Supplemental figures Kourelis et al: *Re-annotation of N. benthamiana gene models.*
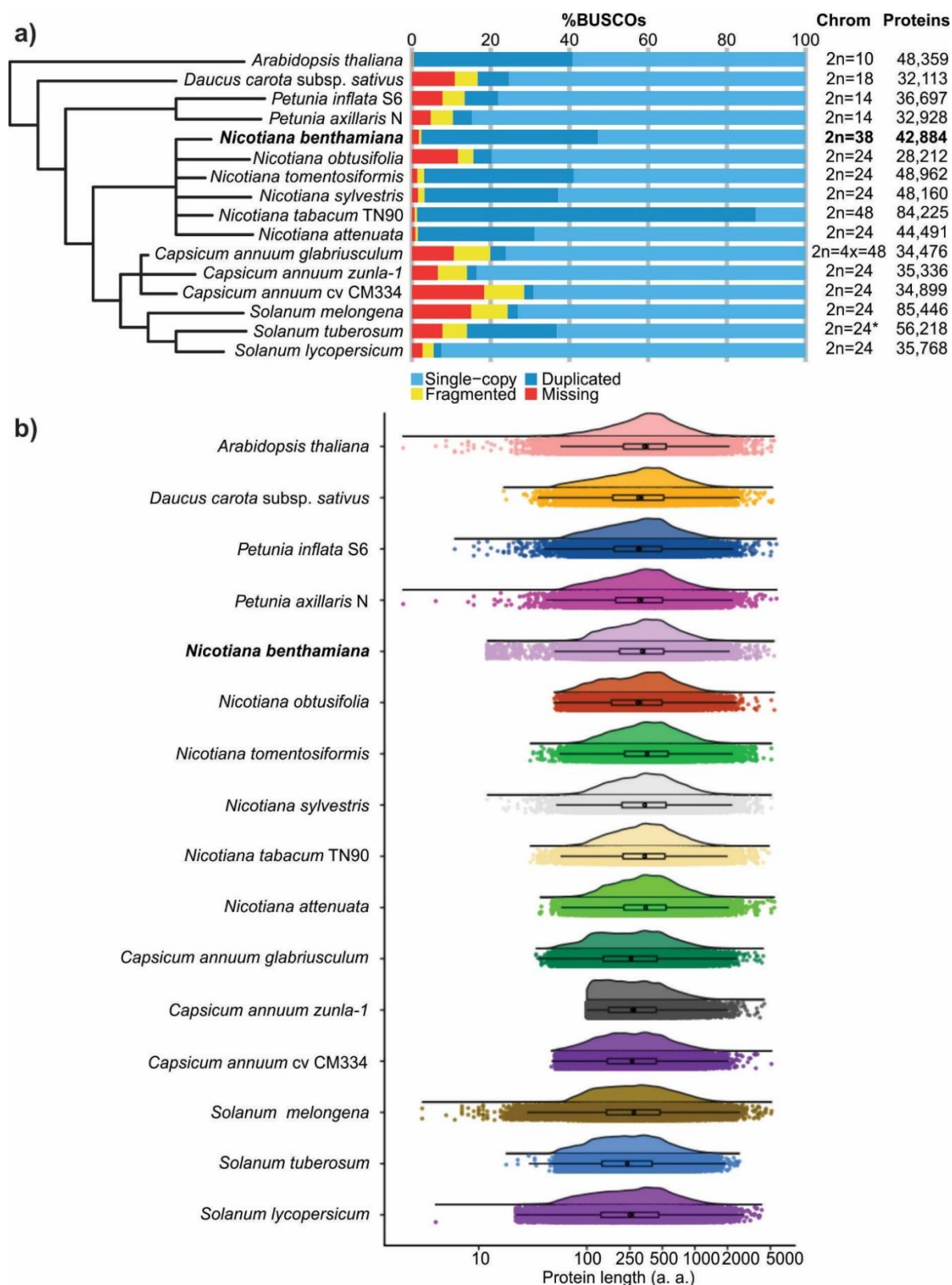

Figure S1: Comparison of Solanaceae proteomes. **a)** Completeness of the predicted proteomes from sequenced Solanaceae genomes was estimated using BUSCO v3 with the embryophyta database. *Certain *Solanum tuberosum* species are polyploid. **b)** Violin and boxplot g of $\log_{10}$ protein length distribution of each predicted proteome. Jittered dots show the raw underlying data.
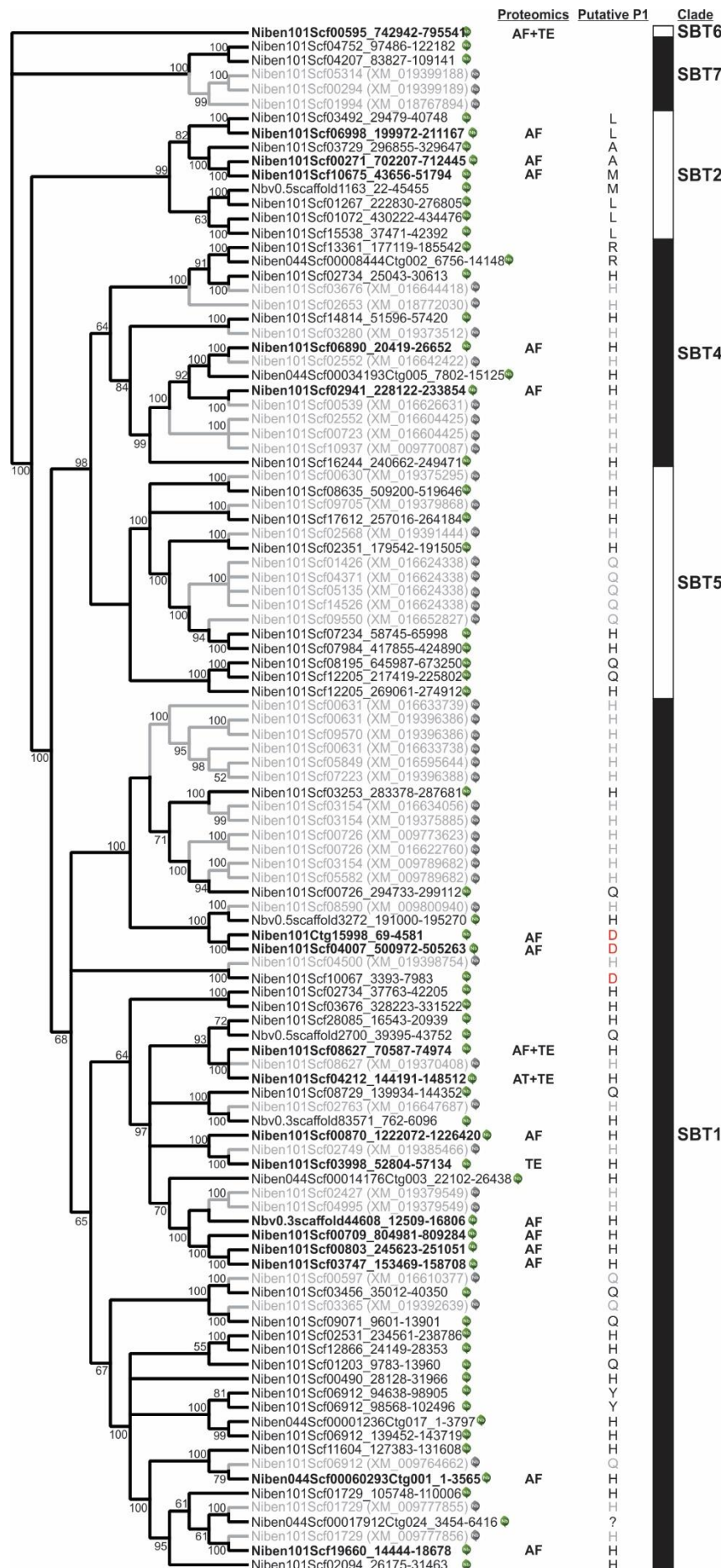
1

Figure S2: Phylogenetic analysis of the subtilisin gene-family from figure 4 with names.

The evolutionary history of the subtilase gene family was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analysed. Putative pseudogenes are indicated in grey. Subtilases identified in apoplastic fluid (AF) and/or total extract (TE) are indicated with yellow and green dots, respectively. Putative P1 based on residue at the prodomain junction is indicated, putative phytaspases are indicated in red. Naming of subtilase clades according to (Taylor and Qiu, 2017).

Figure S3: Phylogenetic analysis of the subtilisin gene-family of tomato and Arabidopsis and including other previously characterized subtilisins.

The evolutionary history of the subtilase gene family was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model. The bootstrap consensus tree inferred from 250 replicates is taken to represent the evolutionary history of the taxa analysed. Putative pseudogenes are indicated in grey. Putative P1 based on residue at the prodomain junction is indicated. Putative phytaspases are indicated in red, confirmed phytaspases in blue. Naming of subtilase clades according to (Taylor and Qiu, 2017).

Other supplemental datasets:

Table S1: GO-SLIM term enrichment complete at p≤0.05

Table S2: MEROPS family term enrichment complete

Table S3: Gene-model comparison

Supplemental dataset 1: NbC gene-models database fasta nucleotide sequence

Supplemental dataset 2: NbC gene-models database gff3 annotation

Supplemental dataset 3: NbC predicted CDS

Supplemental dataset 4: NbC predicted proteome

Supplemental dataset 5: PFAM, SignalP, ApoplastP, and Sma3 v2 annotation of the NbC predicted proteome