka, R. Fonseka, and J. Prohens. 2012. Diversity and relationships of eggplants from three geographically distant secondary centers of diversity. *PLoS One* 7: e41748.

IUCN SSC Antelope Specialist Group. 2016. *Aepyceros melampus*. The IUCN Red List of Threatened Species 2016: e.T550A50180828. [online] Website https://doi.org/10.2305/iucn.uk.2016-2.rlts.t550a50180828.en. [accessed 9 February 2018].

Kadlec, M., D. U. Bellstedt, N. C. Le Maitre, and M. D. Pirie. 2017. Targeted NGS for species level phylogenomics: ‘made to measure’ or ‘one size fits all’? *PeerJ* 5: e3569.

Kartzinel, T. R., P. A. Chen, T. C. Coverdale, D. L. Erickson, W. J. Kress, M. L. Kuzmina, D. I. Rubenstein, et al. 2015. DNA metabarcoding illuminates dietary niche partitioning be African large herbivores. *Proceedings of the National Academy of Sciences* 112: 8019–8024.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kaushik,P.,J.Prohens,S.Vilanova,P.Gramazio,andM.Plazas.2016.Phenotyping of eggplant wild relatives and interspecific hybrids with conventional and phenomics descriptors provides insight for their potential utilization in breeding. *Frontiers in Plant Science* 7: 677.

Kearse,M.,R.Moir,A.Wilson,S.Stones-Havas,M.Cheung,S.Sturrock,S.Buxton, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Knapp, S., and M. S. Vorontsova. 2013. From introduced American weed to Cape Verde Islands endemic – the case of *Solanum rigidum* Lam. (*Solanum*, subgenus *Leptostemonum*). *PhytoKeys* 25: 35–46.

Knapp, S., M. S. Vorontsova, and J. Prohens. 2013. Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): New understanding of species names in a complex group. *PLoS One* 8: e57039.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.

Lester, R. N., and S. M. Z. Hasan. 1991. Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *Solanum incanum*, in Africa and Asia. *In* J. G. Hawkes, R. N. Lester, M. Nee, and N. Estrada [eds], Solanaceae III: taxonomy, chemistry and evolution, 369–387. Royal Botanic Gardens Kew, Kew, UK.

Linder, H. P., H. M. de Klerk, J. Born, N. D. Burgess, J. Fjeldså, and C. Rahbek. 2012. The partitioning of Africa: Statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeography*. 39: 1189–1205.

Melamed, Y., M. E. Kislev, E. Geffen, S. Lev-Yadun, and N. Goren-Inbar. 2016. The plant component of an Acheulian diet at Gesher Benot Ya'aqov, Israel. *Proceedings of the National Academy of Sciences* 113: 14674–14679.

Meyer, R. S., K. G. Karol, D. P. Little, M. H. Nee, and A. Litt. 2012. Phylogeographic relationships among Asian eggplants and new perspectives on eggplant do-mestication. *Molecular Phylogenetics and Evolution* 63: 685–701.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 1–8, New Orleans, Louisiana, USA.

Muñoz-Falcón, J. E., J. Prohens, S. Vilanova, and F. Nuez. 2009. Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. *Annals of Applied Biology* 154: 453–465.

Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences* 104: 19363–19368.

Mutegi, E., A. A. Snow, M. Rajkumar, R. Pasquet, H. Ponniah, M.-C. Daunay, and P. Davidar. 2015. Genetic diversity and population structure of wild/weedy eggplant (*Solanum insanum*, Solanaceae) in southern India: Implication for conservation. *American Journal of Botany* 102: 139–147.

Nylander, J. A. A. 2004. MrModeltest v2. Computer program and documentation distributed by the author, website: https://github.com/nylander/MrModeltest2. [accessed 9 February 2018].

Olet, E. A., and R. Bukenya-Ziraba. 2001. Variation within the *Solanum incanum* complex in Uganda and its relationship with *Solanum cerasiferum*. *In* R. G. van den Berg, G. W. M. Barendse, G. M. van der Weerden, and C. Mariani [eds], Solanaceae V: Advances in Taxonomy and Utilization, 97–108. Nijmegen University Press, Nijmegen, Netherlands.

Pautasso, M., T. F. Döring, M. Garbelotto, L. Pellis, and M. J. Jeger. 2012. Impacts of climate change on plant diseases—opinions and trends. *European Journal of Plant Pathology* 133: 295–313.

Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biology* 14: e1002379.

Peng, Y., H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly un-even depth. *Bioinformatics* 28: 1420–1408.

Pringle, R. M., J. R. Goheen, T. M. Palmer, G. K. Charles, E. DeFranco, R. Hohbein, A. T. Ford, and C. E. Tarnita. 2014. Low functional redundancy among mammalian browsers in regulating an encroaching shrub (*Solanum campylacanthum*) in African savannah. *Proceedings of the Royal Society B, Biological Sciences* 281: 20140390.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6. Computer program and documentation distributed by the author, website: http://beast.community/tracer. [accessed 9 February 2018].

Ranil, R. H. G., J. Prohens, X. Aubriot, H. M. L. Niran, M. Plazas, R. M. Fonseka, S. Vilanova, et al. 2017. *Solanum insanum* L. (subgenus *Leptostemonum* Bitter, Solanaceae), the neglected wild progenitor of eggplant (*S. melongena* L): a review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. *Genetic Resources and Crop Evolution* 64: 1707–1722. [Published electronically in 2016.]

Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

Ripple, W. J., T. M. Newsome, C. Wolf, R. Dirzo, K. T. Everatt, M. Galetti, M. W. Hayward, et al. 2015. Collapse of the world's largest herbivores. *Science Advances* 1: e1400103.

Särkinen, T., L. Bohs, R. G. Olmstead, and S. Knapp. 2013. A phylogenetic frame-work for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* 13: 214.

Shi, C., N. Hu, H. Huang, J. Gao, Y.-J. Zhao, and L.-Z. Gao. 2012. An improved chloroplast DNA extraction procedure for whole plastid genome sequenc-ing. *PLoS One* 7: e31468.

Solanaceae Source Database. 2017. Solanaceae Source: a global taxonomic resource for the nightshades [online]. Website http://www.solanaceaesource.org/ [accessed 9 February 2018].

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stamatakis, A. 2016. The RAxML v8.2.X Manual. [online] Website https://sco.h-its.org/exelixis/resource/download/NewManual.pdf. [accessed 9 February 2018].

Stern, S. R., M. de F. Agra, and L. Bohs. 2011. Molecular delimitation of clades within New World species of the "spiny solanums" (*Solanum* subgenus *Leptostemonum*). *Taxon* 60: 1429–1441.

Syfert, M. M., N. P. Castañeda-Álvarez, C. K. Khoury, T. Särkinen, C. C. Sosa, H. A. Achicanoy, V. Bernau, et al. 2016. Crop wild relatives of the brinjal eggplant (*Solanum melongena*; Solanaceae): poorly represented in genebanks and many species at risk of extinction. *American Journal of Botany* 103: 635–651.

Thiers, B. 2017. Index Herbariorum: A global directory of public herbaria and associated staff. [online] Website http://sweetgum.nybg.org/science/ih/. [ac-cessed 9 February 2018].

Tonti-Filippini, J., P. G. Nevill, K. Dixon, and I. Small. 2017. What can we do with 1000 plastid genomes? *Plant Journal* 90: 808–818.

Uno, K. T., T. E. Cerling, J. M. Harris, Y. Kunimatsu, M. G. Leakey, M. Nakatsukasa, and H. Nakaya. 2011. Late Miocene to Pliocene carbon isotope record of differential diet change among East African herbivores. *Proceedings of the National Academy of Sciences* 108: 6509–6514.

Uno, K. T., P. J. Polissar, K. E. Jackson, and P. B. deMenocal. 2016. Neogene bi-omarker record of vegetation change in eastern Africa. *Proceedings of the National Academy of Sciences* 113: 6355–6363.

van Loo, M., W. Hintsteiner, E. Pötzelsberger, S. Schüler, and H. Hasenauer. 2015. Intervarietal and intravarietal genetic structure in Douglas-fir: nuclear SSRs bring novel insights into past population demographic processes, phylogeography, and intervarietal hybridization. *Ecology and Evolution* 5: 1802–1817.

Vorontsova, M. S., and S. Knapp. 2016. A revision of the spiny solanums, *Solanum* subgenus *Leptostemonum* (Solanaceae) in Africa and Madagascar. *Systematic Botany Monographs* 99: 1–432.

Vorontsova, M. S., S. Stern, S. Bohs, and S. Knapp. 2013. African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. *Botanical Journal of the Linnean Society* 173: 176–193.

Wang, J.-X., T.-G. Gao, and S. Knapp. 2008. Ancient Chinese literature reveals pathways of eggplant domestication. *Annals of Botany* 102: 891–897.

Weese, T. L., and L. Bohs. 2010. Eggplant origins: out of Africa, into the Orient. *Taxon* 59: 49–56.

White, F. 1983. The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. UNESCO, Paris, France.

White, F. 1993. The AETFAT chorological classification of Africa: history, meth-ods and applications. *Bulletin du Jardin Botanique National de Belgique* 62: 225–281.

Williams, A. V., J. T. Miller, I. Small, P. G. Nevill, and L. M. Boykin. 2016. Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Acacia*. *Molecular Phylogenetics and Evolution* 96: 1–8.

Wysocki, W. P., L. G. Clark, L. Attigala, E. Ruiz-Sanchez, and M. R. Duvall. 2015. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylog-enomic analysis. *BMC Evolutionary Biology* 15: 50.

Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using Bruijin graphs. *Genome Research* 18: 821–829.

APPENDIX 1. Summary of species, country, voucher (collector and number), herbarium where vouchers are housed (acronyms follow Index Herbariorum (Thiers, 2017)) and GenBank accession numbers for taxa used in the preliminary Sanger study, provided in the following order: ITS, *waxy* and *ndhF-rpL32*. Dashed lines indicate that the region was not sampled for this accession. Newly generated sequences are indicated with an asterisk following the accession number.

Solanum aethiopicum L., Tanzania, *Vorontsova et al. 156* (BM), MH283730*, –, MH283767*. *S. aethiopicum* L., *Olmstead S-74* (WTU), –, AY996378, –. *S. agnewiorum* Voronts., Kenya, *Vorontsova et al. 195* (BM), KU719948, KU719870, KU720049. *S. anguivi* Lam., Madagascar, *Vorontsova et al. 1302* (K), MH283726*, –, MH283766*. *S. anguivi* Lam., *Bohs 3266* (UT), –, HQ721963, –. *S. aureitomentosum* Bitter, Zimbabwe, *Chase 5271* (BM), KU719949, KU719871, MH283770*. *S. campylacanthum* Hochst. ex A.Rich., Tanzania, *Vorontsova et al. 158* (BM), MH283727*, MH283749*, KU719997. *S. campylacanthum* Hochst. ex A.Rich., Ethiopia, Gambela, *Friis 8107* (K), MH283731*, MH283750*, MH283768*. *S. campylacanthum* Hochst. ex A.Rich., Ethiopia, Wukro, *Friis et al. 10427* (K), MH283735*, MH283754*, MH283773*. *S. campylacanthum* Hochst. ex A.Rich., Ethiopia, Dubluk, *Friis 11156* (K), MH283736*, MH283755*, MH283774*. *S. campylacanthum* Hochst. ex A.Rich., South Africa, *Westfall 1661* (K), MH283738*, MH283757*, MH283776*. *S. campylacanthum* Hochst. ex A.Rich., Namibia, *Kolberg & Kubirske 600* (K), MH283739*, MH283758*, MH283777*. *S. cerasiferum* Dunal, Ethiopia, Pawi, *Friis et al. 7733* (K), MH283732*, MH283751*, MH283769*. *S. cerasiferum* Dunal, Ethiopia, Mengi, *Friis et al. 12670* (K), MH283737*, MH283756*, MH283775*. *S. cerasiferum* Dunal, Central African Republic, *Fay 2715* (K), MH283740*, MH283759*, MH283778*. *S. dasyphyllum* Schumach. & Thonn., Tanzania, *Vorontsova et al. 151* (BM), MH283728*, –, KU720000. *S. dasyphyllum* Schumach. & Thonn., *Cipollini 7* (UT), –, AY996406, –. *S. glabratum* Dunal, South Arabia, *Miyazaki 606I28* (A), KU719913, KU719834, KU719994. *S. incanum* L., Kenya, *Vorontsova et al. 203* (BM), MH283734*, MH283753*, MH283772*. *S. incanum* L., Burkina Faso, *Sanou & Traore BUR48* (K), MH283741*, MH283760*, MH283779*. *S. incanum* L., Oman, *Mandaville 6523* (BM), MH283742*, MH283761*, MH283780*. *S. incanum* L., Pakistan, *Popov 30* (BM), MH283743*, MH283762*, MH283781*. *S. insanum* L., Madagascar, *Vorontsova et al. 644* (BM), KU719914, KU719835, KU720003. *S. insanum* L., *Wang 2052* (PE), KU719884, KU719813, KU719964. *S. insanum* L., India, *Sampath Kumar et al. 941* (MH), KU719940, KU719862, KU720041. *S. lanzae* J.-P.Lebrun & Stork, Kenya, *Vorontsova et al. 783* (BM), KU719921, KU719842, KU720015. *S. lichtensteinii* Willd., Namibia, *Crawford et al. 139* (K), KU719947, KU719869, KU720048. *S. linnaeanum* Hepper & P.-M.L. Jaeger, Morocco, *Jury et al. 13209* (BM), KU719950, KU719872, KU720050. *S. macrocarpon* L., Tanzania, *Tepe et al. 2770* (BM), MH283729*, –, KU720004. *S. macrocarpon* L., *Olmstead S-88* (WTU), –, AY996436, –. *S. melongena* L., Thailand, *Meeboonya et al. RM294* (BM), MH283747*, –, MH283785*. *S. melongena* L., *Olmstead S-91* (WTU), –, AY562959, –. *S. nigriviolaceum* Bitter, Kenya, *Vorontsova et al. 57* (BM), HQ721868, HQ721993, –. *S. polhillii* Voronts., Kenya, *Vorontsova et al. 15* (BM), MH283733*, MH283752*, MH283771*. *S. richardii* Dunal, Madagascar, *Vorontsova et al. 633* (BM), KU719915, KU719836, KU720006. *S. rigidum* Lam., Cape Verde, *Malato-Beliz 141* (LISC), MH283746*, –, MH283784*. *S. supinum* Dunal, Botswana, *Kabelo 170* (K), MH283745*, MH283764*, MH283783*. *S. trilobatum* L., Thailand, *Meeboonya et al. RM243* (BM), MH283725*, MH283748*, MH283765*. *S. umtuma* Voronts. & S.Knapp, South Africa, *Nevhutalu 921* (K), MH283744*, MH283763*, MH283782*. *S. usambarense* Bitter & Dammer, Tanzania, *Vorontsova et al. 166* (BM), KU719923, KU719844, KU720023. *S. violaceum* Ortega, India, *Sampath Kumar et al. 945e* (MH), KU719941, KU719863, KU720042.