

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

Xavier Aubriot, Sandra Knapp, Mindy M Syfert, Péter Poczai, Sven Buerki

▶ To cite this version:

Xavier Aubriot, Sandra Knapp, Mindy M Syfert, Péter Poczai, Sven Buerki. Shedding new light on the origin and spread of the brinjal eggplant (Solanum melongena L.) and its wild relatives. American Journal of Botany, 2018, 105 (7), pp.1175-1187. 10.1002/ajb2.1133 . hal-01862320

HAL Id: hal-01862320 https://univ-rennes.hal.science/hal-01862320

Submitted on 12 Jul2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Title	Shedding new light on the origin and spread of the brinjal eggplant (Solanum melongena L.) and its wild relatives		
Authors	Aubriot, X; Knapp, S; Syfert, MM; Poczai, P; Buerki, S		
Date Submitted	2018-08-18		

1Shedding new light on the origin and spread of the brinjal 2eggplant (*Solanum melongena* L.; Solanaceae) and its wild 3relatives

4

5Xavier Aubriot^{1, 2}, Sandra Knapp¹, Mindy Syfert¹, Péter Poczai³, Sven Buerki^{1, 4}

71 Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, 8England, UK.

92 Unité Mixte de Recherche 6553 Écosystèmes, Biodiversité, Évolution (ECOBIO), Observatoire des 10Sciences de l'Univers de Rennes, Centre National de la Recherche Scientifique, Université de Rennes 111, Rennes CEDEX, France.

123 Botany Unit, Finnish Museum of Natural History, University of Helsinki, PO Box 7, Helsinki FI-1300014, Finland.

144 Department of Biological Sciences, Boise State University, 1910 University Drive, Boise, Idaho 1583725, U.S.A.

16

17Authors for correspondence: Xavier Aubriot, Tel: +33 623 744 393, Email: xavier.aubriot@univ-18rennes1.fr; Sven Buerki, Tel: +1 208 426 3262, Email: svenbuerki@boisestate.edu

19

20ABSTRACT

21**PREMISE OF THE STUDY:** While brinjal eggplant (*Solanum melongena* L.) is the second most 22important solanaceaous vegetable crop, we lack firm knowledge of its evolutionary relationships. This 23in turn limits efficient use of crop wild relatives in eggplant improvement. Here, we examine the 24hypothesis of linear step-wise expansion of the eggplant group from Africa to Asia.

25**METHODS:** We use museum collections to generate nuclear and full-plastome data for all species of 26the eggplant clade. We combine a phylogenomic approach with distribution data to infer a 27biogeographic scenario for the clade.

28**KEY RESULTS:** The eggplant clade has Pleistocene origins in northern Africa. Dispersions to 29tropical Asia gave rise to *Solanum insanum*, the wild progenitor of the eggplant, and to Africa distinct 30lineages of widespread and southern-African species. Results suggest that spread of species to 31southern Africa is recent and was likely facilitated by large mammalian herbivores feeding on 32*Solanum* fruits (African elephant and impala).

33**CONCLUSIONS:** Rather than a linear 'Out Of Africa' sequence, our results are more consistent with 34an initial dispersion event into Asia, and subsequent wide dispersion and differentiation across Africa

35driven by large mammalian herbivores. Our evolutionary results will impact future work on eggplant 36domestication and use of wild relatives in breeding of this increasingly important solanaceous crop. 37

KEY WORDS Eggplant (aubergine), crop wild relatives (CWRs), phylogenomics, chloroplast 39genome, biogeography, molecular dating, dispersion, mammalian herbivores

40INTRODUCTION

41 Crop wild relatives (CWRs) are likely to play a significant role in securing 21st century food 42security (Dempewolf et al., 2014), due to their potential for use in breeding to withstand adverse 43impacts of climate change, increasing scarcity of nutrients, water and other inputs, and new pests and 44diseases (Hopkins and Maxted, 2010). Understanding the phylogenetic relationships, geographic 45origins and spread of CWRs is critical for improving their use in future plant breeding in the face of 46environmental and climate change.

The brinjal eggplant, *Solanum melongena* L. (also known as the aubergine; Solanaceae) is the 48second most important economically important solanaceaous vegetable crop after tomato (FAO, 492017). The brinjal eggplant has a long history of cultivation (see Daunay and Janick, 2007; Wang et 50al., 2008), and is thought to have been domesticated in Asia several times (Daunay and Hazra, 2012; 51Meyer et al., 2012). Increasing genetic diversity in the cultivated crop is a major target in brinjal 52eggplant breeding (Muñoz-Falcón et al., 2009). Improvements in yield, fruit and postharvest quality 53(Vilanova et al., 2010), resistance to plant pathogens (Pautasso et al., 2012; Bebber et al., 2014), and 54introduction of abiotic stress traits such as tolerance to increased drought, temperatures, and extreme 55precipitation are all important, especially in the face of climate change model predictions of higher 56temperatures and extreme weather patterns worldwide (Cattivelli et al., 2008). Crop wild relatives are 57an increasingly important source for traits and genes for brinjal eggplant improvement (Daunay et al., 581991; Muñóz-Falcón et al., 2009; Hurtado et al., 2012; Kaushik et al., 2016), and knowledge about 59their distribution, traits and threat status has been the focus of recent studies (Syfert et al., 2016; 60Kinsell et al., unpublished).

Solanum L., with *c*. 1,400 species, is one of the largest genera of flowering plants (Frodin, 622004) and has been the subject of much recent taxonomic and phylogenetic work (e.g., see references 63in Särkinen et al., 2013; Vorontsova and Knapp, 2016). The species-level taxonomy of *S. melongena* 64and its wild relatives has long been recognised as complex (Lester and Hasan, 1991; Daunay and 65Hazra, 2012), but recent revision and clarification of taxonomic limits in the group (Knapp et al., 662013; Vorontsova and Knapp, 2016) has provided a set of species identities for use in further analyses. 67Previous analyses recognised only two species in the Eggplant clade (e.g., Lester and Hasan, 1991; 68Weese and Bohs, 2010), but further work with African solanums suggested the clade comprised ten 69species characterized in part by their andromonoecious breeding system (see Table 1 in Knapp et al., 702013). Even though the number of taxa recognised increased, several extremely widespread species 71(e.g., *S. campylacanthum* A.Rich. across eastern Africa) include much geographical variation, 72suggesting a role for ecological structuring in these taxa (Ranil et al., 2016; Vorontsova and Knapp, 732016).

74 Solanum melongena and its wild relatives are members of the largest clade of the genus, the

75Leptostemonum Clade (the 'spiny' solanums; Bohs, 2005; Weese and Bohs, 2010; Särkinen et al., 762013), and within that, a monophyletic group comprising almost all Old World spiny solanums (Stern 77et al., 2011; Vorontsova et al., 2013; Aubriot et al., 2016). The Eggplant clade has been recognised as 78monophyletic in previous analyses (e.g., Stern et al., 2011; Weese and Bohs, 2010), but sampling of 79Old World taxa was limited. Using the same set of genebank accessions Weese and Bohs (2010) tested 80Lester's (Lester and Hasan, 1991) evolutionary scenario of brinjal eggplant evolution, but showed that 81the weedy species *S. linnaeanum* Hepper & P.M.-L.Jaeger also belonged to the group. Their analysis 82(Weese and Bohs, 2010) supported the scenario of an African origin for the group with step-wise 83expansion to tropical Asia and migration back to the Middle East as feral forms. Vorontsova et al. 84(2013) expanded the dataset by including many African taxa and still recovered a monophyletic 85Eggplant clade, but with little internal resolution. Their results placed the narrow Kenvan endemic *S*. 86*aqnewiorum* Voronts. within the monophyletic Eggplant clade, a surprising result given its small fruit 87 and hermaphroditic flowers. Inclusion of more species from Africa and southeast Asia (Aubriot et al., 882016) revealed that two additional African species with hermaphroditic flowers were members of the 89monophyletic group (S. lanzae J.-P.Lebrun & Stork and S. usambarense Bitter & Dammer). These 90studies used both plastid and nuclear molecular markers, and although the circumscription of the 91Eggplant clade improved, resolution within it was still poor. Although the brinjal eggplant and its wild 92relatives have clear morphological, molecular and eco-geographical differences (see Meyer et al., 932012; Knapp et al., 2013; Ranil et al., 2016), they have been difficult to resolve using traditional 94molecular markers (Vorontsova et al., 2013; Aubriot et al., 2016).

95 High-throughput sequencing and DNA library preparation improvements have made obtaining 96whole plastid genomes from total DNA extraction affordable for many non-model organisms driving a 97rapid increase in the number of genomes in public sequence repositories (see Tonti-Filippini et al., 982017). It has been suggested that complete plastid genomes can be used as 'superbarcodes' thus 99providing an abundance of characters to increase resolution in poorly resolved phylogenies (Williams 100et al., 2016).

Here we use sequences of whole chloroplast genomes to improve circumscription of and 102resolution in the Eggplant clade, and to test hypotheses concerning the origin of the close relatives of 103the cultivated eggplant, *Solanum melongena*. To account for morphological and geographical 104variability of the species, we designed a two-fold approach. First, we used Sanger data to test species 105circumscriptions for four widespread species of the Eggplant clade; we investigated two nuclear loci, 106ITS and the *waxy* (GBSSI) gene, and one plastid intergenic spacer, *ndhF-rpL32*, to check for 107topological incongruences. Second, based on this first analysis, we generated a plastome phylogeny of 108the Eggplant clade that we used to re-assess Lester and Hasan's (1991) hypothesis of linear step-wise 109migration from Africa to Asia and use ancestral areas reconstruction to investigate biogeography and 110^{dispersion} of the Eggplant clade as currently circumscribed. We relate dispersion and spread in 111^{widespread} eastern African species to large mammalian herbivores that feed on *Solanum* fruits 112(Pringle et al., 2014; Kartzinel et al., 2015). Our results have implications for future research on wild 113^{sources} of genetic variability for eggplant improvement, particularly in the context of global climate 114^{change} and efforts to develop the genetic basis for sustainable pest control.

115

116MATERIALS AND METHODS

117 Taxon sampling—Sampling for preliminary Sanger analysis—For Sanger analysis, we 118 sampled across the distribution range of four widespread species (S. campylacanthum, S. cerasiferum 119Dunal, S. incanum L. and S. insanum L.) in order to test species monophyly and reveal any geographic 120patterns. For the widespread and weedy eastern African species *S. campylacanthum*, we sampled three 121specimens from the northernmost part of its distribution area (Ethiopia), one from Tanzania and two 122 from the southernmost part (Namibia and South-Africa). Both S. cerasiferum and S. incanum are 123distributed across northern sub-Saharan Africa, with *S. incanum* also extending into regions of Middle 124East and central Asia as far as Pakistan (Vorontsova and Knapp, 2016). For S. cerasiferum we used 125 specimens from Ethiopia and the Central African Republic, and for *S. incanum* one specimen from the 126 westernmost part of the distribution range (Burkina Faso), two from the centre (Kenya and Oman), as 127 well as one from the easternmost part (Pakistan). Solanum insanum, the wild progenitor of the 128 cultivated eggplant (Meyer et al., 2012; Ranil et al., 2016), is distributed across tropical Asia, but also 129 occurs in Madagascar; it does not occur in Africa with no populations in East Asia and Africa (Ranil et 130al., 2016); we sampled two specimens from tropical Asia (China and India) and one individual from 131Madagascar. Our sampling is not designed to explore the structure of relationships between the 132 populations of S. insanum and S. melongena; this will require population level sampling across the 133range of both taxa (e.g., Meyer et al., 2012; W. Jinxiu, pers. comm.).

To place our study within the wider framework of Old World spiny solanums, we also 135sampled closely related species (sensu Vorontsova et al., 2013, Aubriot et al., 2016). A clade of four 136tropical Asian species (*"Solanum violaceum* and relatives" sensu Aubriot et al., 2016), as well as three 137African species (*S. nigriviolaceum* Bitter, *S. polhillii* Voronts. and *S. supinum* Dunal) were found to be 138closely related to the Eggplant clade (Vorontsova et al., 2013; Aubriot et al., 2016); therefore we 139included in our sampling one accession of *S. violaceum* Ortega and one accession for each of these 140three African species. Two informal groupings of a large number of African and Asian species, the 141'Anguivi grade' and the 'Climbing clade' (Vorontsova et al., 2013; Aubriot et al., 2016) were also 142sampled in order to better test the circumscription of the Eggplant clade. We selected six species of the 143'Anguivi grade', including the two cultivated African solanums, the Gboma eggplant (*S. macrocarpon* 144L.) and the Scarlet eggplant (*S. aethiopicum* L.), as well as their wild progenitors (*S. dasyphyllum* 145^{Schumach. & Thonn. and *S. anguivi* Lam. respectively). Most species of the 'Anguivi grade' have 146^{geographic} distributions centred in continental eastern Africa, but several species occupy disjunct 147^{regions} such as the Seychelles and India (see Vorontsova et al., 2013; Vorontsova and Knapp, 2016; 148^{Aubriot} et al., 2016 for details on geographic distributions). To identify potential long-distance 149^{dispersal} events, we sampled one species with a tropical Asian distribution (*S. trilobatum* L.) and 150^{another} that occurs within the Middle Eastern region (*S. glabratum* Dunal). We used an eastern 151^{African} scandent shrub from the 'Climbing clade', *S. richardii* Dunal, to root all the trees. Sampling 152^{for} this analysis includes 36 accessions (24 species), accounting for all 13 species of the Eggplant 153^{clade} and for 11 additional spiny solanum species from Africa and Asia.}

154

Sampling for whole chloroplast genomes analysis—After using Sanger sequencing for initial Seasessment 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 Is8two sets of Illumina reads at our disposal. If a species was not shown to be monophyletic based on Species analyses, one accession per lineage was included. All outgroup species used in the 160preliminary analysis were included in the full-plastome analysis, with the exception of *S*. 161*nigriviolaceum* for which we had insufficient DNA extract available.

162

DNA extraction, amplification, sequencing and data assembly—Silica gel dried or herbarium 164material was ground using a Mixer Mill MM 300 with glass beads (Qiagen Inc., Valencia, CA, USA) 165and with molecular grade sand. Total genomic DNA was extracted using the same two-step protocol 166than the one described by Aubriot et al. (2016). We have also extracted chloroplast DNA from fresh 167leaves of *Solanum melongena* 'Black Beauty' following Shi et al. (2012).

Amplified DNA regions used in the preliminary Sanger analysis were selected based on 169previous broad-scope studies on *Solanum* phylogenetics (Särkinen et al., 2013; Aubriot et al., 2016). 170These latter have shown that the plastid intergenic spacer *ndhF-rpL32*, the internal transcribed spacer 171(ITS) and the granule-bound starch synthase I (GBSSI) or *waxy* gene are particularly fitted for 172resolving phylogenetic relationships as the species level. Considering the abundant literature on DNA 173barcoding of landplants and on phylogenetics of *Solanum*, the use of these three rapidly evolving and 174*a priori* unlinked DNA regions should be reliable for testing current species boundaries. We 175downloaded 62 sequences from GenBank, and generated 61 for this study (see Appendix). For each 176additional accession, the three DNA regions were amplified and sequenced following the same 177procedures and techniques than those described in Aubriot et al. (2016). Sanger reads were assembled 178and edited in Geneious v.7.1.9 (Biomatters Ltd., Auckland, New Zealand). All sequences were 179automatically aligned with MAFFT v.7.305 (Katoh and Standley, 2013), using the L-INS-i algorithm. 180Newly generated sequences are deposited in GenBank and listed in the Appendix.

Starting quantities of total genomic DNA for the 25 samples used in the phylogenomic study 182were determined by measurement with a Qubit 2.0 Fluorometer (Life Technologies Corp., Carlsbad, 183CA). DNAs were diluted to approximately 3.8 ng/ μ l and libraries were prepared using the TruSeq 184Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA) following manufacturer instructions. 185Sequencing was performed at the Natural History Museum (London) sequencing facility using the 186Illumina MiSeq platform (2 × 250 bp) and loading concentration of 6 pM. Illumina adapters were 187removed, and all reads were quality-trimmed and quality-filtered using Trimmomatic 0.35 (Bolger et 188al., 2014) with default settings.

189

Plastome assembly, annotation and alignment—All paired reads were *de-novo* assembled 191with IDBA-UD 1.1.1 (Peng et al., 2012), using maximum k-value = 250 bp and minimum k-value = 19250 bp. Resulting contigs were uploaded into Geneious v.7.1.9 (Kearse et al., 2012) where they were 193quality-checked and length-filtered. Segments of the *de novo* plastomes were then circularized by 194matching end points. A high coverage plastid genome (800×) was also assembled *de novo* for 195*Solanum melongena* 'Black Beauty' from the collected reads using the built-in Geneious assembler 196platform with zero mismatches and gaps allowed among the reads. A similar procedure was conducted 197using Velvet v1.2.10 (Zerbino and Birney, 2008) with k-mer length 37, minimum contig length 74 and 198default settings by applying 800× upper coverage limit. The results of the two *de novo* methods were 199compared and inspected. Sanger-based gap closure and IR junction verification was carried out 200following Moore et al. (2007).

Sequence inconstancies were checked by mapping each set of reads to their respective 202plastome as described in Wysocki et al. (2015). Assembled plastomes were aligned to the completely 203annotated bittersweet (*Solanum dulcamara* L.) plastome available on GenBank (NC035724; 204Amiryousefi et al., unpublished). Annotations were transferred from the references to the assembled 205plastomes using a similarity threshold of 80%, and were then checked and refined manually. All 25 206plastomes were aligned with MAFFT v.7.305 using the FFT-NS-2 algorithm, and the alignment was 207inspected for structural mutations. The pairwise distances between each pair of plastomes were 208calculated and analysed using the Kimura Maximum Composite Likelihood model implemented in 209MEGA 7.0 (Kumar et al., 2016).

210

Phylogenetic analyses and dating—Sanger and genomic alignments were subjected to both 212maximum likelihood (ML) and Bayesian inference (BI) analyses. All analyses were run via the 213CIPRES Science Gateway (Miller et al., 2010) using RAxML-HPC v.8.1.24 (Stamatakis, 2014) for 214ML and MrBayes 3.2.6 (Huelsenbeck and Ronquist, 2001) for BI. In all analyses gaps were treated as 215missing data (N) and Solanum richardii served as the outgroup. Sanger alignments were analysed both 216separately and as combined datasets. RAxML analyses were run by using the rapid bootstrap 217algorithm with 1,000 bootstrap iterations. The GTR + G rate substitution model was applied to each 218matrix in order to fit the substitution models implemented in BI (see below); the proportion of 219 invariable sites was not estimated as it is not recommended by RAxML developer (Stamatakis, 2016). 220Bootstrap replicate trees were used to draw bipartitions on the optimal ML tree; nodes with bootstrap 221support values (BS) below 50% were pruned to obtain the ML 50% majority-rule tree. For BI 222analyses, the best-fitting nucleotide substitution model inferred with MrModeltest v.2.3 (Nylander, 2232004) via estimation of the Akaike information criterion was assigned to each of DNA region. The 224 regions *ndhF-rpL32* and ITS followed a GTR + I + G substitution model, whereas the GTR + I model 225better fitted for the waxy region. MrBayes analyses constituted of two independent parallel runs of 226 four Markov chains each, run for 10 million generations and trees were sampled every 1000 227 generations. Adequate mixing of Markov chains and convergence of the two runs were confirmed with 228Tracer v1.6 (Rambaut et al., 2014); after a 10% burn-in, remaining trees were used to generate a 50% 229Bayesian majority-rule consensus tree. Congruence between the topologies obtained for each of the 230three markers was visually inspected before concatenating them into a combined matrix that was 231 divided into three partitions, to which the best-fitted substitution models were applied. The combined 232matrix was run under the same conditions detailed above for the single marker matrices.

We used the same BI and ML approaches to analyse the phylogenomic dataset. The plastome 234matrix was analysed with MrModeltest, and the GTR + I + G nucleotide substitution model was 235selected as best-fitting the data. For same reasons than stated above, this model was implemented in 236MrBayes whereas the GTR + G rate substitution model was preferred for RAxML analyses. All 237conditions of MrBayes and RAxML runs were identical to those implemented for analyses of Sanger 238alignments.

Relative divergence times, together with the topology were estimated using BEAST v.1.8.4 240(Drummond et al., 2012), with a GTR + I + G rate substitution model applied to the plastome matrix. 241A Yule speciation tree prior and a relaxed uncorrelated clock-model that allows rates to vary 242independently along branches (Drummond et al., 2006) were used with all other parameters set to 243default. We used two temporal constraints to calibrate the BEAST analyses: (1) age of the Eggplant 244clade (mean = 1.54 Myr; standard deviation = 0.5), (2) age of the root (mean = 3.38 Myr; standard 245deviation = 0.5). These were derived from the Solanaceae-wide phylogeny of Särkinen et al. (2013), 246with uncertainty regarding these dates incorporated by assigning normal prior distributions to the two 247calibration points (Couvreur et al., 2008; Evans et al., 2014). Four independent BEAST analyses were 248run, each 100 million generations, sampling every 10,000 generations. Convergence of all parameters 249was assessed using Tracer 1.5 (Rambaut et al., 2014), the first 10% of each of the four Markov chains 250was removed as burn-in, and chains were combined with LogCombiner 1.7.2. (Drummond et al., 2512012) to calculate the maximum clade credibility tree.

252

253*Biogeographic inference and dispersion analysis*—Species distributions were determined from 254collection data stored in Solanaceae Source Database (2017, http://www.solanaceaesource.org/); 255cultivated species (*Solanum melongena, S. aethiopicum, S. macrocarpon*) were excluded from the 256biogeographic analysis. Because several species are widespread, with distribution areas that 257encompass all eastern Africa (*S. campylacanthum*) or that extend from western Africa to Pakistan (*S. 258incanum*), circumscription of biogeographic areas that are meaningful in terms of patterns of species 259distribution and that are coherent climatically and biogeographic evidence (see below): (A) 261Tropical Asian region; (B) Somalian and Middle East region; (C) Sudanian region; (D) Congolian 262region; (E) Zambesian region; (F) Southern African region; (G) Malagasy region (see Fig. 1A). Four 263of these (C, D, E and F) directly correspond to those of Linder et al. (2012). Linder's regions are 264statistically defined using patterns from groups of vascular plants and vertebrates, and were shown to 265be robust between groups. This biogeographic framework is well suited for these solanums with 10 of 26623 taxa included in our analysis restricted to one region, and several clades that are characteristic of a 267set of connected biogeographic regions (Southern African clade; Fig. 1B).

Boundaries of the biogeographic regions used in this study are subject to some uncertainty due contrasting climatic patterns (e.g. the wet aseasonal Congolian region vs the drier and more 270seasonal Sudanian region). Areas around the Tanzania/Kenya border in east Africa correspond to a 271zone of high mountainous complexity (i.e. Albertine Rift and Kilimanjaro Massif), and have been 272previously assigned to either the Somalian or the Zambesian regions, depending on the data types used 273(White, 1983, 1993; Linder et al., 2012). This region has also been identified as a centre of diversity 274for African spiny solanums (Syfert et al., 2016; Vorontsova and Knapp, 2016). In addition, two of our 275sampled species, *S. polhillii* and *S. usambarense*, are restricted to this transitional region (Vorontsova 276and Knapp, 2016) and assigning species to particular biogeographic region was challenging; therefore 277we accounted for this complexity by assigning these two species to both the Somalian and the 278Zambesian regions.

Handling of the Middle East area presents another challenge as this region does not account endemic species of spiny solanums in our sampling (Aubriot et al., 2016). Only *S. glabratum sland S. incanum* are present in the Middle East, both have distribution ranges centred on northeastern 282Africa (i.e., Sudanian region). The Red Sea is not a strong dispersal barrier, being only about 25 km 283wide at its southernmost and narrowest (natural) point, and we combined the Middle East with the 284Sudanian region, hence delimiting our region B. Similarly, it is difficult to draw the eastern boundary 285 of region B as the Middle East acts as a biogeographic crossroads between Asia and Africa, its biota 286 being of composite origin (Buerki et al., 2012). Among all spiny solanums native of the Middle East 287 region, two are distributed from eastern tropical India to the Arabian Peninsula (*S. pubescens* Willd., 288 S. *virginianum* L.) while others are not present in tropical regions as their distributions are restricted to 289 dry regions between northeastern Africa and northwestern India (*S. cordatum* Forssk., *S. forskalii* 290 Dunal, *S. glabratum*, *S. incanum*, *S. platacanthum* Dunal). Given such distribution patterns, we used 291 the Thar Desert of India as the eastern boundary of zone B (Sudanian region and Middle East). This 292 delimitation also corresponds to the western limit of the Asian monsoon climatic regime, acting as a 293 proxy of the limit between the tropical areas of region A (Tropical Asia) and the mostly dry areas of 294 region B (Holt et al., 2013).

The history of Madagascar and of its neighbouring archipelagos (Mascarenes, Granitic 296Seychelles and Comoros) is characterized by complex geological and climatic events that 297accompanied their isolation from Africa and the Indian subcontinent. The Malagasy endemic flora has 298a composite origin, with elements from Africa, the Indian subcontinent and South-East Asia (Buerki et 299al., 2013). The spiny solanums endemic to Madagascar (not sampled here) are not closely related to 300the Eggplant clade and form a monophyletic clade whose phylogenetic placement is still unresolved 301(Vorontsova et al., 2013; Aubriot et al., 2016) within the larger Old World lineage of spiny solanums. 302We sampled four species that are native to Madagascar and the Mascarenes; two are present in Africa 303and in Madagascar (*S. anguivi* and *S. richardii*), while the two others have a distribution range centred 304on tropical Asia, with *S. insanum* (the wild progenitor of the eggplant) present in Madagascar (Knapp 305et al., 2013; Aubriot *et al.*, 2016; Ranil et al., 2016) and *S. violaceum* recorded from Mauritius 306(Vorontsova and Knapp, 2016). Hence, we treated Madagascar and the neighbouring archipelagos as a 307distinct biogeographic area; the Malagasy region (region G).

The dispersal-extinction-cladogenesis (DEC) likelihood model implemented in Lagrange v. 30920130526 (Ree and Smith, 2008) was used to investigate the biogeographic history of the Eggplant 310clade. The analysis was performed on the BEAST maximum clade credibility tree using the following 311set of parameters: (1) species ranges were coded using presence in the seven biogeographic regions; 312(2) maximum range size at nodes was constrained to 3; (3) areas not formed of adjacent regions were 313excluded from the analysis (adjacency matrix); (4) the dispersal constraints matrix was used to set the 314dispersal rates between the seven biogeographic regions with a value of 1 for adjacent regions (e.g., E 315and F), 0.5 between sea-separated regions (e.g., G) and 0.01 for nonadjacent regions (e.g., A and F). 316Ancestral area reconstruction for each node was plotted on the BEAST maximum clade credibility tree 317using pie charts (see Buerki et al., 2011 for details on the R script).

318 To assess the potential influence of fruit-eating migratory mammals in the dispersion of the 319African widespread Eggplant clade species (*Solanum campylacanthum, S. cerasiferum* and *S.*

320*incanum*), a distribution map of these species together with two of their known large-mammalian seed 321dispersers (African elephant, *Loxodonta africana* (Blumenbach) and impala, *Aepyceros melampus* 322(Lichentenstein); see Pringle et al., 2014; Kartzinel et al., 2015) was reconstructed using the R 323packages *maps* and *dismo*. Shapefiles of the current distributions of the African elephant and impala 324were downloaded from the IUCN Red List of Threatened Species website (IUCN SSC Antelope 325Specialist Group, 2016; Blanc, 2008). Using georeferenced herbarium specimens, we gathered data on 3261,150 populations of *S. campylacanthum*, *S. cerasiferum* and *S. incanum*. A Venn diagram showing 327the overlap between the distributions of the widespread Eggplant clade species and the African 328elephant and impala was inferred using the R package *VennDiagram*.

329

330RESULTS

Sanger reconstructions and species delimitation—Visual comparison of 50% majority-rule 332trees obtained from single marker matrices with BI and ML showed no strongly supported topological 333conflict (BS > 75% and PP > 0.9) (only the BI consensus trees shown; see Appendix S1a,b,c in the 334Supplemental Data with the online version of this article). Topologies are overall all poorly resolved 335with the exception of few strongly supported nodes; they only differed in the degree of resolution, 336with the ITS consensus trees exhibiting higher proportions of supported nodes than those inferred from 337waxy and *ndhF-rpL32*.

338 The combined Sanger dataset yielded a total of 3,121 base pairs (bp) of which 256 bp were 339variable. There were no supported conflicts between the BI and ML 50% majority-rule trees, with the 340BI tree showing a much greater resolution than the ML one, albeit often with poorly supported nodes. 341The combined topology (Appendix S1d) shows that the Eggplant clade as previously defined (Voront-342sova et al., 2013; Aubriot et al., 2016) is monophyletic (PP = 0.99) and it includes the Cape Verde en-343demic *Solanum rigidum* Lam. Within the Eggplant clade, two of the widespread species are resolved 344as monophyletic (*S. campylacanthum*, *S. cerasiferum*) and one as paraphyletic (*S. incanum*). The three 345accessions of S. incanum from the eastern part of the species distribution range (Kenya, Oman and 346Pakistan) form a strongly supported lineage (BS = 99%; PP = 1) that excludes the accession from 347Burkina Faso; the latter is sister to the clade formed by the eggplant (S. melongena) and its wild pro-348genitor (*S. insanum*). Solanum melongena is part of a polytomy that includes all three accessions of *S.* 349*insanum*; this confirms previous phylogenetic results (Vorontsova et al., 2013; Aubriot et al., 2016). 350Affinities of the species of the Eggplant clade remain largely unresolved using these markers, with the 351exception of a group of species from southern Africa (S. aureitomentosum Bitter, S. lichtensteinii 352Willd., S. linnaeanum, S. umtuma Voronts. & S.Knapp) forming a strongly supported group (BS = 35390%; PP = 1) that had been recognised in previous analyses (Vorontsova et al., 2013; Aubriot et al., 3542016).

355

356 Plastome phylogenomic analyses—Plastome lengths ranged from 155,017 to 155,820 bp, 357 with a mean coverage between $10.2 \times$ and $254.8 \times$; see Table 1 and Appendix S2 for details on the 358plastomes. The plastome alignment was 159,227 bp in length, of which 3,890 bp were variable. There 359were no supported conflicts between the BI topologies obtained with BEAST and MrBayes, and the 360RAxML 50% majority-rule tree (see Appendix S3). Phylogenetic trees are all fully resolved with the 361exception of a polytomy in the MrBayes and RAxML 50% majority-rule trees. For the MrBayes topo-362logy all other nodes have PP \geq 0.99 (Appendix S3b). For the BEAST and RAXML trees, 13 out of 15 363nodes within the Eggplant clade are strongly supported (*i.e.*, BS \geq 87% and PP \geq 0.97; Fig. 1B and 364Appendix S3a,c). Plastome trees confirm results from the Sanger analyses, but provide much better 365resolution of relationships in the clade (Fig. 1B). The Eggplant clade (*sensu* Vorontsova et al., 2013; 366Aubriot et al., 2016) is recovered as monophyletic; it is sister (BS = 82%; PP = 0.94) to a lineage 367 formed by two African (Solanum polhillii and S. supinum) and one tropical Asian species (S. trilob-368*atum*), partly confirming previous phylogenetic results (Vorontsova et al., 2013; but see Aubriot et al., 3692016 for an alternative position of *S. trilobatum*). The first-branching lineage, here called 'Kilimanjaro 370clade' (Fig. 1B), groups three species from the eastern African region (S. aqnewiorum, S. lanzae, S. 371*usambarense*). Affinities within this clade are still unresolved; in the BEAST analysis *S. agnewiorum* 372branches first but this result is not supported (PP = 0.41). The next branching lineage is formed by the 373eggplant and its wild progenitor, S. insanum, the two accessions of the cultivated eggplant being 374 grouped together (Fig. 1B). This lineage is sister to a monophyletic group that includes all remaining 375species of the Eggplant clade and is composed of two sister clades: (1) the Southern African clade 376(Fig. 1B) with four species from southern Africa (S. aureitomentosum Bitter, S. lichtensteinii, S. lin-377*naeanum* and *S. umtuma*), and (2) the Widespread clade (Fig. 1B) that includes three species with 378large distribution ranges (S. campylacanthum, S. cerasiferum and S. incanum) and the Cape Verde is-379lands endemic S. rigidum. Within the Southern African clade, S. aureitomentosum branches first, fol-380lowed by S. linnaeanum that is sister to a moderately supported grouping formed by S. lichtensteinii 381and *S. umtuma* (BS = 83%; PP = 1). *Solanum campylacanthum* is the earliest diverging species of the 382Widespread clade; it is sister to a lineage in which S. cerasiferum branches first, followed by one of 383the two accession of *S. incanum* (from Kenya) that is in turn sister to a grouping formed by the Cape 384Verdean S. rigidum and the second accession of S. incanum (from Burkina Faso). 385

386 *Historical biogeography of the Eggplant clade and dispersion analysis*—The most likely 387biogeographic scenario for the Eggplant clade is displayed in Figure 1b. The Eggplant clade originated 388in area B (Somalian and Middle East region) sometimes during the early Pleistocene. The Kilimanjaro 389clade maintained this distribution range with the exception of *Solanum usambarense*, which expanded 390its distribution range into area E (Zambesian region). There is a high level of biogeographic uncer-391tainty on the node separating the cultivated eggplant plus its progenitor (*S. insanum*) from the rest of 392the species in the clade. This is due to a disjunct distribution between these two clades (Fig. 1B). The 393most parsimonious scenario invokes two dispersal events leading to the most common ancestor of the 394whole clade therefore achieving a widespread distribution across regions BE and A (Tropical Asian re-395gion). The disjunct distribution of *S. insanum* in A and G (Malagasy region) most likely results from a 396long-distance dispersal event from A to G during the last million years. The majority of inferred dis-397persal events took place between the Southern African and Widespread clades. Southward and west-398ward dispersal from B has been active in these two clades over the last 2 Myrs, and F (Southern 399African region) was colonized during the past 1 million years. In addition to having acted as a cradle, 400area B has been a pivotal region in shaping the current distribution of the eggplant clade. The inferred 401dispersion scenarios are illustrated in Fig. 1A.

402 The distribution of *Solanum campylacanthum*, *S. cerasiferum* and *S. incanum* together with 403their large mammalian seed dispersers is displayed in Fig. 2A. The distribution of African elephants 404and impalas covers 53.7% (617 in total) of the populations of these three widespread species (Fig. 2B). 405The elephant range currently only covers 24.3% (150) of the three *Solanum* species populations, while 406the impala distribution covers 94.4% (583); most *Solanum* populations only overlap the impala range 407(467 populations) (Fig. 2B).

408

409DISCUSSION

While studies focusing on the morpho-ecological and genetic diversity of the eggplant and its 411wild relatives are increasing (e.g., Gramazio et al., 2016; Kaushik et al., 2016; Ranil et al., 2016; 412Syfert et al., 2016; Acquadro et al., 2017) our study is the first designed to test phylogenetic and 413biogeographic scenarios in the Eggplant clade (*sensu* Aubriot et al., 2016). Our sampling for Sanger 414and genomic studies allowed us to test recent species delimitations and build a phylogeny for the 415Eggplant clade that is well resolved internally. This study strongly supports *Solanum insanum*, the 416only wild Eggplant clade species currently distributed within the original centre of domestication of *S*. 417*melongena* (Meyer et al., 2012), as its most likely progenitor (Fig. 1A,B). The Somalian and Middle 418East region is identified as the cradle of the Eggplant clade with recent dispersions to southern/western 419Africa and tropical Asia; distribution data suggest that large savannah herbivores contributed to the 420geographical expansion of several of the species (Fig. 2A,B). The results do not, however, provide a 421definitive biogeographic scenario for the disjunct distribution between *S. insanum* + *S. melongena* and 422the rest of the species of the Eggplant clade that are all restricted to Africa and western Asia (see node 42310 in Appendix S4a).

425 *Phylogenetic relationships and species concepts in the eggplant clade*—Within the Eggplant 426clade, four species concepts are particularly challenging: Solanum campylacanthum, S. cerasiferum, S. 427*incanum* and *S. insanum* (the three first species belong to the Widespread clade, the last is sister to *S.* 428*melongena*; Fig. 1B). Widespread across Africa and Asia, these species are also extremely 429polymorphic, with morphological characters that vary within and between populations (Knapp et al., 4302013; Ranil et al., 2016). Olet and Bukenya-Ziraba (2001) suggested that S. campylacanthum (as "S. 431*incanum*) and *S. cerasiferum* belong to the same biological species; our phylogenetic results support 432the monophyly of these taxa and confirm they represent two distinct lineages. Sanger and full-433 plastome data unambiguously confirm S. insanum as the wild progenitor of the cultivated brinjal 434eggplant. While the two eggplant accessions are sister in the plastome topology, pairwise distances 435between the two plastomes of the eggplant and plastome of *S. insanum* are identical (1E-4) and *c.* 25 436 times smaller than the mean interspecific pairwise distances of the dataset (2.8E-3). Strong plastome 437 identity between the brinjal eggplant and its wild progenitor is not surprising as they are known to 438interbreed relatively freely when in sympatry (e.g., Davidar et al., 2015; Mutegi et al., 2015). Our 439 results also illustrate the fact that, while the cultivated eggplant and *S. insanum* present distinctive 440morphological features (Ranil et al., 2016), diagnostic molecular data are still lacking. Solanum 441*incanum* is recovered as paraphyletic, with the sample from the westernmost part of the distribution 442range (Burkina Faso) that does not group unambiguously with the other S. incanum accessions. 443Genomic data resolve the Cape Verde endemic S. rigidum nested within S. incanum, strongly 444supported as sister to the sample from Burkina Faso. The plastid genome of *S. rigidum* is very similar 445to that of the two accessions of S. incanum; mean pairwise distance (MPD) between these three 446plastomes is 3.5E-4. The phylogeny does not reflect morphological characters; S. rigidum is 447morphologically similar to S. cerasiferum and readily distinguishable from S. incanum by its deeply 448lobed leaves and sparser indumentum. This result can be interpreted in two ways. Solanum rigidum 449may be a recent morpho-geographical segregate of the continental S. incanum, both taxa having 450conserved their ancestral chloroplast genome. Internal transcribed spacer (ITS) data (Appendix S1a) 451resolve *S. rigidum* as sister to *S. macrocarpon* (gboma eggplant) and *S. dasyphyllum*, albeit this result 452 is poorly supported (PP = 0.82); this would suggest a hybrid origin via chloroplast capture with 453contributions from *S. incanum* (as the maternal parent of the cross) and arguably from *S. macrocarpon* 454or its wild progenitor, S. dasyphyllum (see also Knapp and Vorontsova, 2013). Chloroplast capture 455 revealed by whole plastid genome sequences has been reported in the tomato clade (Solanum sect. 456Lycopersicon; Poczai et al., unpublished) as well as among species of Gossypium (Chen et al., 2017) 457 and Gaultheria (Zhang et al., 2017). Obtaining broader geographic sampling of Eggplant clade 458species, particularly S. rigidum and accessions from western Africa, using data suited to this 459taxonomic depth, such as nuclear SSRs or SNPs (e.g., van Loo et al., 2015; Acquadro et al., 2017) 460may help better resolve circumscription of *S. incanum* and *S. rigidum*.

461 Our inferred plastome phylogeny provides much better resolution at the interspecific level 462 within the Eggplant clade than previous phylogenetic investigations that either lacked taxonomic 463 representativeness (Meyer et al., 2012; Acquadro et al., 2017) or phylogenetic resolution (Vorontsova 464et al., 2013; Aubriot et al., 2016) or both (Weese and Bohs, 2010). The topology we obtained, with the 465exception of one internal nodes (Fig. 1B), is fully resolved with strongly supported nodes. This 466plastome phylogeny confirms several phylogenetic hypotheses while it contradicts others. Taxonomic 467 composition and phylogenetic structure of the Eggplant clade as found by Aubriot et al. (2016) is 468confirmed but with the addition of *Solanum rigidum* and much improved resolution and support. Three 469species that were found in polytomy at the base of the Eggplant clade (*S. agnewiorum*, *S. lanzae* and *S.* 470usambarense) here are resolved as a basal Kilimanjaro clade; the Southern African clade and 471Widespread clade, whose compositions globally overlap clades that were previously poorly supported 472 with Sanger data (see Aubriot et al., 2016; Vorontsova et al., 2013) are here resolved with highly 473supported nodes. Phenetic results obtained by Acquadro et al. (2017) with a data set of 12,859 nuclear 474SNPs, but with limited taxonomic sampling (6 Eggplant clade spp.) are not fully congruent with our 475plastome topology; comparisons are difficult due to differences in sampling and methodological 476approaches.

477

478 Effect of savannahs and mammalian dispersers on biogeography of the Eggplant clade— 479Our biogeographic results show that the Eggplant clade originated recently, during the last *c*. 3 Myr, in 480the Somalian and Middle East region (Fig. 1A,B). Spread then occurred mostly within Africa; the 481progenitor of the eggplant, *Solanum insanum*, is the only wild species that is not present in continental 482Africa, but that has a Malagasy-Southeast Asian distribution. Dispersion to tropical Asia that gave rise 483to *S. insanum* is likely to have happened during the Early Pleistocene via the Middle East region; we 484suggest dispersion to Madagascar was a later event, possibly linked to recent human migration (see 485Crowther et al., 2016). Our results challenge Lester and Hasan's (1991) hypothesis of an eastern 486African origin for the 'ancestors of S. melongena' with subsequent spread to tropical Asia via a step-487wise expansion through Middle East. Our biogeographic analysis instead shows that the origin of the 488Eggplant clade lies in northeastern Africa and the Middle East, with spread both south (Widespread 489and Southern African clades) and east (S. insanum + S. melongena). Importantly, it shows that the 490tropical Asian lineage of *S. insanum* did not proceed from a step-wise expansion through Middle East, 491but instead from an early dispersion from Africa, unrelated to the southwards spread of the 492Widespread and Southern African clades. This first attempt to provide a testable biogeographic 493 framework for the Eggplant clade shows that its history consists in a wide dispersion across the 494African continent rather than an 'Out of Africa' phenomenon. Nesting of Asian species such as S.

495*insanum* in otherwise African lineages, suggesting dispersal events in African lineages are not rare in 496Old World spiny solanums; *S. trilobatum* and *S. violaceum* also show this pattern in our plastome 497phylogeny (see also Aubriot et al., 2016).

Species of the Eggplant clade have a recent biogeographic history and several species have 499large and complex distribution ranges, which makes it difficult to untangle biotic and abiotic factors 500that could explain evolutionary dynamics in the clade. Many spiny solanums are plants of disturbed 501areas, and most species of the Eggplant clade are weedy and occur in nitrogen-rich habitats of arid-502areas or gaps in tropical forests, in a wide range of altitudes (Vorontosva and Knapp, 2016; Knapp and 503Vorontsova, 2013). Only the three species of the basal Kilimanjaro clade are habitat-restricted; they 504are found in higher elevation (> 1,200 m) montane woodlands (Vorontosva and Knapp, 2016). These 505are also either hermaphroditic (*Solanum lanzae, S. usambarense*) or only weakly andromonoecious (*S.* 506*agnewiorum*) with small fruits, while strong andromonecy characterizes all other species of the 507Eggplant clade (Vorontosva and Knapp, 2016; Vorontsova et al., 2013), with larger fruits. Our data 508suggest that the combination of breeding system and ecology (see below) in this lineage have been 509important factors in the dramatic expansion of its distribution range over the last 2 Myrs.

The increased geographical spread of members of the Eggplant clade coincides with profound 510 511habitat changes in Africa. Decrease in atmospheric CO₂ concentrations and aridification beginning in 512the late Miocene led to global increase in the biomass of plants using C_4 photosynthesis. This resulted 513in the extension of savannahs and fire-prone environments, first in northeastern Africa and then in 514South Africa (Cerling et al., 1997; Uno et al., 2016). Recent studies on hominin environments (Cerling 515et al., 2011; Uno et al., 2016) found that woody grassland savannahs became persistent in northeastern 516Africa at least 6 Myr BP, with major increases in the abundance of savannah grazing mammalian s 517 over this period (Uno et al., 2011). Open grassland and wooded grassland savannahs are typical 518habitats for members of both the South-African and Widespread clades, and large savannah 519mammalians such as African elephants and impalas have been recorded as important herbivores and 520seed dispersers of Eggplant clade species (Pringle et al., 2014; Kartzinel et al., 2015). In Africa 521elephants and impalas have native distributions that overlap those of members of the Eggplant clade, 522and particularly species of the Widespread clade (see Fig. 2A,B). Although the historical range of 523A frican elephants fully encompassed the current distribution of the Eggplant clade species (see Ripple 524et al., 2015), it appears that today the impala is likely to be more important for the seed dispersal of 525widespread species. We suggest that the extensive geographical expansion over the last 2–3 Myr and 526the impressive ranges of some of the extant species (e.g., *Solanum campylacanthum, S. cerasiferum, S.* 527*incanum*) is related to the dispersal of seeds by these large, migratory mammalian herbivores. 528Contraction of ranges of these large mammalian seed dispersers could ultimately contribute to 529isolation of populations previously more intermixed, leading to differentiation and speciation. Human

530use of these species could have also contributed to their recent wide dispersal and spread; many 531species of the clade are used medicinally and for tanning hides (Bitter, 1923; Ranil et al., 2016; 532Vorontsova and Knapp, 2016). Fragments of *Solanum* seeds have been recorded from a 780,000 yr old 533hominin site in Israel (Melamed et al., 2016), and archaeological remains may provide additional 534information about early interactions between hominins and Eggplant clade species.

535 Disentangling the relationship between complex changes in ecology and differentiation in the 536Eggplant clade, and in spiny solanums more broadly, is challenging, especially in the light of recent 537profound environmental change; new types of data will be required to understand the evolutionary 538success of this cosmopolitan and very diverse group.

539

Eggplant domestication, still a biogeographic puzzle—Our understanding of the evolutionary 541history and biogeography of the eggplant and its wild relatives is now better understood following the 542analyses conducted here using whole plastome and nuclear sequence data. The two nuclear regions we 543analysed here do not contradict plastome data, but nuclear genomic data for all the species of the 544Eggplant clade will be necessary to assess phylogenetic discordance and detect historical gene flow. 545Combination of plastome and denser nuclear genomic data would also open the way to in-depth 546investigation of the genomic basis of environment-specific and lineage-specific adaptations in 547eggplant CWR (e.g., Pease et al., 2016). This will ultimately provide plant breeders and physiologists 548with an estimate of the potential for selection from standing variation across eggplant CWR, 549potentially revolutionizing genetic improvement of eggplants.

Our study also shows that even with sampling across the full diversity of the Eggplant clade 551and a robust phylogenomic framework, our understanding of the biogeographic history of the Eggplant 552clade can still be improved; in particular dispersion of *Solanum insanum* lineages to tropical Asia and 553emergence and domestication of the brinjal eggplant. Meyer et al. (2012) suggested that *S. melongena* 554was domesticated several times from *S. insanum* (as *S. incanum*); use of *S. insanum* as a medicine 555(Ranil et al., 2016) and extensive movement of peoples across southeastern Asia support this idea. 556Combining sampling for *S. insanum* through its whole geographical range with the use of recently de-557veloped targeted high-throughput sequencing (Kadlec et al., 2017) will provide a promising approach 558for reconstructing reliable scenarios of evolution of brinjal eggplant from its wild progenitor, *S. in*-559*sanum*.

560ACKNOWLEDGEMENTS

561We thank the curators of A, BM, K, LISC and MH for permission to sample DNA from specimens 562held in their collections; Pr. H. Balslev (Aarhus University), the Bangkok Forest Herbarium (BKF), 563Pr. P. Chantaranothai and the Botanical team of the Applied Taxonomic Research Center of Khon 564Kaen University for field assistance in Thailand; Dr. D. Narasimhan (Madras Christian College), Dr. 565V. Sampath Kumar, Dr. G.V.S Murthy, Dr. P. Lakshminarasimhan, Dr. A. Benniamin, G. 566Gnanasekaran (Botanical Survey of India), P. Kale and R. Singh (BLAT) for their invaluable 567assistance during visits of Indian herbaria. We thank M.S. Vorontsova, M.-C. Daunay, M.C. Duarte 568and J. Prohens for their comments that significantly improved the manuscript; we thank the NHM 569sequencing facility team, particularly A. Briscoe and K. Hopkins, for assistance with molecular work. 570This work was funded in part by the Natural History Museum's Natural Resources and Hazards 571Initiative and by the Natural History Museum's Departmental Investment Fund.

572**REFERENCES**

573

574Acquadro, A., L. Barchi, P. Gramazio, E. Portis, S. Vilanova, C. Comino, M. Plazas, et al. 2017. Cod-575ing SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS* 576*One* 7: e0180774, doi: 10.1371/journal.pone.0180774.

577

578Aubriot, X., P. Singh, and S. Knapp. 2016. Tropical Asian species show the Old World clade of "spiny 579solanums" (the Leptostemonum Clade: Solanaceae) is not monophyletic. *Botanical Journal of the Lin-*580*nean Society* 180: 1–27, doi:10.1111/boj.12412.

581

582Bebber, D. P., T. Holmes, and S. J. Gurr. 2014. The global spread of crop pests and pathogens. *Global* 583Ecology and Biogeography 23: 1398–1407, doi: 10.1111/geb.12214.

584

585Bitter, G. 1923. Solana Africana Part IV. *Repertorium specierum novarum regni vegetabilis* 16: 1– 586320. Dahlem bei Berlin, Berlin, Germany.

587

588Blanc, J. 2008. Loxodonta africana. The IUCN Red List of Threatened Species 2008: 589e.T12392A3339343. [online] Website http://dx.doi.org/10.2305/IUCN.UK.2008.RLT-590S.T12392A3339343.en. [accessed 9 February 2018].

591

592Bohs, L. 2005. Major clades in *Solanum* based on *ndhF* sequences. *In* R. C. Keating, V. C. Hollowell, 593and T. B. Croat [eds], A festschrift for William G. D'Arcy: the legacy of a taxonomist, 27–49. Mis-594souri Botanical Garden Press, St. Louis, Missouri, USA.

595

596Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina se-597quence data. *Bioinformatics* 30: 2114–2120, doi: 10.1093/bioinformatics/btu170.

598

599Buerki, S., F. Forest, N. Alvarez, J. A. A. Nylander, N. Arrigo, and I. Sanmartín. 2011. An evaluation 600of new parsimony-based versus parametric inference methods in biogeography: a case study using the 601globally distributed plant family Sapindaceae. *Journal of Biogeography* 38: 531–550, doi: 10.1111/ 602j.1365-2699.2010.02432.x.

603

604Buerki, S., S. Jose, S. R. Yadav, P. Goldblatt, J. C. Manning, and F. Forest. 2012. Contrasting biogeo-605graphic and diversification patterns in two mediterranean-type ecosystems. *PLoS One* 7: e39377, doi: 60610.1371/journal.pone.0039377. 607

608Buerki, S., D. S. Devey, M. W. Callmander, P. B. Phillipson, and F. Forest. 2013. Spatio-temporal his-609tory of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society* 171: 304–329, 610doi: 10.1111/boj.12008.

611

612Cattivelli, L., F. Rizza, F. W. Badeck, E. Mazzucotelli, A. M. Mastrangelo, E. Francia, C. Marè, et al. 6132008. Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. 614*Field Crops Research* 105: 1–14, doi: 10.1016/j.fcr.2007.07.004.

615

616Cerling, T. E., J. M. Harris, B. J. MacFadden, M. G. Leakey, J. Quade, V. Eisenmann, and J. R. 617Ehleringer. 1997. Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389: 618153–158, doi: 10.1038/38229.

619

620Cerling, T. E., J. G. Wynn, S. A. Andanje, M. I. Bird, D. K. Korir, N. E. Levin, E. Mace, et al. 2011. 621Woody cover and hominin environments in the past 6 million years. *Nature* 476: 51–56, doi: 10.1038/622nature10306.

623

624Chen, Z., C. E. Grover, P. Li, Y. Wang, H. Nie, Y. Zhao, M. Wang, et al. 2017. Molecular evolution 625of the plastid genome during diversification of the cotton genus. *Molecular Phylogenetics and Evolu-*626*tion* 112: 268–276, doi: 10.1016/j.ympev.2017.04.014.

627

628Couvreur, T. L. P., L. W. Chatrou, M. S. M. Sosef, and J. E. Richardson. 2008. Molecular phylogenet-629ics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology* 6: 54, doi: 63010.1186/1741-7007-6-54.

631

632Crowther, A., L. Lucas, R. Helm, M. Horton, C. Shipton, H. T. Wright, S. Walshaw, et al. 2016. An-633cient crops provide first archaeological signature of the westward Austronesian expansion. *Proceed*-634*ings of the National Academy of Sciences* 113: 6635–6640, doi : 10.1073/pnas.1522714113.

636Daunay, M.-C., R. N. Lester, and H. Laterrot. 1991. The use of wild species for the genetic improve-637ment of brinjal (*Solanum melongena* L.) and tomato (*Lycopersicon esculentum* Miller). *In* J. G. 638Hawkes, R. N. Lester, M. Nee, and N. Estrada [eds], Solanaceae III: taxonomy, chemistry, evolution, 639389–412. Royal Botanic Gardens Kew, Kew, UK.

641Daunay, M.-C. and J. Janick. 2007. History and iconography of eggplant. *Chronica Horticulturae* 47: 64216–22.

643

644Daunay, M.-C., and P. Hazra. 2012. Eggplant. *In* P. Hazra, and K. V. Peter [eds], Handbook of vegeta-645bles, 257–322. Studium Press, Houston, Texas, USA.

646

647Davidar, P., A. A. Snow, M. Rajkumar, R. Pasquet, M.-C. Daunay, and E. Mutegi. 2015. The potential 648for crop to wild hybridization in eggplant (*Solanum melongena;* Solanaceae) in southern India. *Ameri*-649*can Journal of Botany* 102:128–138, doi: 10.3732/ajb.1400404.

650

651Dempewolf, H., R. J. Eastwood, L. Guarino, C. K. Khoury, J. V. Müller, and J. Toll. 2014. Adapting 652agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. 653*Agroecology and Sustainable Food Systems* 38: 369–377, doi: 10.1080/21683565.2013.870629. 654

655Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dat-656ing with confidence. *PLoS Biology* 4: e88, doi: 10.1371/journal.pbio.0040088.

657

658Drummond, A. J., M. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti 659and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973, doi: 10.1093/molbev/mss075. 660

661Evans, M., X. Aubriot, D. Hearn, M. Lanciaux, S. Lavergne, C. Cruaud, P. P. Lowry II PP, et al. 2014. 662The evolution of succulence: insights from a remarkable radiation in Madagascar. *Systematic Biology* 66363: 698–711, doi: 10.1093/sysbio/syu035.

664

665FAO. 2017. The statistical database (FAOSTAT). [online] Website http://www.fao.org/faostat/en/. 666[accessed 9 February 2018].

667

668Frodin, D. G. 2004. History and concepts of big plant genera. *Taxon* 53: 753–776.

669

670Gramazio, P., J. Blanca, P. Ziarsolo, F. J. Herraiz, M. Plazas, J. Prohens, and S. Vilanova. 2016. Tran-671scriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two 672close relatives of the common eggplant (*Solanum melongena*) with interest for breeding. *BMC Ge*-673*nomics* 17: 300, doi: 10.1186/s12864-016-2631-4.

675Holt, B. G., J.-P. Lessard, M. K. Borregaard, S. A. Fritz, M. B. Araujo, D. Dimitrov, P.-H. Fabre, et al. 6762013. An update of Wallace's zoogeographic regions of the world. *Science* 341: 343–343, doi : 67710.1126/science.1228282.

678

679Hopkins, J., and N. Maxted. 2010. Crop wild relatives: plant conservation for food security. Natural 680England, Peterborough, UK.

681

682Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. 683*Bioinformatics* 17: 754–755.

684

685Hurtado, M., S. Vilanova, M. Plazas, P. Gramazio, H. H. Fonseka, R. Fonseka, and J. Prohens. 2012. 686Diversity and relationships of eggplants from three geographically distant secondary centers of diver-687sity. *PLoS One* 7: e41748, doi: 10.1371/journal.pone.0041748.

688

689IUCN SSC Antelope Specialist Group. 2016. Aepyceros melampus. The IUCN Red List of Threatened 690Species 2016: e.T550A50180828. [online] Website http://dx.doi.org/10.2305/IUCN.UK.2016-2.RLT-691S.T550A50180828.en. [accessed 9 February 2018].

692

693Kadlec, M., D. U. Bellstedt, N. C. Le Maitre, and M. D. Pirie. 2017. Targeted NGS for species level 694phylogenomics: 'made to measure' or 'one size fits all'? *PeerJ* 5: e3569, doi: 10.7717/peerj.3569. 695

696Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Im-697provements in performance and usability. *Molecular Biology and Evolution* 30: 772–780, doi: 69810.1093/molbev/mst010.

699

700Kartzinel, T. R., P. A. Chen, T. C. Coverdale, D. L. Erickson, W. J. Kress, M. L. Kuzmina, D. I. 701Rubenstein, et al. 2015. DNA metabarcoding illuminates dietary niche partitioning be African large 702herbivores. *Proceedings of the National Academy of Sciences* 112: 8019–8024, doi: 10.1073/703pnas.1503283112.

704

705Kaushik, P., J. Prohens, S. Vilanova, P. Gramazio, and M. Plazas. 2016. Phenotyping of eggplant wild 706relatives and interspecific hybrids with conventional and phenomics descriptors provides insight for 707their potential utilization in breeding. *Frontiers in Plant Science* 7: 677, doi: https://doi.org/10.3389/708fpls.2016.00677.

710Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012.
711Geneious Basic: an integrated and extendable desktop software platform for the organization and ana712lysis of sequence data. *Bioinformatics* 28: 1647–1649, doi: 10.1093/bioinformatics/bts199.
713

714Knapp, S., and M. S. Vorontosova. 2013. From introduced American weed to Cape Verde Islands en-715demic – the case of *Solanum rigidum* Lam. (*Solanum*, subgenus *Leptostemonum*). *PhytoKeys* 25: 35– 71646, doi: 10.3897/phytokeys.25.4692.

717

718Knapp, S., M. S. Vorontsova, and J. Prohens. 2013. Wild relatives of the eggplant (*Solanum melon-*719*gena* L.: Solanaceae): New understanding of species names in a complex group. *PLoS One* 8: e57039, 720doi: 10.1371/journal.pone.0057039.

721

722Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis 723version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874, doi: 10.1093/mol-724bev/msw054.

725

726Lester, R. N., and S. M. Z. Hazan. 1991. Origin and domestication of the brinjal eggplant, *Solanum* 727*melongena*, from *Solanum incanum*, in Africa and Asia. *In* J. G. Hawkes, R. N. Lester, M. Nee, and N. 728Estrada [eds], Solanaceae III: taxonomy, chemistry and evolution, 369–387. Royal Botanic Gardens 729Kew, Kew, UK.

730

731Linder, H. P., H. M. de Klerk, J. Born, N. D. Burgess, J. Fjeldså, and C. Rahbek. 2012. The partition-732ing of Africa: Statistically defined biogeographical regions in sub-Saharan Africa. *Journal of Biogeo-*733*graphy*. 39: 1189–1205, doi: 10.1111/j.1365-2699.2012.02728.x.

734

735Melamed, Y., M. E. Kislev, E. Geffen, S. Lev-Yadun, and N. Goren-Inbar. 2016. The plant component 736of an Acheulian diet at Gesher Benot Ya'aqov, Israel. *Proceedings of the National Academy of Sci*-737*ences* 113: 14674–14679, doi: 10.1073/pnas.1607872113.

738

739Meyer, R. S., K. G. Karol, D. P. Little, M. H. Nee, and A. Litt. 2012. Phylogeographic relationships
740among Asian eggplants and new perspectives on eggplant domestication. *Molecular Phylogenetics*741*and Evolution* 63: 685–701, doi: 10.1016/j.ympev.2012.02.006.
742

743Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES science gateway for inference 744of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 7451–8, New Orleans, Louisiana, USA.

746

747Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genome-scale data to re-748solve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sci*-749*ences* 104: 19363–19368, doi: 10.1073/pnas.0708072104.

750

751Muñoz-Falcón, J. E., J. Prohens, S. Vilanova, and F. Nuez. 2009. Diversity in commercial varieties 752and landraces of black eggplants and implications for broadening the breeders' gene pool. *Annals of* 753*Applied Biology* 154: 453–465, doi: 10.1111/j.1744-7348.2009.00314.x.

754

755Mutegi, E., A. A. Snow, M. Rajkumar, R. Pasquet, H. Ponniah, M.-C. Daunay, and P. Davidar. 2015. 756Genetic diversity and population structure of wild/weedy eggplant (*Solanum insanum*, Solanaceae) in 757southern India: Implication for conservation. *American Journal of Botany* 102: 139–147, doi: 10.3732/ 758ajb.1400403.

759

760Nylander, J. A. A. 2004. MrModeltest v2. Computer program and documentation distributed by the 761author, website: https://github.com/nylander/MrModeltest2. [accessed 9 February 2018].

762

763Olet, E. A., and R. Bukenya-Ziraba. 2001. Variation within the *Solanum incanum* complex in Uganda 764and its relationship with *Solanum cerasiferum*. *In* R. G. van den Berg, G. W. M. Barendse, G. M. van 765der Weerden, and C. Mariani [eds], Solanaceae V: Advances in Taxonomy and Utilization, 97–108. 766Nijmegen University Press, Nijmegen, Netherlands.

767

768Pautasso, M., T. F. Döring, M. Garbelotto, L. Pellis, and M. J. Jeger. 2012. Impacts of climate change 769on plant diseases – opinions and trends. *European Journal of Plant Pathology* 133: 295–313, doi: 77010.1007/s10658-012-9936-1.

771

772Pease, J. B., D. C. Haak, M. W. Hahn, and L. C. Moyle. 2016. Phylogenomics reveals three sources of 773adaptive variation during a rapid radiation. *PLoS Biology* 14: e1002379, doi:10.1371/journal. 774pbio.1002379.

776Peng, Y., H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin. 2012. IDBA-UD: A de novo assembler for 777single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 28: 1420–7781408, doi: 10.1093/bioinformatics/bts174.

779

780Pringle, R. M., J. R. Goheen, T. M. Palmer, G. K. Charles, E. DeFranco, R. Hohbein, A. T. Ford, et al. 7812014. Low functional redundancy among mammalian browsers in regulating an encroaching shrub 782(*Solanum campylacanthum*) in African savannah. *Proceedings of the Royal Society B, Biological Sci*-783*ences* 281: 20140390, doi: 10.1098/rspb.2014.0390.

784

785Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6. Computer program and 786documentation distributed by the author, website: http://beast.community/tracer. [accessed 9 February 7872018].

788

789Ranil, R. H. G., J. Prohens, X. Aubriot, H. M. L. Niran, M. Plazas, R. M. Fonseka, S. Vilanova, et al. 7902016. *Solanum insanum* L. (subgenus *Leptostemonum* Bitter, Solanaceae), the neglected wild progeni-791tor of eggplant (*S. melongena* L): a review of taxonomy, characteristics and uses aimed at its enhance-792ment for improved eggplant breeding. *Genetic Resources and Crop Evolution* 64: 1707–1722, doi: 79310.1007/s10722-016-0467-z.

794

795Ree, R. H., and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by 796dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14. 797

798Ripple, W. J., T. M. Newsome, C. Wolf, R. Dirzo, K. T. Everatt, M. Galetti, M. W. Hayward, et al. 7992015. Collapse of the world's largest herbivores. *Science Advances* 1: e1400103. doi: 10.1126/800sciadv.1400103.

801

802Särkinen, T., L. Bohs, R. G. Olmstead, and S. Knapp. 2013. A phylogenetic framework for evolution-803ary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* 13: 214, 804doi: 10.1186/1471-2148-13-214.

805

806Shi, C., N. Hu, H. Huang, J. Gao, Y.-J. Zhao, and L.-Z. Gao. 2012. An improved chloroplast DNA ex-807traction procedure for whole plastid genome sequencing. *PLoS One* 7: e31468, doi: 10.1371/jour-808nal.pone.0031468.

810Solanaceae Source Database. 2017. Solanaceae Source: a global taxonomic resource for the night-811shades [online]. Website http://www.solanaceaesource.org/ [accessed 9 February 2018].

812

813Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large 814phylogenies. *Bioinformatics* 30: 1312–1313, doi: 10.1093/bioinformatics/btu033.

815

816Stamatakis, A. 2016. The RAxML v8.2.X Manual. [online] Website https://sco.h-its.org/exelixis/re-817source/download/NewManual.pdf. [accessed 9 February 2018].

818

819Stern, S. R., M. de F. Agra, and L. Bohs. 2011. Molecular delimitation of clades within New World 820species of the "spiny solanums" (*Solanum* subgenus *Leptostemonum*). *Taxon* 60: 1429–1441. 821

822Syfert, M. M., N. P. Castañeda-Álvarez, C. K. Khoury, T. Särkinen, C. C. Sosa, H. A. Achicanoy, V. 823Bernau, et al. 2016. Crop wild relatives of the brinjal eggplant (*Solanum melongena*; Solanaceae): 824poorly represented in genebanks and many species at risk of extinction. *American Journal of Botany* 825103: 635–651, doi:10.3732/ajb.1500539.

826

827Thiers, B. 2017. Index Herbariorum: A global directory of public herbaria and associated 828staff. [online] Website http://sweetgum.nybg.org/science/ih/. [accessed 9 February 2018]. 829

830Tonti-Filippini, J., P. G. Nevill, K. Dixon, and I. Small. 2017. What can we do with 1000 plastid gen-

831omes? Plant Journal 90: 808–818, doi: 10.1111/tpj.13491.

832

833Uno, K. T., T. E. Cerling, J. M. Harris, Y. Kunimatsu, M. G. Leakey, M. Nakatsukasa, and H. Nakaya. 8342011. Late Miocene to Pliocene carbon isotope record of differential diet change among East African 835herbivores. *Proceedings of the National Academy of Sciences* 108: 6509–6514, doi: 10.1073/ 836pnas.1018435108.

837

838Uno, K. T., P. J. Polissar, K. E. Jackson, and P. B. deMenocal. 2016. Neogene biomarker record of ve-839getation change in eastern Africa. *Proceedings of the National Academy of Sciences* 113: 6355–6363, 840doi: 10.1073/pnas.1521267113.

841

842van Loo, M., W. Hintsteiner, E. Pötzelsberger, S. Schüler, and H. Hasenauer. 2015. Intervarietal and 843intravarietal genetic structure in Douglas-fir: nuclear SSRs bring novel insights into past population

844demographic processes, phylogeography, and intervarietal hybridization. *Ecology and Evolution* 5: 8451802–1817, doi: 10.1002/ece3.1435.

846

847Vilanova, S., M. Blasco, M. Hurtado, J. E. Muñoz-Falcón, J. Prohens, and F. Nuez. 2010. Develop-848ment of a linkage map of eggplant based on a *S. incanum* x *S. melongena* backcross generation. *In* J. 849Prohens, and A. Rodriguez-Burruezo [eds] Advances in genetics and breeding of capsicum and egg-850plant, 435–439. Editoral de la Universitat Politècnica de València, Valencia, Spain.

851

852Vorontsova, M. S., S. Stern, S. Bohs, and S. Knapp. 2013. African spiny *Solanum* (subgenus *Leptoste*-853*monum*, Solanaceae): a thorny phylogenetic tangle. *Botanical Journal of the Linnean Society* 173: 854176–193, doi:10.1111/boj.12053.

855

856Vorontsova, M. S., and S. Knapp. 2016. A revision of the spiny solanums, *Solanum* subgenus *Lep*-857*tostemonum* (Solanaceae) in Africa and Madagascar. *Systematic Botany Monographs* 99: 1–432.

858

859Wang, J. X., T. G. Gao, and S. Knapp. 2008. Ancient Chinese literature reveals pathways of eggplant 860domestication. *Annals of Botany* 102: 891–897.

861

862Weese, T. L., and L. Bohs. 2010. Eggplant origins: out of Africa, into the Orient. *Taxon* 59: 49–56. 863

864Williams, A. V., J. T. Miller, I. Small, P. G. Nevill, and L. M. Boykin. 2016. Integration of complete 865chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Aca*-866*cia*. *Molecular Phylogenetics and Evolution* 96: 1–8, doi: 10.1016/j.ympev.2015.11.021. 867

868^{White, F. 1983.} The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AET-869^{FAT/UNSO} vegetation map of Africa. UNESCO, Paris, France.

870

871White, F. 1993. The AETFAT chorological classification of Africa: history, methods and applications.872Bulletin du Jardin Botanique National de Belgique 62: 225–281.

873

874Wysocki, W. P., L. G. Clark, L. Attigala, E. Ruiz-Sanchez, and M. R. Duvall. 2015. Evolution of the 875bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC Evolutionary Bio*-876*logy* 15: 50, doi: 10.1186/s12862-015-0321-5.

878Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using Bruijin 879graphs. *Genome Research* 18: 821–829, doi: 10.1101/gr.074492.107. 880

881Zhang, M.-Y., P. W. Fritsch, P.-F. Ma, H. Wang, L. Lu, D.-Z. Li. 2017. Plastid phylogenomics and 882adaptive evolution of *Gaultheria* series *Trichophyllae* (Ericaceae), a clade from sky islands of the Hi-883malaya-Hengduan Mountains. *Molecular Phylogenetics and Evolution* 110: 7–18.

884**TABLE 1** Summary of GenBank accession numbers, species name, collection information and plastome composition and coverage for taxa used in the 885NextGen analyses. Herbarium codes in column 4 follow Index Herbariorum (Thiers, 2017). Abbreviation in columns 7 to 10 are as follows: herb. = 886herbarium material; sil. = silica gel-dried material; LSC = large single copy region; SSC = Small single copy region; IRs = inverted repeat regions. **[In-887formation intended for the review process only:** The sequences generated for the present study are in the process of being submitted to GenBank. 888Once the GenBank accession numbers for all these sequences have been generated, we will update the table].

GenBank Mean Mate-Total Name in topologies Species name Voucher Collection location LSC SSC IRs accescoverial length sions rage Vorontsova et al. 18,56 155,608 86,219 Solanum aethiopicum Tanzania, Morogoro, Tegetero sil. 25,413 Х Solanum aethiopicum L. 152.5 3 156 (BM) Solanum aqnewiorum Vo-Vorontsova et al. 18,58 7 Solanum agnewiorum Kenya, Eastern, Nyambene 155,586 86,163 25,418 19.5 Х herb. ronts. 195 (BM) Madagascar, Fianarantsoa, Vorontsova et al. 18,55 155,577 86,188 25,415 Х Solanum anguivi Solanum anquivi Lam. sil. 124.6 1302 (K) a Itremo Massif Chase N.C. 5271 Solanum aureitomen-Solanum aureitomentosum Zimbabwe, Southern Rhodesia, 18,49 7 155,303 85,950 25.428 Х herb. 38 tosum Bitter (BM) Odzi Solanum campylacan-Solanum campylacanthum Vorontsova et al. 18,50 Tanzania, Tanga, Mkata Kideleko 155,017 85,667 25.422 54.5 sil. Х thum Hochst. ex A.Rich. 158 (BM) Central African Republic. 18,51 25,384 Х Solanum cerasiferum Solanum cerasiferum Dunal Fay 2715 (K) herb. 155,413 86,129 17.2 6 Bamingui-Bangoran, Bozobo Solanum dasyphyllum Schu-Vorontsova et al. Solanum dasyphyllum Tanzania, Morogoro, Tegetero 155,716 86,314 25,433 62.8 Х sil. 6 mach. & Thonn. 151 (BM) Miyazaki 606I28 18,51 Solanum qlabratum Dunal Saudi Arabia, Asir, Abah 155,368 86,081 25,385 20.2 Х Solanum glabratum herb. (A) Vorontsova et al. 18,53 9 25,421 Х Solanum incanum 1 Solanum incanum L. Kenya, Eastern, Laisamis herb. 155,657 86,276 105.3 203 (BM) Sanou & Traore 18,50 8 Solanum incanum 2 155,617 86,269 25,420 92.5 Х Solanum incanum L. Burkina Faso, Houet, Saguere herb. BUR48 (K) Sampath Kumar et 18,49 8 Х Solanum insanum Solanum insanum L. India, Tamil Nadu, Thamaraikulam sil. 155,570 86,240 25,416 254.8 al. 941 (MH) Solanum lanzae Vorontsova et al. 18,19 Kenya, Rift Valley, Kifuko Ranch 25,409 Х Solanum lanzae sil. 155,269 86,253 175 783 (K) J.-P.Lebrun & Stork

х	Solanum lichtensteinii	Solanum lichtensteinii Willd.	Crawford et al. FC139 (K)	Namibia, Kunene, Outjo	herb.	155,574	86,217	18,51 5	25,421	30.3
х	Solanum linnaeanum	Solanum linnaeanum Hepper & PM.L. Jaeger	Jury et al. 13209 (BM)	Morocco, Rabat-Salé-Kénitra, Mou- lay Bousselham	herb.	155,574	86,219	18,50 9	25,423	73.5
х	Solanum macrocarpon	Solanum macrocarpon L.	Tepe et al. 2770 (BM)	Tanzania, Morogoro, Ruvu	sil.	155,820	86,398	18,55 4	25,434	55.5
х	Solanum melongena 1	Solanum melongena L.	Meeboonya et al. RM294 (BM)	Thailand, Chumphon, Pang Wan	sil.	155,583	86,250	18,50 1	25,416	44.6
х	Solanum melongena 2	Solanum melongena L.	Poczai H07653 (H)	Finland, Helsinki, Kumpula Botan- ical Garden	sil.	155,569	86,236	18,50 1	25,416	800
х	Solanum polhillii	Solanum polhillii Voronts.	Vorontsova et al. 15 (BM)	Kenya, Rift Valley Mt. Suswa	herb.	155,422	86,044	18,54 4	25,417	196.9
х	Solanum richardii	Solanum richardii Dunal	Vorontsova et al. 633 (K)	Madagascar, Fianarantsoa, Van- gaindrano	sil.	155,187	86,113	18,24 4	25,415	118
х	Solanum rigidum	Solanum rigidum Lam.	Malato-Beliz 141 (LISC)	Cape Verde, Maio	herb.	155,594	86,244	18,51 0	25,420	10.2
х	Solanum supinum	Solanum supinum Dunal	Kabelo 170 (K)	Botswana, Ghanzi, Okwa Valley	herb.	155,769	86,362	18,54 9	25,429	111.9
х	Solanum trilobatum	Solanum trilobatum L.	Meeboonya et al. RM243 (BM)	Thailand, Samut Songkhram, Bueng Yi San	sil.	155,289	86,218	18,24 5	25,413	95.8
х	Solanum umtuma	<i>Solanum umtuma</i> Voronts. & S.Knapp	Nevhutalu 921 (K)	South Africa, KwaZulu-Natal, Swart Umfolozi	herb.	155,541	86,173	18,51 4	25,427	29.5
х	Solanum usambarense	Solanum usambarense Bitter & Dammer	Vorontsova et al. 166 (BM)	Tanzania, Tanga, Magamba Forest Reserve	sil.	155,411	86,028	18,55 1	25,416	42.3
x	Solanum violaceum	Solanum violaceum Ortega	Sampath Kumar et al. 945e (MH)	India, Tamil Nadu, Palar Dam	sil.	155,748	86,309	18,60 9	25,415	147.2

890APPENDIX

891Summary of species, country, voucher (collector and number), herbarium acronym (codes follow 892Thiers 2017) and GenBank accession numbers for taxa used in the preliminary Sanger study, provided 893in the following order: ITS, *waxy* and *ndhF-rpL32*. Dashed lines indicate that the region was not 894sampled for this accession. Newly generated sequences are indicated with an asterisk following the ac-895cession number. [**Information intended for the review process only:** GenBank numbers for se-896quences already published are provided. The sequences generated for the present study are in the pro-897cess of being submitted to GenBank and we have used our laboratory extraction ID (XA plus three di-898gits) to refer to them. Once the GenBank accession numbers for all these sequences have been gener-899ated, we will update the table where appropriate].

900

901Solanum aethiopicum L., Tanzania, Vorontsova et al. 156 (BM), XAS232*, -, XAS232*. S. ae-902thiopicum L., Olmstead S-74 (WTU), -, AY996378, -. S. agnewiorum Voronts., Kenya, Vorontsova 903et al. 195 (BM), KU719948, KU719870, KU720049. S. anquivi Lam., Madagascar, Vorontsova et al. 9041302 (K), XAS218*, -, XAS218*. S. anguivi Lam., Bohs 3266 (UT), -, HQ721963, -. S. aureito-905mentosum Bitter, Zimbabwe, Chase N.C. 5271 (BM), KU719949, KU719871, XAS298*. S. campyla-906canthum Hochst. ex A.Rich., Tanzania, Vorontsova et al. 158 (BM), XAS220*, XAS220*, 907KU719997. S. campylacanthum Hochst. ex A.Rich., Ethiopia, Friis 8107 (K), XAS290*, XAS290*, 908XAS290*. S. campylacanthum Hochst. ex A.Rich., Ethiopia, Friis et al. 10427 (K), XAS313*, 909XAS313*, XAS313*. S. campylacanthum Hochst. ex A.Rich., Ethiopia, Friis 11156 (K), XAS314*, 910XAS314*, XAS314*. S. campylacanthum Hochst. ex A.Rich., South Africa, Westfall 1661 (K), 911XAS317*, XAS317*, XAS317*. S. campylacanthum Hochst. ex A.Rich., Namibia, Kolberg & Ku-912birske 600 (K), XAS318*, XAS318*, XAS318*. S. cerasiferum Dunal, Ethiopia, Friis et al. 7733 913(K), XAS291*, XAS291*, XAS291*. S. cerasiferum Dunal, Ethiopia, Friis et al. 12670 (K), 914XAS315*, XAS315*, XAS315*. S. cerasiferum Dunal, Central African Republic, Fay 2715 (K), 915XAS322*, XAS322*, XAS322*. S. dasyphyllum Schumach. & Thonn., Tanzania, Vorontsova et al. 916151 (BM), XAS223*, -, KU720000. S. dasyphyllum Schumach. & Thonn., Cipollini 7 (UT), -, 917AY996406, -. S. glabratum Dunal, South Arabia, Miyazaki 606128 (A), KU719913, KU719834, 918KU719994. S. incanum L., Kenya, Vorontsova et al. 203 (BM), XAS309*, XAS309*, XAS309*. S. 919incanum L., Burkina Faso, Sanou & Traore BUR48 (K), XAS325*, XAS325*, XAS325*. S. in-920canum L., Oman, Mandaville 6523 (BM), XAS327*, XAS327*, XAS327*. S. incanum L., Pakistan, 921Popov 30 (BM), XAS328*, XAS328*, XAS328*. S. insanum L., Madagascar, Vorontsova et al. 644 922(BM), KU719914, KU719835, KU720003. S. insanum L., Wang 2052 (PE), KU719884, KU719813, 923KU719964. S. insanum L., India, Sampath Kumar et al. 941 (MH), KU719940, KU719862, 924KU720041. S. lanzae J.-P.Lebrun & Stork, Kenya, Vorontsova et al. 783 (BM), KU719921, 925KU719842, KU720015. S. lichtensteinii Willd., Namibia, *Crawford et al.* 139 (K), KU719947, 926KU719869, KU720048. S. linnaeanum Hepper & P.-M.L. Jaeger, Morocco, Jury et al. 13209 (BM), 927KU719950, KU719872, KU720050. S. macrocarpon L., Tanzania, *Tepe et al.* 2770 (BM), XAS227*, 928–, KU720004. S. macrocarpon L., Olmstead S-88 (WTU), –, AY996436, –. S. melongena L., Thail-929and, Meeboonya et al. RM294 (BM), XAS343*, –, XAS343*. S. melongena L., Olmstead S-91 930(WTU), –, AY562959, –. S. nigriviolaceum Bitter, Kenya, Vorontsova et al. 57 (BM), HQ721868, 931HQ721993, –. S. polhillii Voronts., Kenya, Vorontsova et al. 15 (BM), XAS308*, XAS308*, 932XAS308*. S. richardii Dunal, Madagascar, Vorontsova et al. 633 (BM), KU719915, KU719836, 933KU720006. S. rigidum Lam., Cape Verde, Malato-Beliz 141 (LISC), XAS337*, –, XAS337*. S. su-934pinum Dunal, Botswana, Kabelo 170 (K), XAS335*, XAS335*, XAS335*, S. trilobatum L., Thail-935and, Meeboonya et al. RM243 (BM), XAS130*, XAS130*, XAS130*, S. umtuma Voronts. & 936S.Knapp, South Africa, Nevhutalu 921 (K), XAS34*, XAS34*, XAS334*, S. usambarense Bitter & 937Dammer, Tanzania, Vorontsova et al. 166 (BM), KU719923, KU719844, KU720023. S. violaceum 938Ortega, India, Sampath Kumar et al. 945e (MH), KU719941, KU719863, KU720042.

940FIGURE LEGENDS

941FIGURE 1 Phylogeny and biogeography of the Eggplant clade based on whole chloroplast genome 942sequences. (A) Map showing the seven biogeographic areas used to infer the biogeographic history of 943the Eggplant clade. Dotted lines indicate instances of complex biome transitions. Arrows illustrate the 944dispersion events inferred from the biogeographic analysis. (B) Full-plastome dated phylogeny of the 945Eggplant Clade (consensus of 4 BEAST analyses; 159,227 bp matrix). All nodes have $PP \ge 0.97$ and $946BS \ge 87\%$ except for the nodes designated with *. Green bars correspond to 95% HDP intervals for 947the dates of the nodes (Myr). Geographical distribution for each terminal is indicated using the biogeo-948 graphic regions subdivision. The most probable ancestral area is figured at each node of the Eggplant 949clade; high level of biogeographic uncertainty is indicated with dotted lines. Pie-charts with relative 950probabilities of ancestral states at each node are provided in Appendix S4a,b. Hypothesised dispersion 951(*D*) and extinction (*E*) events are figured above the branches of the tree. Within the Eggplant clade, 952clades are given provisional names based on their distributions. Names of cultivated species are in 953bold. Pictures from top to bottom and left to right are fruits of the cultivated eggplant (Solanum mel-954ongena), S. incanum, S. linnaeanum, S. insanum (eggplant wild relative) and S. agnewiorum. Photo-955graph credits: (S. melongena, S. insanum) X. Aubriot; (S. incanum, S. agnewiorum) M.S. Vorontsova; 956(*S. linnaeanum*) S. Knapp.

957

958**FIGURE 2** Overlap between the distribution of the widespread Eggplant clade species in Africa and 959two 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 961Fig. 1B for more details on this clade). The current distributions of the African elephant and impala 962are also displayed on this map. **(B)** Venn diagram displaying the overlap of distributions between pop-963ulations of *S. campylacanthum* + *S. cerasiferum* + *S. incanum* and the African elephant and impala. 964Overall, the two animals cover 53.7% of the distribution of the three *Solanum* species.

966SUPPLEMENTAL DATA LEGENDS

APPENDIX S1 50% majority-rule trees from the separate and combined Bayesian analyses of the 968ITS, *waxy* and *ndhF-rpL32* datasets.

APPENDIX S2 Chloroplast genome maps of the 25 accessions sequenced; a description of the 970general plastid structure is also supplied.

APPENDIX S3 50% majority-rule trees from the RAxML, MrBayes and BEAST analyses of the full-972plastome dataset.

APPENDIX S4 Biogeographical scenarios for the Eggplant clade inferred by Lagrange plotted onto 974the BEAST maximum clade credibility tree; probabilities of splits for each node are also provided.