Complete plastid genome sequence of African nightshade
(Solanum scabrum) and comparative plastomics across Solanales

Sablok, Gaurav

2019-08-19

http://hdl.handle.net/10138/305082

Complete plastid genome sequence of African nightshade (*Solanum scabrum)* and comparative plastomics across Solanales

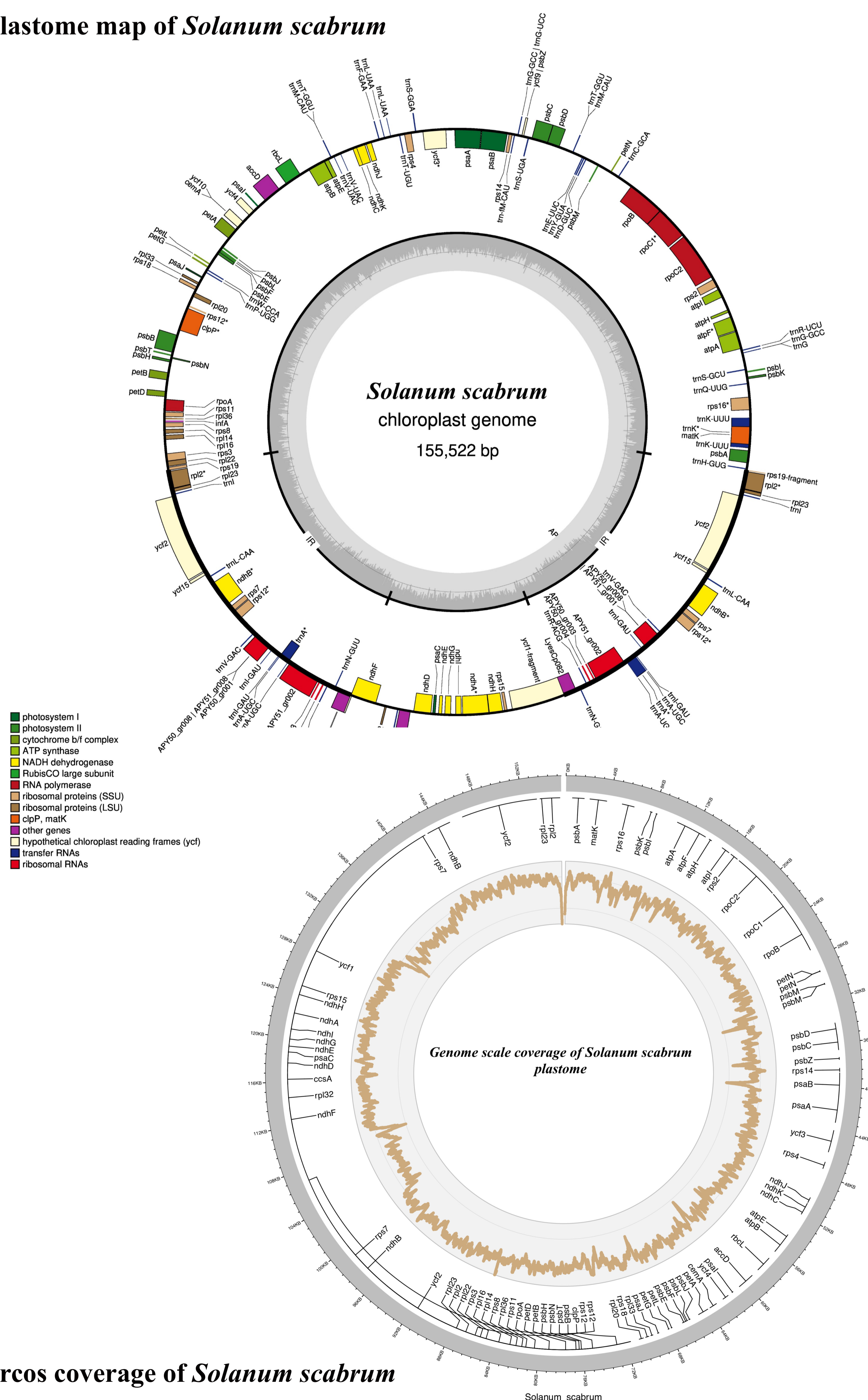Gaurav Sablok[1,2], Andres Orejuela[3], Xiaolan He[1], Jaakko Hyvönen[1,2*], Tiina Särkinen[3], Péter Poczai[1]

[1]Finnish Museum of Natural History, Helsinki, Finland; [2]Organismal and Evolutionary Biology (OEB) Research Programme, University of Helsinki, Finland; [3]Royal Botanic Garden Edinburgh, University of Edinburgh. *Email: jaakko.hyvonen@helsinki.fi
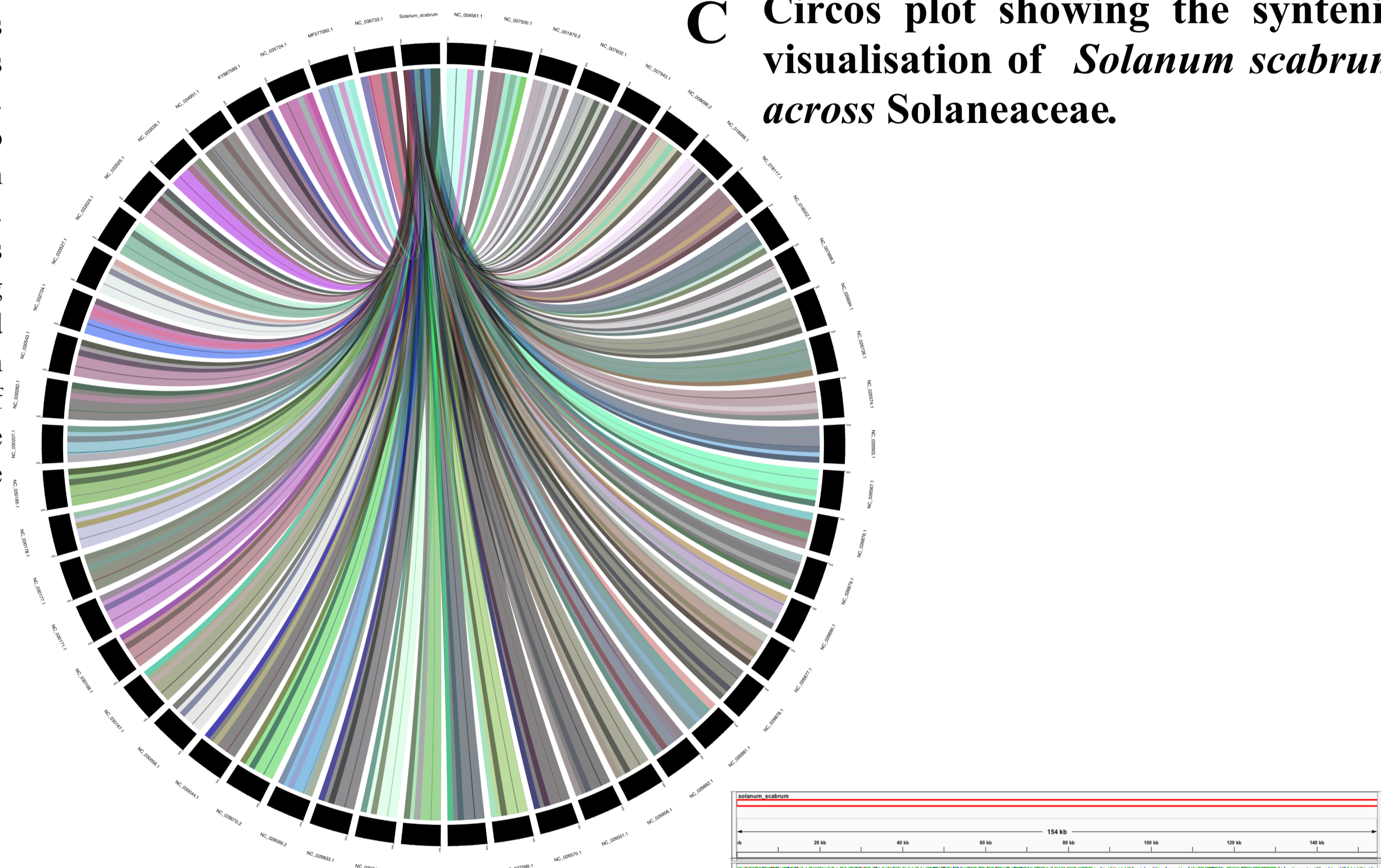
Introduction:

Solanaceae is a particularly interesting angiosperm family, not only because they include many major crop species such as potato, tomato, eggplant, pepper, tobacco and ornamentals like petunias, but also because numerous species are used as biological models. They have been widely used for understanding crop genetics and plant genome evolution in general. Sequencing efforts have been concentrated mostly to sequence genomes of important crop species of Solanaceae to understand the links between wild and cultivated members of the family. We present the complete plastome of African nightshade (*Solanum scabrum* Mill.) a hexaploid (2n = 6x = 72) species of the *S. nigrum* L. complex or *Solanum* sect. *Solanum,* a widely cultivated species across Africa. Recent studies highlight *S. scabrum* as a "super-vegetable" for its nutritional and environmental benefits with potential of global importance. The leaves and berries are the source of coloring plant extracts, inks and dyes, and they are rich in proteins, fibres, iron, vitamins and amino acids. Using 12,413,264 paired end reads deposited in the sequence read archive (SRA) we have assembled the plastid genome sequence of African nightshade with an estimated coverage of 123×. The plastid genome sequence had a total size of 155,522 bp, typical of Solanaceae with a large single copy (LSC) region of 85,896 bp and small single copy (SSC) region of 18,406 bp while the IRs comprised of 25,610 bp. We illustrate the role of *Solanum scabrum* and its comparative plastomics across Solanaceae and Convolvulaceae to understand the plastomics of Solanales.

Material and Methods:

Solanum scabrum plastome assembly was performed using the previously available sequencing reads available from NCBI SRA archive (SRR6664741). Genomic reads were mapped to the available Solanaceae plastomes using the Bowtie2 with an --end-to-end, --sensitive (-D 15 -R 2 -N 0 -L 22 -i S,1,1.15) parameters and the corresponding SAM and BAM files were sorted using the SAMtools (Li et al. 2009) and BAMTools (Barnett et al. 2011). Plastome mapped reads were used further for assembly using SPADES denovo assembler (Bankevich et al. 2012) with a kmer range of 55,67,71 and – careful mode. Assembled contigs were checked for coverage by realigning the reads to the assembled contigs and low coverage contigs were filtered out. Contigs were aligned to the *Solanum tuberosum* (NC_008096.2) plastome using Mummer (Kurtz et al. 2004). Coverage of the assembled plastome was checked by realining the genomic reads and using BAMTools (Barnett et al. 2011) and BEDTools (Quinlan and Hall, 2010). Plastome annotation were performed using the GESEQ (Tillich et al. 2017). Synteny of the assembled plastome was evaluated using the LASTz against the Solanaceae plastomes and Circos (Krzywinski et al. 2009) for visualization. Codon alignments were done using the MACSE and internal and terminal codons were masked prior to the supermatrix, which was used for the phylogeny contruction using the IQTREE (Nguyen et al. 2015) with model selection using ModelFinder (Kalyaanamoorthy et al. 2017) and ultrafast bootstrap using UFBoot2 (Hoang et al. 2018).

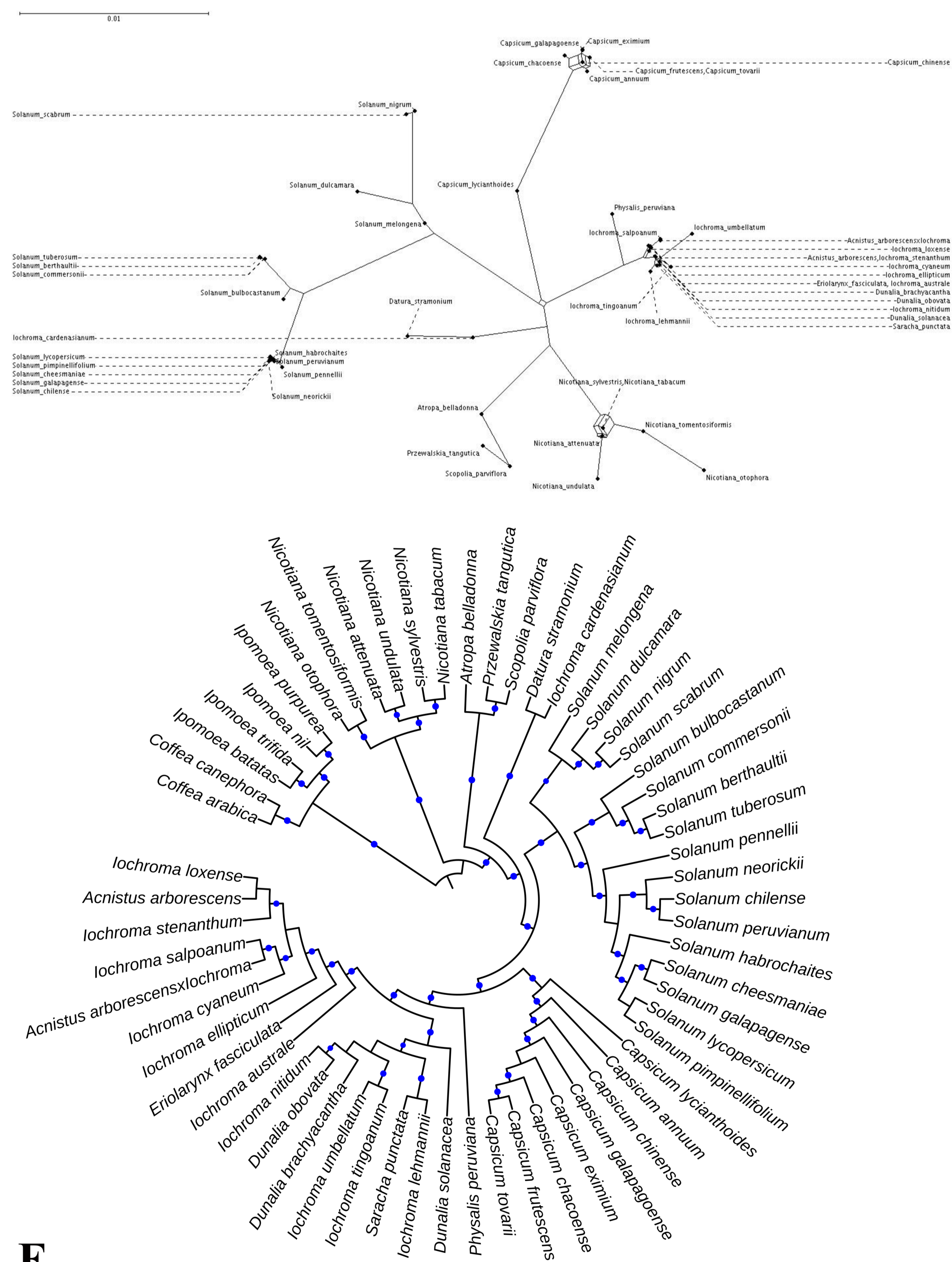**A** Plastome map of *Solanum scabrum*

B Circos coverage of *Solanum scabrum*

C Circos plot showing the syntenic visualisation of *Solanum scabrum across* Solaneaceae.

D Structural mapping of insertion/ deletion events of *Solanum scabrum across* Solaneaceae.

E Network analysis of *Solanum scabrum across* Solaneaceae. using Convolvulaceae and Rubiaceae as outgroups.

F Phylogenetic placement of *Solanum scabrum across* Solaneaceae using Convolvulaceae and Rubiaceae as outgroups.

References: Li et al. Bioinformatics (2009) 25: 2078-9; Barnett et al. Bioinformatics (2011) 27:1691-1692.; Tillich et al. Nucleic Acids Res. (2017) 45(W1): W6–W11; Kurtz et al. Genome Biol. (2004) 5:R12; Quinlan et al. Bioinformatics (2010) 26:841–842; Bankevich J Comput Biol. (2012) 19:455–477; Krzywinski et al. Genome Res. (2009) 19:1639-1645; Hoang et al. Mol. Biol. Evol. (2018) 35:518-522; Kalyaanamoorthy et al. Nat. Methods (2017) 14:587-589; Nguyen et al. Mol. Biol. Evol. (2015) 32:268-274.