Phylogenetic Relationship among Eggplant Solanum L. and Related Species in Southern Nigeria as Revealed by Nuclear and Chloroplast Genes

M.O. Sifau, L.A. Ogunkanmi, K.O. Adekoya, B.O. Oboh and O.T. Ogundipe

1Department of Cell Biology and Genetics, 2Department of Botany, University of Lagos, Lagos, Nigeria 3Laboratory of Molecular Biology, Biotechnology Unit, National Centre for Genetic Resources and Biotechnology (NACGRAB), PMB 5382, Ibadan, Nigeria

Abstract: The common name “Eggplant” is given to vegetables of several Solanum L. species that are important for human diet and health. The taxonomy and phylogenetic relationships among these taxa are currently unclear due to its large size and tropical center of diversity. This study assessed the genetic diversity in forty nine eggplant and related Solanum species’ genetic resources in Southern Nigeria using one nuclear (ITS) and two chloroplast genes. Analysis of DNA sequence data from the three regions (ITS, trnl C-trnL D and trnL E-trnF) gave a high level of genetic variability (polymorphism) among the samples studied. A bootstrap value of 100 was observed between S. macrocarpon L. and S. torvum Sw. and between S. macranthus A. Rich. and S. indicum L., closely followed by a value of 99 between S. aethiopicum L. and S. dasypHYLLUM Schum. and Thonn. All these are indications of a close relationship between these species and a possibility of a common ancestor is strongly proposed. Solanum torvum was often separated out from S. melongena and even where they grouped together they have a low bootstrap value of 3 which is an indication of distant relatedness. The high level of intra and inter specific variations displayed within eggplant accessions and between its relatives as reported in this study could be effectively used in genetic improvement of cultivated eggplant varieties as well as in situ and ex situ conservation.

Key words: Eggplant, variability, ITS, chloroplast, polymorphism, Nigeria

INTRODUCTION

Solanum L. is the largest genus of the family Solanaceae with over 2,000 species. Members of the genus are as varied morphologically as they are diverse in number and distribution ecogeographically. It includes perennial shrubs; vines, herbs, or trees, with or without spines, glabrous or pubescent with unbranched or branched, often glandular hairs (Mueller et al., 2005). The common name “Eggplant” is ascribed to vegetable Solanum which encompasses three closely related cultivated species that belong to Solanum subgenus Leptostemonum: Solanum melongena, S. aethiopicum and S. macrocarpon (Daunay et al., 2001; Doganlar et al., 2002). The name eggplant comes from the shape (egg-like) of the vegetable’s fruit (Doganlar et al., 2002).

Despite the importance of the genus Solanum, phylogenetic relationships among the taxa are currently unclear (Daunay and Lester, 1989). Solanum taxonomy has been complicated largely due to species large number, overlapping ecogeographical distribution (Levin et al., 2005), morphological plasticity, similarity of genomes (Okoli, 1988) and existence of swarms of natural hybrids (Obute et al., 2006; Oyelana and Ugborogh, 2008). The inconsistencies and misconceptions generated by these factors have made past attempts at taxonomically resolving the complexities associated with the genus difficult (Knapp et al., 2004; Levin et al., 2005). However, Sifau et al. (2014) has provided some species database of the vegetable Solanum and related species in Southwestern Nigeria with emphasis on variation patterns.

Several projects have focused on phylogenetic relationships in the entire Solanum genus and in some of its component clades. Levin et al. (2006) stated that many of the subgenera and sections within Solanum are not yet valid; they are still presently used provisionally as the phylogeny of this genus is not fully resolved yet and many species have not been re-evaluated. Cladistic analyses of DNA sequence data suggest that the present subdivisions and rankings are largely invalid. Far more subgenera would seem to warrant recognition with
subgenus *Leptostemonium* being the only one that can at present be clearly subdivided into sections, having 10
clades and 21 sections (Levin et al., 2006). Traditional
taxonomists have recognized three subfamilies within the
Solanaceae namely: Solanooideae, Nolanoideae and
Cestroideae Levin et al. (2006). However Hunziker (1979)
excluded the Nolanoideae (i.e., genus *Nolana*) from the
Solanaceae and expanded the number of subfamilies to six, namely, Solanooideae, Cestroideae, Juaulilloideae, Salpiglossoideae, Schizanthoideae and Anthocerecidioideae.

Subsequent molecular data from chloroplast ndhF
sequences identified about 13 major clades within
*Solanum* (Bohs and Olmstead, 1997; Bohs, 2005). An
overall phylogeny of *Solanum* using molecular data
from three nuclear and chloroplast genes showed that
few of these subgenera comprise monophyletic groups
(i.e., with common ancestors); but indicated that
*Solanum* may be comprised of about 12-15 major clades.
These have been given informal names and are
themselves the subject of more detailed phylogenetic
studies (Weese and Bohs, 2007). Molecular data and
phylogenetic analyses thus challenged the traditional
view of Solanaceae subfamilies (Olmstead and Palmer,
1992; Olmstead and Sweere, 1994; Fay et al., 1998) but the
precise number of monophyletic groups in the family and
their names and circumscription are still under
investigation.

Most previous studies provide information about
phylogenetic relationships using single-locus data
(Bohs and Olmstead, 1997). Such single-locus-based
species trees contained some errors (Takahata and
Nei, 1985). Substantial number of erroneous trees can be
avoided by using multilocus methods or Restriction
Fragment Length Polymorphism (RFLP). Other studies
used DNA sequence data from nuclear regions such as
ITS and granule-bound starch synthase gene (GBSSI or
waxy) or chloroplast regions (trnF-trnL and trnS-trnG) or
combinations of these data (Levin et al., 2005, 2006).
These methods provide species trees based on the
phylogenetically relevant information contained in many
loci in the whole genome (Hampl et al., 2001).

The objective of this study was to use multilocus
methods to establish phylogenetic relationships among
eggplant *Solanum* and related species in Southern Nigeria
using sequence data obtained from one nuclear (ITS) and
two chloroplast gene regions trnL-trnF and trnE-trnF.
This involved the amplification and sequencing of the
Internal Transcribed Spacer (ITS) region of nuclear
ribosomal DNA, composed of ITS1, the 5.8S gene and
ITS2 and the non-coding chloroplast DNA (cpDNA)
regions trnE-trnF spacer plus trn intron from the
extracted total genomic DNA.

**MATERIALS AND METHODS**

**Sample collection and identification:** Samples of eggplant
*Solanum* and related species were collected in different
locations within Southern part of Nigeria, especially in
areas known for eggplant endemism and diversity.

Voucher specimens were prepared from the samples
following the method of Ogundipe et al. (2008) and sent
to Forestry Herbarium Ibadan (FHI) where they were
authenticated (Table 1). These specimens were then
deposited at both the University of Lagos Herbarium
(LUH) and Forestry Herbarium Ibadan (FHI) for reference
purposes.

**Total genomic DNA extraction and amplification:** Total
genomic DNA extraction was carried out on young fresh

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Sample name</th>
<th>Site of collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR46ndi</td>
<td><em>S. indicum</em> subsp. <em>distichum</em> var. <em>distichum</em> Schumann and Thonn</td>
<td>Ogbang village, Cross river</td>
</tr>
<tr>
<td>CR47melon</td>
<td><em>S. melongena</em> L.</td>
<td>Ogbang village, Cross river</td>
</tr>
<tr>
<td>CR36ndi</td>
<td><em>S. indicum</em> subsp. <em>distichum</em> var. <em>distichum</em> Schumann and Thonn</td>
<td>Edmond village, Obubra, Cross river</td>
</tr>
<tr>
<td>CR50macra</td>
<td><em>S. macrocarpon</em> Lam.</td>
<td>Edibo-Holyen junction, Off marian road, calabar, Cross river</td>
</tr>
<tr>
<td>OS21torv</td>
<td><em>S. torvum</em> Sw.</td>
<td>Iwo, Osun</td>
</tr>
<tr>
<td>ED31melon</td>
<td><em>S. melongena</em> L.</td>
<td>Ugo town, Edo</td>
</tr>
<tr>
<td>ED22macra</td>
<td><em>S. macrocarpon</em> Lam.</td>
<td>Ugo town, Edo</td>
</tr>
<tr>
<td>OY29torv</td>
<td><em>S. torvum</em> Sw.</td>
<td>Alakia Ibadan, Oyo</td>
</tr>
<tr>
<td>ED30melon</td>
<td><em>S. melongena</em> L.</td>
<td>Ugo town, Edo</td>
</tr>
<tr>
<td>OY27melon</td>
<td><em>S. melongena</em> L.</td>
<td>Alakia Ibadan, Oyo</td>
</tr>
<tr>
<td>CR37torv</td>
<td><em>S. torvum</em> Sw.</td>
<td>Okonmu, Akamkpa (Calabar-Born road), Cross river</td>
</tr>
<tr>
<td>CR59torv</td>
<td><em>S. torvum</em> Sw.</td>
<td>Edonan Village, Obubra, Cross river</td>
</tr>
<tr>
<td>OY17adath</td>
<td><em>S. aethiopicum</em> L.</td>
<td>Ighobe, Oyo</td>
</tr>
<tr>
<td>OY19melon</td>
<td><em>S. melongena</em> L.</td>
<td>Ighobe, Oyo</td>
</tr>
<tr>
<td>OG04dasy</td>
<td><em>S. dysphylidium</em> Schum. and Thonn</td>
<td>Joga orise, Ogun</td>
</tr>
<tr>
<td>OY18bescab</td>
<td><em>S. scaratum</em> L.</td>
<td>Ighobe, Oyo</td>
</tr>
<tr>
<td>LA24igole</td>
<td><em>S. gilo</em> Raddii</td>
<td>Agbowo-Rossi, Lagos</td>
</tr>
</tbody>
</table>
leaves dried with silica gel and kept in sample bags, using the modified Cetyltrimethylammonium bromide (CTAB) extraction protocol (Doyle and Doyle, 1990) followed by additional purification done in a silica-column inserted into vacuum manifold connected to a vacuum pump using QIA quick purification kit. Verification of the quality of the purified DNA samples was achieved by electrophoresis on a 1% agarose gel.

Polymerase Chain Reaction (PCR) involved the amplification of the Internal Transcribed Spacer (ITS) region of nuclear ribosomal DNA and the non-coding chloroplast DNA (cpDNA) regions trnL-trnF spacer plus trnL intron from the extracted total genomic DNA. A positive (a known sample which has worked previously for the region being amplified) and negative (PCR mix but no DNA added) were used as a control for contamination. A master mix including every cocktail except the template DNA (of each sample) was made up in a 1.5 mL eppendorf tube, allowing for one extra sample.

The Internal Transcribed Spacer (ITS) region of nuclear ribosomal DNA, composed of ITS1, the 5.8S gene and ITS2, was amplified using two primers ITS1eul (5′-GTC CAC TGA ACC TTA CTA TTG AG-3′) and ITS4 (5′-TCC GCC GCT TAT GGA TAT GC-3′) as described by White et al. (1990) and Baldwin et al. (1995). The fragment size amplified was between 1236-1280 bp. The following PCR profile was used: 94°C for 2 min, 72°C for 3 min (initial strand separation); 30 cycles at 94°C for 1 min (Denaturation), 50°C for 1 min (annealing), 72°C for 1 min and 30s (primer extension); with a final extension at 72°C for 7 min. PCR products were cleaned using the QIAquick PCR purification kit (Qiagen, Valencia, California, USA).

The non-coding cpDNA regions trnL-trnF spacer plus trnL intron are adjacent in the chloroplast genome but were amplified separately. TrnL includes region between trnL-trnC and trnL includes region between trnL-trnF. Primers used for amplification were UniC (5′-CGAAAATCGGTAGACGCTACG) and UniF (5′-ATTTGAACTGTTGACACGAG) of Taberlet et al. (1991). The fragment size amplified was between 750-800 bp. The thermal cycler program included initial denaturing at 94°C for 3 min; 28 cycles at 94°C for 1 min (Denaturation), 48°C for 1 min (annealing), 72°C for 1 min (primer extension) ending with a final extension at 72°C for 7 min and final hold at 4°C. The last cycle in both cases was followed by 5 min extension at 72°C to allow complete extension of the PCR products with a final hold at 4°C till electrophoresis.

During electrophoresis, 5 µL of each of PCR product were mixed with 3 µL of 10X loading dye (0.25% bromophenol blue, 0.25% xylene cyanol and 40% sucrose, w/v) and spun briefly in a micro centrifuge before loading on a 1.5% agarose gel which has been previously stained with safe view. This was run for 1½ h at 110 mA. The gel was viewed and photographed (with the aid of eyeproctor) using a Gel Documentation and Analysis Systems (UVdoc, GA-9000/9010 Version 12).

**Cycle sequencing of PCR product:** Cycle sequencing was to increase the sensitivity of the DNA sequencing process and permits the use of very small amounts of DNA starting material. Amplification of selected regions was achieved in a 10 µL reaction mixtures containing 0.5 µL pink juice (Big Dye Terminator, Applied Biosystems Inc.), 3.0 µL 5X sequencing buffer (Bioline), 0.75 µL primer (1:10 dilution; forward or reverse for each primer pair) and varying quantities of purified PCR products depending on the strength of the reaction, made up to 10 µL with sterile water. The amplification of ITS region was improved by the addition of 4% Dimethyl Sulfoxide (DMSO) in the total volume of the sequencing mix. Cycle sequencing was done in a Gene Amp® PCR System 9700 Thermocycler (Applied Biosystems Inc.) using the following programme: Initial denaturation at 95°C for 30 sec followed by one cycle of denaturation at 95°C for 60 sec, annealing at 55°C for 30 sec and extension at 72°C for 60 sec; ran for 30 cycles and thereafter held the reaction mixture for 7 min at 72°C to allow complete extension of the PCR products with a final hold at 4°C.

Circle sequencing products were then purified using the QIAquick PCR purification kit and thereafter sequenced on an automated ABI PRISM® 3730 DNA Analyzer. Primers used included ITS4, ITS5HP (for ITS regions) and UniC and UniF (for trnL-trnF spacer plus trnL intron regions). Thaler sequences were analyzed and edited and a consensus sequence for each sample (contigs) was constructed using the program Sequencher version 4.5. Gaps in the sequence were deleted and alignment was done by uploading the sequence to the NCBI database. Non-aligned regions in the sequence were also removed prior to phylogenetic analysis.

**Statistical analysis:** The data sets obtained from the sequenced nuclear (ITS) and chloroplast (trnL-trnF spacer plus trnL intron) gene regions of the samples were subjected to phylogenetic analysis by Cladistic method using PhyML_3.0.1_win32 option of SeaView4 software. PhyML (Phylogenetic estimation using Maximum Likelihood) was used to analyze each data set separately prior to combining.

**RESULTS**

Variations in fruit and flower traits of some samples studied are as shown in Fig. 1 and 2. Fruits usually occur as berry, fleshy, usually green, white or purple in colour...
Fig. 1(a-g): Variation among eggplant fruit samples, (a) Small, round and greenish of *S. nigrum*, (b) Small, round and purplish of *S. scabrum*, (c) Big, round and greenish of *S. macrocarpon*, (d) Egg-shaped or ellipsoid, white with green stripe of *S. gilo*, (e) Big, obovate and purplish of *S. melongena*, (f) Big, obovate and white turning yellow of *S. melongena* and (g) Big, round and greenish with thorny calyx of *S. dasyphyllum*

Fig. 2(a-g): Variations among some eggplant samples flower traits, (a) Rotate and purplish white of *S. melongena*, (b) Semi-stellate and white of *S. melongena*, (c) Semi-stellate and white of *S. gilo*, (d) Stellate and white of *S. gilo*, (e) Rotate and purplish of *S. melongena*, (f) Rotate and whitish of *S. macrocarpon* and (g) Semi-stellate and purplish white of *S. macrocarpon*
Fig. 3: DNA profile of trnL-E-trnF regions, L: 100 bp DNA ladder, -ve: Negative control, 1-20 represents amplicon from samples

Fig. 4: Phylogenetic relationships among eggplant *Solanum* species using *Capsicum annuum* L. as the outgroup. Bootstrap supports are indicated above branches; five clades and four ungrouped (a, b, c and d) samples are also recognized from the analysis

turning orange or red as they become ripe. Flowers are usually perfect, (4-) 5-merous, actinomorphic or zygomorphic; calyx campanulate, sometimes accrescent in fruit; corolla rotate, campanulate, stellate, or urceolate, white, green, yellow, pink, or purple.

The quality of the PCR products for all samples were verified on a 1.5% agarose gel before proceeding to sequencing and the PCR result for trnL-trnF spacer was as shown in Fig. 3. Nucleotide sequences generated from sequencing of ITS region and trnL-trnF spacer plus trnL intron were analyzed with Seaview4 software to draw the phylogenetic tree using the Maximum Likelihood (PhyML) option of the software. *Capsicum annuum* was used as an outgroup and the resulting phylogenetic tree is shown in Fig. 4. Figure 4 shows five major clades (1 to 5) and four ungrouped [A to D, corresponding to *S. gilo* (LA23), *S. torvum* (OS21 and CR37) and *S. melongena* (ED31), respectively]. Clade 1 contains five accessions while
Clade 4 is the lowest with two samples. Each of the remaining three clades contained three accessions, respectively. The highest bootstrap value of 100 was observed between *S. macrocarpon* and *S. torvum* and between *S. macranthum* and *S. indicum*. This was closely followed by a value of 97 between *S. scabrum* (OY18) and *S. gilo* (LA24). However, low bootstrap values were observed among those found in clades 1 and 2.

**DISCUSSION**

Molecular analysis based on sequence of one nuclear (ITS region: ITS1, 5.8S, ITS2) and two plastid (non-coding trnL intron and intergenic spacer trnL-trnF) markers resulted in a well resolved phylogenetic relationships, with results strongly suggesting that most of the *Solanum* accessions in this study are monophyletic (i.e., with common ancestor). This finding is in agreement with the findings of Olmstead and Palmer (1997) and Levin et al. (2005). Olmstead and Palmer (1997) had reported that phylogenetic classification of *Solanum* species using chloroplast DNA restriction site variation revealed that the Old World and Australian species within Subgenus *Leptostemonum* form a monophyletic clade.

Levin et al. (2005) also found out that majority of *Solanum* taxa comprise a monophyletic lineage, though some taxa are polyphyletic. However, these are in sharp contrast with Weese and Bohs (2007) and Bohs and Olmstead (1997) who, based on their findings, suggested that most traditionally recognized *Solanum* subgenera are not monophyletic. Of the three gene regions studied, trnL-trnF spacer was found useful for phylogenetic inference as indicated by both a high percentage of parsimony-informative sites as well as a low level of homoplaspy. The observed high bootstrap value of 100 between *S. macrocarpon* (ED32) and *S. torvum* (OY29) and between *S. macranthum* (CR50) and *S. indicum* (CR38), followed closely by a value of 97 between *S. scabrum* (OY18) and *S. gilo* (LA24) is an indication of a close relationship among those species concerned and possibility of having a common ancestor.

Meanwhile, the observed low bootstrap values in clades 1 and 2 showed that, though members may have a common ancestor, each one has diverged so that they share little or no resemblance to each other. It should be noted as well that *S. torvum* that is separated from *S. melongena* but they grouped together at low bootstrap value; an indication of distant relatedness. Mace et al. (1999), Furini and Wunder (2004) and Ishikii et al. (2008) also found out that *S. torvum* section Torva was separated out from section Melongena and other related *Solanum* species. However, this observation contrasted with earlier findings of Sifan et al. (2014) who observed relatedness between *S. melongena* and *S. torvum* due to their occurrence in the same clade based on data derived from RAPD analysis.

**CONCLUSION**

In conclusion, the examined species were resolved as a monophyletic group from the outgroup *Capsicum*. However, high level of specific variations was displayed between some Nigerian eggplant species particularly between *S. melongena* and *S. aethiopicum* and between *S. macrocarpon* and *S. indicum* that were clearly separated or recognized by a high bootstrap values. The results of this study could be effectively used in genetic improvement of both wild and cultivated eggplant varieties as well as *in situ* and *ex situ* conservation.

**ACKNOWLEDGEMENT**

The authors express their sincere appreciation to the University of Lagos Central Research Committee for providing in part Grant (No: CRC 2005/03, UNILAG) for this study.

**REFERENCES**


