PCP
PLANT & CELL PHYSIOLOGY

Genome Sequence and Analysis of *Nicotiana benthamiana*, the Model Plant for Interactions between Organisms

Ken-ichi Kurotani[1], Hideki Hirakawa[2], Kenta Shirasawa[2], Yasuhiro Tanizawa[3], Yasukazu Nakamura[3], Sachiko Isobe[2,*] and Michitaka Notaguchi[1,4,*]

[1]Bioscience and Biotechnology Center, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan
[2]Department of Frontier Research and Development, Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba, 292-0818 Japan
[3]Research Organization of Information and Systems, National Institute of Genetics, Yata, Mishima, 411-8540 Japan
[4]Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601 Japan

*Corresponding authors: Sachiko Isobe, E-mail, sisobe@kazusa.or.jp; Michitaka Notaguchi, E-mail, notaguchi.michitaka.c4@f.mail.nagoya-u.ac.jp
(Received 11 August 2022; Accepted 7 December 2022)

Regular Paper

Nicotiana benthamiana **is widely used as a model plant for dicotyledonous angiosperms. In fact, the strains used in research are highly susceptible to a wide range of viruses. Accordingly, these strains are subject to plant pathology and plant–microbe interactions. In terms of plant–plant interactions, *N. benthamiana* is one of the plants that exhibit grafting affinity with plants from different families. Thus, *N. benthamiana* is a good model for plant biology and has been the subject of genome sequencing analyses for many years. However, *N. benthamiana* has a complex allopolyploid genome, and its previous reference genome is fragmented into 141,000 scaffolds. As a result, molecular genetic analysis is difficult to perform. To improve this effort, de novo whole-genome assembly was performed in *N. benthamiana* with Hifi reads, and 1,668 contigs were generated with a total length of 3.1 Gb. The 21 longest scaffolds, regarded as pseudomolecules, contained a 2.8-Gb sequence, occupying 95.6% of the assembled genome. A total of 57,583 high-confidence gene sequences were predicted. Based on a comparison of the genome structures between *N. benthamiana* and *N. tabacum*, *N. benthamiana* was found to have more complex chromosomal rearrangements, reflecting the age of interspecific hybridization. To verify the accuracy of the annotations, the cell wall modification genes involved in grafting were analyzed, which revealed not only the previously indeterminate untranslated region, intron and open reading frame sequences but also the genomic locations of their family genes. Owing to improved genome assembly and annotation, *N. benthamiana* would increasingly be more widely accessible.**

Keywords: Cell wall enzymes • Genome assembly • *Nicotiana benthamiana* • *Nicotiana tabacum* • Polyploidy • Pseudomolecule

Introduction

Nicotiana benthamiana is one of the most widely used experimental models in plant biology, especially the interaction between different living organisms (Bombarely et al. 2012, Bally et al. 2018). *Nicotiana benthamiana* has been employed in plant pathology studies owing to its susceptibility to diverse plant diseases, particularly viral diseases. *Nicotiana benthamiana* has a 72-bp insertion, including a termination codon, in its RNA-dependent RNA polymerase (*RdRP*) gene (*Rdr1*), causing premature termination of the open reading frame (ORF) of the gene (Yang et al. 2004). *Nicotiana benthamiana* plants transformed with the *RdRP* gene of *Medicago truncatula* displayed resistance to the tobacco mosaic virus of the genus, *Tobamovirus*, indicating that defects in the *Rdr1* gene may be a factor in the high RNA virus susceptibility of *N. benthamiana* (Yang et al. 2004). *Nicotiana benthamiana* is also a popular platform for recombinant protein production. In fact, non-infectious virus-like particles have been used for the manufacture of COVID-19 vaccines (van Herpen et al. 2010, Bally et al. 2018, Mamedov et al. 2021). Our previous studies revealed that interfamily grafting, which is usually considered difficult to establish, can be achieved with *N. benthamiana* (Notaguchi et al. 2020, Kurotani and Notaguchi 2021). In these studies, the high activity of cell wall modification enzymes was suggested through transcriptome and functional analyses. A molecular genetic study revealed that the upregulation of *Glycosyl hydrolase 9B3* (*GH9B3*) encoding β-1,4-glucanase in *N. benthamiana* after grafting with intra- and interfamily partners was important to achieve cell–cell adhesion at the graft boundary. Transcriptome analysis revealed that other cell wall modification enzymes, such as β-1,3-glucanase, xyloglucan hydrolase and expansin, were also upregulated during grafting in *N. benthamiana*. These observations revealed a characteristic feature of *N. benthamiana*. A similar characteristic was found in *Petunia*,

Plant Cell Physiol. 64(2): 248–257 (2023) doi:https://doi.org/10.1093/pcp/pcac168, Advance Access publication on 8 February 2023, available online at https://academic.oup.com/pcp

the closest genus of *Nicotiana* in Solanaceae (Kurotani et al. 2022). These plants may partially conserve similar cell wall modifications. Thus, *N. benthamiana* can serve as a model for various cellular events in plants, which includes not only the recognition and defense mechanisms of plant viruses, bacteria and other foreign invaders but also individual recognition, damage response and wound repair mechanisms at the cellular level in plants.

The Solanaceae family is widely distributed throughout the temperate and tropical continents. This family includes many of the most important agricultural species in the world, such as potato, tomato, eggplant, petunia and tobacco. Several specific groups have been defined in the genus, *Nicotiana*. *Suaveolentes*, the section to which *N. benthamiana* belongs, is mainly found in Australia. The *Nicotiana* genus has been repeatedly interbred, and the progeny of the paternal species of the *Suaveolentes* section has been demonstrated to belong to the *Sylvestres* section (Knapp et al. 2004, Bally et al. 2018). On the other hand, the maternal lineage might be the ancestor of the *Noctiflorae* section; however, the details are less clear. *Nicotiana tabacum*, the most widely cultivated common tobacco species, is included in the *Nicotiana* section. The paternal lineage of the *Nicotiana* section is considered to be of the *Sylvestres* section, while the maternal lineage is attributed to the *Tomentosae* section (Schiavinato et al. 2020). The establishment of *Suaveolentes* section including *N. benthamiana* by hybridization was estimated to have occurred approximately 10 million years ago (MyA) (Bally et al. 2018) and is thus older than the 200,000 years for *N. tabacum* (Leitch et al. 2008). The basic haploid chromosome number of the genus, *Nicotiana*, is $n = 12$. As *N. tabacum* is an allotetraploid, it has $n = 2\times = 24$ (Leitch et al. 2008). In contrast, *N. benthamiana* is thought to have a complex ancient allopolyploid genome consisting of 19 chromosomes (Wheeler 1945, Narayan 1987). For allotetraploid, two homeologs derived from each of the two subgenomes per gene were expected to be found. However, *N. benthamiana* appears to have evolved through many hybridizations among its polyploid ancestors. Furthermore, some genomic regions and chromosomes are most likely to be lost.

The use of *N. benthamiana* is becoming increasingly active, with two research consortia independently publishing assembled draft genome sequences (Bombarely et al. 2012, Naim et al. 2012). One published the *N. benthamiana* genome and transcriptome sequencing consortium (benthgenome) assembly (v0.5), and the other published the Sol Genomics assembly (v1.01). Both draft genome assemblies were constructed using Illumina-based sequencing with only minor differences. The quality of the genome assemblies, as measured by the benchmarking universal single-copy orthologs (BUSCO) score, which predicts the percentage of completely and accurately assembled coding genes, was very similar for both, with 95–96% and >90% of the genes annotated. However, owing to the complex allopolyploid structure of the genome, only finely divided scaffolds have been decoded (141,000 scaffolds), and the chromosome level has not been reached. The linkage status between genes is unknown, or sequence information on the regulatory regions of gene expression is lacking, which has been a barrier to genetic analysis.

Table 1 Statistics of the assembled *N. benthaminana* genome and CDS sequences, Nbe.v1.

	Genome (Nbe.v1)		Gene
	All	Pseudo-molecules	HC, CDS
Number of sequences	1,668	21	57,583
Total length (bp)	2,926,135,461	2,798,934,012	61,180,166
Average length (bp)	1,754,278	133,282,572	1,062
Maximum length (bp)	184,452,736	184,452,736	14,934
Minimum length (bp)	1,000	66,383,842	150
N50 length (bp)	141,754,421	141,754,421	1,407
Gaps (%)	0.004	0.004	0.000
GC%	38.1	37.9	42.7
BUSCOs (%) v3, obd10			
Complete	99.6	99.6	95.9
Complete single-copy	33.9	34.1	43.5
Complete duplicated	65.7	65.5	52.4
Fragmented	0.1	0.1	1.3
Missing	0.3	0.3	2.8

In this study, we aimed to complement sequence information close to the chromosomal level. The obtained sequence information was validated via a comparison with the sequence information of other Solanaceae species for which genomic information is available. We performed phylogenetic analyses of several genes that were suggested to be responsible for the high grafting ability of *N. benthamiana*.

Results

Sequencing and assembly

A total length of 48.9 Gb of sequence data was obtained from the paired-end (PE) library (**Supplementary Table S1**). The distribution of distinct *k*-mers ($k = 17$) revealed a single peak at multiplicities of 13, indicating high homozygosity of the 'lab strain' genome (**Supplementary Fig. S1**). Based on the identified peal, the estimated genome size of the 'lab strain' was 3,120.2 Mb.

A total length of 150.1 Gb of Hifi reads was generated from the seven SMRT cells. The read coverage of the lab strain genome was 48.1×. Hifiasm assembly generated 1,525 primary and 3,342 associate haplotigs (**Supplementary Table S2**). The total length of the primary contigs was 2,926.0 Mb, with an N50 length of 42.5 Mb. The primary contig sequences were then scaffolded with Omni-C reads using 3D-DNA. The number of sequences was 1,668, which was slightly higher than that of the Hifiasm primary sequence. The N50 length was 141.8 Mb, 3.3-fold longer than that of the Hifiasm primary sequences. The resulting assembled sequences were designated as Nbe.v1 (**Table 1**, **Supplementary Table S2**).

The assembled genome covered 93.8% of the *N. benthamiana* genome, as estimated by Jellyfish (3,120.2 Mb).

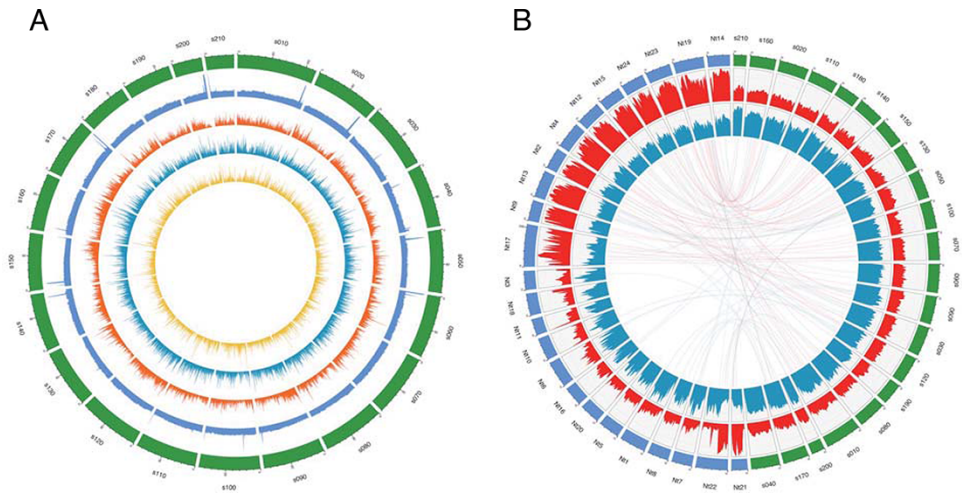

Fig. 1 The *N. benthamiana* genome. (A) The distribution of retrotransposons and genes in 21 scaffolds. Circles indicate the concentration of Ty3-gypsy (yellow) and Ty1-copia (blue), gene content (orange) and G+C% (light blue) from inner to outer. (B) Coverage of regions of high homology with *N. sylvestris* and *N. tomentosiformis* on *N. benthamiana* and *N. tabacum*. Circles indicate the coverage of *N. sylvestris* (blue) and *N. tomentosiformis* (red) from inner to outer. The center of the image represents the linkage of SSRs. Blue and red lines indicate SSRs on *N. sylvestris* and *N. tomentosiformis*, respectively. S010–S210 represent the 21 longest scaffolds of *N. benthamiana*, and Nt1–Nt24 represent the 24 chromosomes of *N. tabacum*.

The assembly quality of Nbe.v1 was assessed by mapping the sequences to 1,375 BUSCOs (**Table 1**). The percentage of identified complete BUSCOs was 99.6%, indicating the high quality of the assembled genome. The percentages of complete single copies and duplications were 33.9% and 65.7%, respectively. The relatively high number of duplicates reflects the allopolyploid genome structure of *N. benthamiana*.

Of the 1,668 assembled sequences, the length of the 21 longest scaffolds ranged from 66.4 to 184.5 Mb. The 21 longest scaffolds were regarded as pseudomolecules, which reflected the chromosomes of *N. benthamiana* (**Supplementary Table S3**). However, as the chromosome number of *N. benthamiana* was reported as 19 (Kourelis et al. 2018), sequences of a few chromosomes were considered to be separated into two or three pseudomolecule sequences. The total length of the 21 pseudomolecules was 2,798.9 Mb, occupying 95.6% of the assembled genome (**Table 1**).

Gene prediction and annotation

A total of 1.4 Gb Iso-Seq sequences were obtained from various organs under different treatments (**Supplementary Table S1**). The sequences were then clustered and collapsed, and a total of 84,934 confident-complete gene sequences, which revealed ORF regions on the sequences, were identified with a total length of 101.1 Mb by ANGEL (**Supplementary Table S4**). Meanwhile, 84,754 genes with a total length of 75.5 Mb were predicted on Nbe.v1 by BRAKER2. The sets of genes predicted by ANGEL and BRAKER2 were then merged, and the resulting 84,570 sequences with a total length of 75.5 Mb were designated as predicted genes of Nbe.v1.

The 84,570 sequences were then classified as high-confidence (HC), low-confidence (LC) or transposable element (TE) sequences based on the evidence level. The numbers of predicted gene sequences classified as HC, LC and TE were 57,583, 18,116 and 8,871, respectively (**Table 1**, **Supplementary Table S4**). The percentage of complete BUSCOs in HC was 95.9%, whereas those in LC and TE were 0.0% and 1.9%, respectively, suggesting that most of the protein-coding gene sequences were classified as HC.

The distribution of retrotransposons and genes in 21 long scaffolds was confirmed (**Fig. 1A**). The Ty1-copia and Ty3-gypsy groups were evenly distributed throughout the entire area in both scaffolds; however, the gene distribution tended to be markedly denser closer to the ends. In addition, telomere repeat sequence searches confirmed the presence of the 5′ terminal regions of Nbe.v1.s00010, Nbe.v1.s00020, Nbe.v1.s00060, Nbe.v1.s00130 and Nbe.v1.s00160 and the 3′ end of Nbe.v1.s00030. This newly assembled *N. benthamiana* genome, referred to as Nbe.v1, and the constructed annotation data can be accessed at NbenBase (https://nbenthamiana.jp).

Linkage and synteny analysis of *N. benthamiana* and *N. tabacum*

For comparison of the *N. benthamiana* and *N. tabacum* genomes, the coverage of regions of high homology with *Nicotiana sylvestris* and *Nicotiana tomentosiformis* was examined (**Fig. 1B**). The ancestral species of *N. sylvestris* is the common parent of *N. benthamiana* and *N. tabacum*, whereas the ancestral species of *N. tomentosiformis* is the other parent of *N. tabacum*. As previously reported, chromosomes derived from *N. sylvestris* and *N. tomentosiformis* were relatively clear in *N. tabacum*. In contrast, in *N. benthamiana*,

N. sylvestris–derived sequences were almost equally distributed in all scaffolds (**Fig. 1B**).

Subsequently, simple sequence repeat (SSR) markers identified in *N. tabacum* (Tong et al. 2016) were searched in the genome of *N. benthamiana* to derive chromosomal correspondence. Using the BLAT program (Kent 2002) and in silico PCR with the *N. tabacum* SSR primer set as queries, 4,622 and 40,717 SSRs were detected in the *N. benthamiana* and *N. tabacum* genomes, respectively. Of these SSRs, 131 SSRs, identified as unique to both *N. benthamiana* and *N. tabacum*, were plotted on the *N. benthamiana* scaffold (**Fig. 1B**, **Supplementary Fig. S2 and Table S5**). Of the 131 unique SSRs, 63 were derived from the ancestral species of *N. sylvestris*. Of the 24 chromosomes in *N. tabacum*, Nt1, Nt3, Nt5, Nt6, Nt7, Nt8, Nt10, Nt11, Nt16, Nt18 and Nt20 were considered to be predominantly derived from the ancestral species of *N. sylvestris*; Nt2, Nt4, Nt9, Nt12, Nt13, Nt14, Nt15, Nt19, Nt21, Nt23 and Nt24 from the ancestral species of *N. tomentosiformis* and Nt17 and Nt23 from mixtures of the two parental species (Sierro et al. 2014, Edwards et al. 2017). In contrast to *N. tabacum* in which the chromosomes derived from the two ancestral species were relatively clearly distinguishable, in *N. benthamiana*, SSRs derived from the ancestral species of *N. sylvestris* were scattered across most scaffolds (**Supplementary Fig. S2**).

Synteny analysis was performed between *N. benthamiana* and *N. tabacum* (**Supplementary Figs. S3, S4**). *Nicotiana tabacum* is considered a hybrid of the ancestral species of *N. sylvestris* and *N. tomentosiformis* with chromosomes derived from each. Nt4 has a high coverage of sequences derived from the ancestor of *N. tomentosiformis*, and Nt6 is derived from the ancestor of *N. sylvestris*. Synteny between these chromosomes has been previously reported (Sierro et al. 2014, Edwards et al. 2017). In this study, synteny analysis of *N. benthamiana* and *N. tabacum* genome sequences revealed that two *N. benthamiana* scaffolds, Nbe.v1.s00120 and Nbe.v1.s00150, had synteny with two *N. tabacum* chromosomes, Nt4 and Nt6 (**Supplementary Fig. S3B**). Similarly, Nbe.v1.s00130 and Nbe.v1.s00140 displayed synteny with both Nt12 and Nt16 (**Supplementary Fig. S3C**). Collinear blocks were detected between *N. benthamiana* and *N. tabacum* in all chromosomes (**Supplementary Fig. S4**). These results indicate the retention of two sets of genomic sequences in alloploid species; however, sequences derived from the *Sylvestres* section, the common ancestral progenitor of *N. benthamiana* and *N. tabacum*, were found to be quite scattered and distributed in the *N. benthamiana* chromosomes.

Comparative analysis of *N. benthamiana* and other Solanaceae family gene sequences

To explore the evolution of the gene family in the Solanaceae family and *Arabidopsis thaliana*, orthologous clustering using the *N. benthamiana* annotation constructed in this study and the five published gene annotations (269,901 total genes) was performed, and 27,624 orthologous groups were defined (**Fig. 2A**). Most orthogroups (11,526) were common to all species, and the second most common category was the orthogroups present in five species of Solanaceae (3,221 orthogroups). The species tree inferred by Knapp et al. (2004) indicated that *N. tabacum* and *N. sylvestris* had the closest relationship, followed by *N. benthamiana*. Both *N. benthamiana* and *N. tabacum* have an ancestral species of *N. sylvestris* as a hybrid parent; however, hybridization of *N. benthamiana* might have occurred earlier than *N. tabacum*, consistent with the results of this analysis (**Fig. 2B**). In addition, analyses of age distributions built from the distribution of silent divergence rate (dS) among *N. benthamiana*, *N. tabacum*, *Solanum lycopersicum* and *A. thaliana*. The dS between *N. benthamiana* and *N. tabacum* was estimated to be about 0.1–0.2, indicating that the divergence occurred at about 3–7 MyA (**Supplementary Fig. S5**).

Rdr1 and related gene homoeologs in *N. benthamiana*

As defects in the *Rdr1* gene have been demonstrated to be one of the reasons for the high susceptibility of *N. benthamiana* to RNA viruses, we confirmed the sequence of the *Rdr1* gene in the sequence determined in this study. Consequently, the *Rdr1* gene is divided into two annotations, *Nbe.v1.s00180g20180* and *Nbe.v1.s00180g20190*, due to truncation of its ORF by insertion of 72 nucleotides, as revealed in a previous study (Yang et al. 2004, **Supplementary Fig. S6**). No other sequences homologous to *Rdr1* at the DNA sequence level were found in the genome, and the closest match at the amino acid level, *Nbe.v1.s00150g34350*, was annotated with *RdRP2* (LOC107773126) in *N. tabacum*. *Nbe.v1.s00150g34350* and *Nbe.v1.s00120g10840*, although on separate chromosomes, are very close at the DNA level, with 97% identity, suggesting that these two genes are homoeologs inherited from the ancestors of *N. benthamiana*, and are considered to be independent of *Rdr1*. The next closest gene at the amino acid level was *Nbe.v1.s00010g38270*, which encodes *SDE1/SGS2*, one of the two *RdRPs* found in *N. benthamiana* in previous reports (Yang et al. 2004), and annotated in the Sol Genomics assembly (Niben1.01) as *Niben101Scf12609g01010.1*. *Nbe.v1.s00010g38270* is 97% identical at the DNA level to *Nbe.v1.s00100g30920*, and both are considered homoeologs.

Genome sequence analysis of genes of interest in interfamily grafting of *N. benthamiana*

Previously, we identified *GH9B3* (*Niben101Scf01184g16001* in Niben1.01, which corresponds to *Nbe.v1.s00150g18250* in Nbe.v1), which plays an important role in the establishment of interfamily grafting in *N. benthamiana* (Notaguchi et al. 2020). The coding sequence (CDS) and translated amino acid sequences of *GH9B3* were identical in Niben1.01 and Nbe.v1. Fortunately, the 2-kb base sequence upstream of the transcription start site did not contain 'N' in Niben1.01 (**Supplementary Fig. S7A**). In our previous report (Notaguchi et al. 2020), 189 genes were identified that showed a pattern of increased expression from 1 to 7 d after grafting in *N. benthamiana* similar to

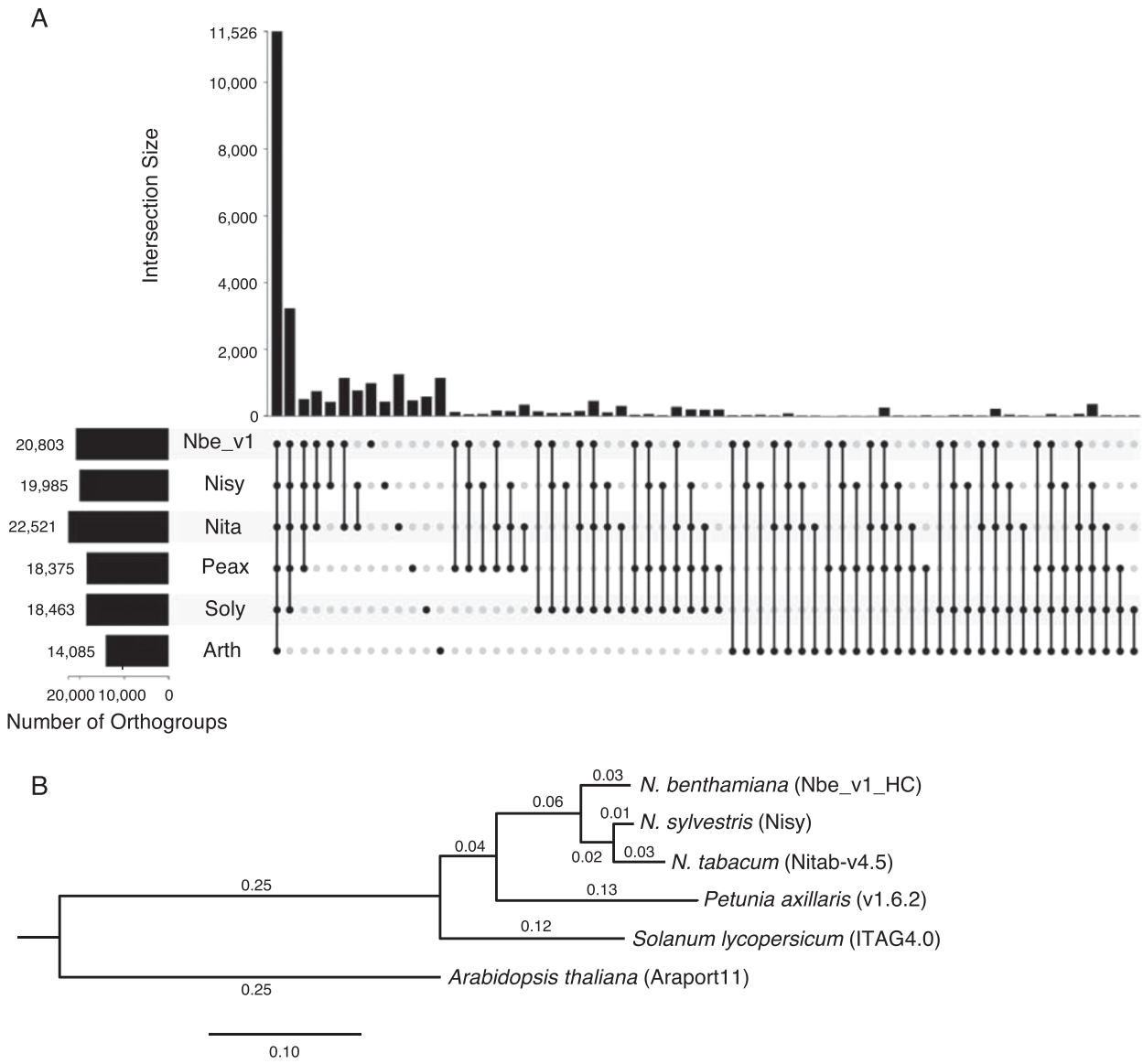

Fig. 2 Comparative and phylogenetic analyses with other plant species at the gene level. (A) Upset plot comparing the shared orthogroup among *N. benthamiana* (Nbe.v1), *N. sylvestris* (Nisy), *N. tabacum* (Nta), *P. axillaris* (Peax), *S. lycopersicum* (Soly) and *A. thaliana* (Arth). (B) A phylogenetic tree constructed among the six plant species. The tree was constructed with the STAG method using OrthoFinder v2.5.4. The values indicate the branch lengths. The scale bar represents the number of differences between the sequences.

GH9B3. We further focused on other cell wall modification enzyme genes in this study. No sequencing deficiencies in the CDS of *xyloglucan endotransglucosylase/hydrolase 28* (*XTH28; Niben101Scf00984g04002* and *Nbe.v1.s00100g19910*) were found in Niben1.01; however, the N regions were described within the third intron and 2.2 kb upstream of the start codon. Our analysis revealed 83 bases of unknown sequences in the introns, and an extra 17 bases of sequences were described in the upstream region due to misassembly in the N regions (**Supplementary Fig. S7A**). In *Expansin B3* (*EXPB3; Niben101Scf04102g03015* and *Nbe.v1.s00190g01690*), only the sequence of the transcription region was described in Niben1.01, and the promoter region was completely unknown; in Nbe.v1, the sequence up to the adjacent gene was revealed (**Supplementary Fig. S7A**). In Niben1.01, seven of the 189 genes also contained unidentified bases in the CDS. In the transcribed region, including introns, the N regions were present in 38 genes. Furthermore, there were 50 genes with undetermined 2-kb promoter regions (**Supplementary Table S6**).

Another gene of interest, *Niben101Scf14799g01012*, was the closest match to *AT4G16260* (*Glycosyl hydrolase superfamily protein*) in the full-length tblastx search against *A. thaliana* genes. *AT4G16260* is a member of the *β-1,3-glucanase* family and is homologous to *BG1* (*AT3G57270*), *BG2* (*AT3G57260*) and *BG3* (*AT3G57240*). The genomic sequence of *Niben101Scf14799g01012* was identical to that of *Nbe.v1.s00020g13490* obtained in

this analysis. However, the predicted CDS sequence in Nbe.v1 was 225 bases (75 amino acids) longer than the annotation in Niben1.01 (**Supplementary Fig. S7A**). A comparison of the actual mapped regions of the RNA-sequencing (RNA-seq) reads shown with the genomics browser suggested that Nbe.v1 annotation was more suitable (**Supplementary Fig. S7B**).

Phylogenetic analysis of homologous genes

The above-mentioned four genes form large gene families in a variety of plants, whereas in *N. benthamiana*, these genes constitute more complex families. Notably, phylogenetic analysis of the Niben1.01 gene annotation and Nbe.v1 annotation by amino acid sequence revealed a minor difference between the sequences of the two annotations, with correspondence for most genes. Owing to the large genetic distance between *A. thaliana* and *N. benthamiana*, it was difficult to clearly show the orthologs of these families (**Supplementary Figs. S8–S11**). In the phylogenetic tree of *β-1,3-glucanase*, the clade containing *A. thaliana BG1–BG3* and *AT4G16260* showed largely separated branches between *A. thaliana* and *N. benthamiana*. Therefore, the homologs of the genes between these plants did not clearly correspond to each other (**Fig. 3A**). Using the CDS of each of the 19 *β-1,3-glucanase* genes in the Nbe.v1 annotation as a query, we performed a homology search in tBlastx against the ARAPORT11 gene model and found that the top hit for 16 genes was *AT4G16260*. These genes were clustered on the four scaffolds of Nbe.v1 (**Fig. 3B**). In contrast, genes homologous to *BG1*, *BG2* and *BG3*, as well as other *β-1,3-glucanase* genes, were found to correspond to one gene in each of the three scaffolds in Nbe.v1 (**Fig. 3B**). No genes homologous to *BG4* and *BG5* were detected in the Nbe.v1. RNA-seq analysis performed in our previous study was mapped against the genome sequence determined in this study using new annotation information. Furthermore, post-grafting gene expression in *N. benthamiana* was confirmed for these *β-1,3-glucanase* genes. Among *N. benthamiana β-1,3-glucanase* genes, *Nbe.v1.s00020g13490* and *Nbe.v1.s00020g16140* were found to be significantly upregulated in both interfamily grafted and self-grafted plants, and *Nbe.v1.s00180g24240*, *Nbe.v1.s00180g24260* and *Nbe.v1.s00020g02080* were slightly upregulated. However, the other genes were scarcely expressed and showed no fluctuation (**Fig. 3C**). *Nbe.v1.s00020g13490* and *Nbe.v1.s00020g16140*, two significantly expressed genes, were located close together in the same Nbe.v1.s00020 scaffold (**Fig. 3B**).

Discussion

Nicotiana benthamiana is one of the most widely distributed experimental plants. The history of *N. benthamiana* was finely reviewed by, in which the origin and derivatives were surveyed literally and genetically. *Nicotiana benthamiana* was first collected >170 years ago; however, there are scant records of its collection from the wild . The survey by revealed the paucity of publicly available genetic resources for *N. benthamiana*. One to five accessions were recorded in the United States Department of Agriculture–Agricultural Research Service National Plant Germplasm System, the Botanical Garden of Nijmegen (Netherlands), the Institut fur Pflanzengenetik und Kulturpflanzenforschung Gatersleben (Germany) and the Australian Plant Genetic Resource Information Service. Nevertheless, there is no proof of the provenance of *N. benthamiana* distributed in the research community. In the study, an amplified fragment length polymorphism analysis was performed for 11 accessions to determine the diversity of varieties, which revealed that the accessions used by the plant research community today might only be a single or collection of closely related accessions. Hence, the accession of *N. benthamiana* sequenced in this study was originally from the University of California, Davis, and might be one of the distributed laboratory accessions. The reported features of *Rdr1* in *N. benthamiana* (Yang et al. 2004) were also identified in our analysis. Loss of function of the *Rdr1* gene of *N. benthamiana* due to the insertion of 72 bases in the ORF is considered to be responsible for its high RNA virus susceptibility. This 72-base insertion was identified in the Nbe.v1 *N. benthamiana* genome reference (**Supplementary Fig. S6**). Furthermore, we confirmed that no other homoeologs corresponding to *Rdr1* were present in *N. benthamiana*. Based on the previous analysis of *N. benthamiana* genes involved in the RNA interference pathway, approximately half were estimated to have a pair of identical and functional homoeologs, approximately one-fourth were estimated to have a pair of non-functional homoeologs and the remaining one-fourth were estimated to have lost their respective homoeologs (Nakasugi et al. 2014, Bally et al. 2018). For *Rdr1* in *N. benthamiana*, the complete loss of homeologs implies that the phenotype of high susceptibility to the RNA virus was acquired by base insertion into one gene.

In a previous study, we showed that *N. benthamiana* could establish interfamily grafting and identified 189 genes whose expression was upregulated in this process (Notaguchi et al. 2020). In the *N. benthamiana* draft sequence (Bombarely et al. 2012), seven of the 189 genes contained nucleotides not yet determined in the CDS, and 38 genes contained unknown nucleotides in the transcribed regions, including introns. Additionally, 50 genes in the promoter region up to 2 kb upstream of the transcription start point were found to have unknown sequences (**Supplementary Table S6**). These unidentified sequences included not only single nucleotides but also many sequences of more than several hundred nucleotides. Particularly in the promoter region, it is difficult to distinguish between a single undetermined nucleotide and a long undetermined region. The uncertain sequence information was an obstacle to the retrieval of promoter sequence information to construct reporter lines for the expression analysis of these genes; however, with the sequence information analyzed in this study, all unidentified regions in the upregulated genes after interfamily grafting were resolved.

Genome sequencing resulted in 1,668 scaffolds, including 21 large scaffolds corresponding to the size of the chromosome. Not all scaffolds had gene-dense regions at both ends, and telomeric repeats were detected only in some scaffolds,

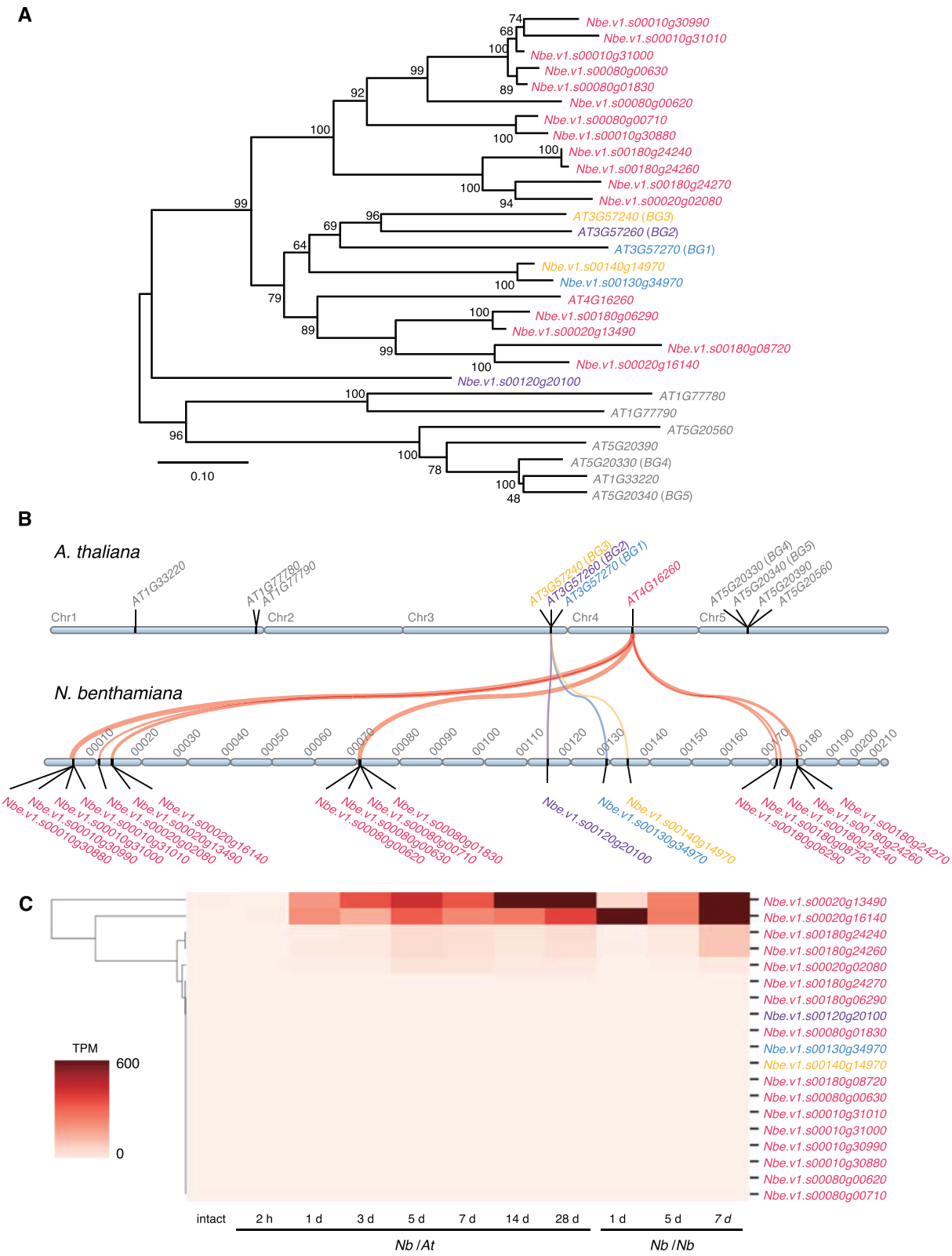

Fig. 3 Phylogenetic analysis and gene expression profiles of *β-1,3-glucanase* genes. (A) Phylogenetic tree of the *β-1,3-glucanase* genes of *A. thaliana* and *N. benthamiana*. The *AT3G57270* gene annotated as *BG1* in *A. thaliana* and the other *A. thaliana* genes whose full-length amino acid sequences are highly homologous to *BG1* were used as queries to extract homologs from the *N. benthamiana* CDS annotations created in this study. The alignment of the *β-1,3-glucanase* genes of *A. thaliana* and *N. benthamiana* was established using ClustalW, and a phylogenetic tree was created using the neighbor-joining method with bootstrap. (B) The new CDS annotation sequence was used as a query to demonstrate the correspondence with the gene with the lowest *E*-value in the *A. thaliana* transcript database using tblastx. (C) Expression heatmaps of the *β-1,3-glucanase* genes of *N. benthamiana*. The expression profiles of intact *N. benthamiana*; 2 h, 1, 3, 5, 7, 14 and 28 d after grafting of *N. benthamiana* onto *A. thaliana* (*Nb*/*At*) and 1, 5 and 7 d after self-grafting of *N. benthamiana* (*Nb*/*Nb*) are displayed.

indicating that they did not cover the entire region as complete chromosomes (**Supplementary Fig. S3**). Full construction of chromosome information using FISH is the subject of future studies. *Nicotiana benthamiana* affiliated with a section *Suaveolentes* is likely to be one of the oldest *Nicotiana* allopolyploids. It was estimated that the section *Suaveolentes* originated >10 MyA through a cross between a section *Noctiflorae* and another section *Sylvestres* (Knapp et al. 2004, Bally et al. 2018). Another study dated the hybridization event at 6 MyA (Clarkson et al. 2017). In this study, diversion between *N. benthamiana* and *N. tabacum* was estimated to be 3–7 MyA. *Nicotiana tabacum* is also an allopolyploid that genome was previously analyzed. The emergence of *N. tabacum* was estimated to have occurred approximately 200,000 years ago (Leitch et al. 2008), and its origin was clearly estimated. A detailed synteny analysis with its ancestral progenitors, *N. sylvestris* (section *Sylvestres*) and *N. tomentosiformis* (section *Tomentosa*), successfully revealed the origin of each chromosome clearly (Sierro et al. 2014, Edwards et al. 2017). The genome analysis of a section *Noctiflorae*, another progenitor of *N. benthamiana* involved, has not yet progressed. Therefore, it was not possible to define the origin of the chromosomes of *N. benthamiana* based on the chromosome sequences of the ancestral species. An alternative comparison with *N. tabacum* showed that the genomic regions derived from the *Sylvestres* section seem to be scattered in many chromosomes. The timing of emergence for *N. benthamiana* was a long time ago and that for *N. tabacum* was recent. Moreover, their progenies were more closed in *N. benthamiana* than those in *N. tabacum*. These could explain the more complex mixture of progenies' chromosomes in the *N. benthamiana* genome compared to the *N. tabacum* genome. Although *N. benthamiana* is an alloploid plant harboring two parent gene copies in most cases, Clustered Regularly Interspaced Short Palindromic Repeats technology, which can target multiple loci using corresponding guide RNAs (Zetsche et al. 2017), will allow us to knock out target genes to investigate the function of genes of interest. Completion of genome construction at the chromosome level will further enhance genetic studies in *N. benthamiana*.

Materials and Methods

Plant materials

The *N. benthamiana* lab strain, derived from the plant research laboratories of the University of California, Davis, was used for whole-genome sequencing. *Nicotiana benthamiana* seeds were surface-sterilized with 5% (w/v) bleach for 5 min, washed three times with sterile water, incubated at 4°C for 3 d and planted on a medium containing half-strength Murashige and Skoog medium, 0.5% (m/v) sucrose and 1% agar. The pH was adjusted to 5.8 with 1 M KOH. Seedlings were grown at 27°C and 100 μmol m^{-2} s^{-1} of illumination under continuous light conditions for *N. benthamiana*. On day 7 after germination, the plants were replanted in a 1:1 mixture of medium and vermiculite. The true leaves were then collected from plants grown for 4 weeks, and genomic DNA was extracted. RNAs for Iso-Seq analysis were also collected from the mature leaves, immature leaves, petioles, stems, stem tips and roots of *N. benthamiana* plants at 4 weeks and flowers at 2 months. The treated sites at 2 h, 3 d and 5 d after *N. benthamiana* grafting and 3 and 5 d after stem wound treatment were also used.

Genome and transcriptome sequencing

PE reads were obtained to estimate genome size and heterozygosity. Genomic DNA was extracted from young leaves using a genomic DNA extraction column (Favorgen Biotech Corp., Ping-Tung, Taiwan). The library was constructed using a Swift 2S® Turbo Flexible DNA Library kit (Cambridge Bioscience, Cambridge, UK) with an expected insert size of 350 bp. Library sequencing was performed using a DNBSEQ-G400RS system (MGI Tech, Shenzhen, China) with a read length of 150 nt (**Supplementary Table S1**).

DNAs for long-read sequences were extracted using a Genomic-tips Kit (QIAGEN, Germantown, MD, USA). A library was prepared using the SMRTbell Express Template Prep Kit 1.0 (Pacific Bioscience, Menlo Park, CA, USA). Size selection of the library was performed using BluePippin (Sage Science, Beverly, MA, USA) to remove DNA fragments <15 kb in length. The library was sequenced using a Sequel II system (Pacific Bioscience) with seven SMRT cells. An Omni-C library was created using the Dovetail Omni-C® Library Preparation Kit (Dovetail Genomics, CA, USA) for chromosome-level scaffolding. Sequencing of the Omni-C library was performed using a DNBSEQ-G400RS system (MGI Tech, Shenzhen, China) with a read length of 150 nt. An Iso-Seq library was constructed using RNAs derived from 12 samples. Iso-Seq Express Template Preparation (Pacific Bioscience) was used for library construction with tag barcodes to identify the samples.

Genome assembly

Genome size and heterogeneity were estimated by *k*-mer frequency analysis using Jellyfish (Marçais and Kingsford 2011). The Hifi reads sequenced by Sequel II were assembled using Hifiasm v0.13. with default parameters (Cheng et al. 2021). The primary contigs were then scaffolded with Omni-C reads using 3D-DNA (Dudchenko et al. 2017). The contact map of the assembled sequences was generated using Juicebox (Durand et al. 2016).

Assembly quality was assessed using the embryophyta_odb10 dataset by BUSCO v5.1.0 (Simão et al. 2015). Repetitive sequences in the assembled genome were identified using RepeatMasker 4.0.7 (http://www.repeatmasker.org/) for known repetitive sequences registered in Repbase (https://www.girinst.org/repbase/) and de novo repetitive sequences defined by RepeatModeler 1.0.11 (http://www.repeatmasker.org/RepeatModeler).

Gene prediction and annotation

The Iso-Seq reads were clustered using the Iso-Seq 3 pipeline implemented in SMRT Link (https://www.pacb.com/support/software-downloads/) and mapped onto the assembled genome sequences using Minmap2 (Li 2018). The obtained isoform sequences were then collapsed into non-redundant isoform sequences using Cupcake ToFU (https://github.com/Magdoll/cDNA_Cupcake). The ORFs on the collapsed sequences were predicted using ANGEL (https://github.com/PacificBiosciences/ANGEL).

In parallel, de novo gene prediction was performed using BRAKER 2 v2.1.5 on the soft-masked genome sequences generated using RepeatMasker. The Iso-Seq collapsed filtered sequences were used as training sets. The gene sequences predicted by ANGEL and BRAKER2 were then merged based on the results of similarity searches against UniProtKB (https://www.uniprot.org) by DIAMOND v2.0.5 with a more sensitive option (Buchfink et al. 2015) and the results of mapping gene sequences onto the assembled genome. The following conditions were used for the selection of the merged sequences: (I) when gene sequences predicted by ANGEL and BRAKER2 hit against the same reference sequences in UniPortKB, sequences with longer hit lengths were selected; (II) when gene sequences predicted by ANGEL and BRKAER2 did not hit against UniProtKB but mapped the same region on the assembled genome, a longer sequence was selected and (III) when either of the gene sequences predicted by ANGEL and BRAKER2 mapped onto the assembled genome, the sequence was selected.

The merged gene sequences were conducted to the following multiple similarity searches: (I) TBLASTN or BLASTP searches against the collapsed filtered Iso-Seq sequences obtained in this study, peptide sequences of *N. benthamiana* (v1.01) (Bombarely et al. 2012) and *N. tabacum* (Nitab4.5) (Edwards et al. 2017) with similarity ≥ 75% and *E*-value ≤1e−20; (II) BLASTP searched against peptide sequences of *A. thaliana* (Araport 11) (Cheng et al. 2017) with similarity ≥25% and *E*-value ≤1e−20; (III) similarity searches against UniportKB by DIAMOND with a more sensitive option with similarity ≥25%, *E*-value ≤1e−20 and (IV) EggNOG search with *E*-value ≤1e−20 (Huerta-Cepas et al. 2018). The genes that met the above-mentioned conditions were classified into HC genes. The genes were classified into transposable elements (TEs) with keywords of TEs. The remaining genes were classified into LC genes. The HC genes were considered as intrinsic genes.

Synteny and orthogroup analyses with other Solanaceae

Genome sequence comparisons were performed using D-Genies (Cabanettes and Klopp 2018). The predicted gene sequences were clustered using the OrthoFinder v2.5 (Emms and Kelly 2019) with those of *N. tabacum* (Nitab4.5), *N. sylvestris* (The National Center for Biotechnology Information accession: GCF_000393655.1) (Sierro et al. 2013), *Petunia axillaris* (v1.6.2) (Bombarely et al. 2016), *S. lycopersicum* (ITAG4.0) (The Tomato Genome Consortium 2012) and *A. thaliana* (Araport 11) (Cheng et al. 2017). Phylogenetic relationships among them were inferred by the Species Tree Inference from All Genes (STAG) method (Emms and Kelly 2018) using OrthoFinder. The phylogenetic tree was constructed by MEGA 7.0.14 (Kumar et al. 1994).

Quantitative analysis of the transcriptome

Grafted or intact plants were harvested at the respective time points. Approximately 10–15 mm of graft junctions or stem tissues were sampled. Each biological replicate comprised pooled tissues from 10 plants. Complementary DNA libraries were prepared using an Illumina TruSeq Stranded Total RNA kit with Ribo-Zero Plant or the Breath Adapter Directional sequencing method (Townsley et al. 2015, Ichihashi et al. 2018) and sequenced for an 86-bp single end with an Illumina NextSeq 500 platform (Illumina, San Diego, CA, USA). Gene expression levels (transcripts per million) were estimated using StringTie version 2.2.0 (Pertea et al. 2015). Details of the RNA preparation and data preprocessing are described in a previous report (Notaguchi et al. 2020).

Supplementary Data

Supplementary data are available at *PCP* online.

Data Availability

The genome assembly data, annotations and gene models are available at the NbenBase (https://nbenthamiana.jp). The obtained genome and transcriptome sequences are available in the DNA Data Bank of Japan (www.ddbj.nig.ac.jp), under the BioProject accession number PRJDB13789.

Funding

Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (22K06181 to K.K. and Y.N., 20H03273 to K.K. and M.N., 21H00368 and 21H05657 to M.N.); Japan Science and Technology Agency (JPMJTR194G to M.N.); New Energy and Industrial Technology Development Organization (JPNP20004 to M.N.).

Acknowledgements

We acknowledge technical assistance by Akiko Watanabe, Takaharu Kimura, Yoshie Kishida, Hisano Tsuruoka, Chiharu Minami, Akiko Komaki, Akiko Obara, Rie Aomiya and Taeko Shibazaki of the Kazusa DNA Research Institute and Miki Matsumoto of Nagoya University.

Author Contributions

K.K., Y.N., S.I. and M.N. conceived the research and designed the experiments. K.K., H.H. and K.S. performed the genome sequencing and analyzed data with the support of Y.N., S.I. and M.N. K.K., H.H., Y.N., S.I. and M.N. wrote the manuscript. K.K., Y.T. and Y.N. conceived the webpage interface and handled the data repository at National Institute of Genomics.

Disclosures

The authors have no conflicts of interest to declare.

References

Bally, J., Jung, H., Mortimer, C., Naim, F., Philips, J.G., Hellens, R., et al. (2018) The rise and rise of *Nicotiana benthamiana*: a plant for all reasons. *Annu. Rev. Phytopathol.* 56: 405–426.

Bombarely, A., Moser, M., Amrad, A., et al. (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* 2: 1–9.

Bombarely, A., Rosli, H.G., Vrebalov, J., Moffett, P., Mueller, L.A. and Martin, G.B. (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol. Plant. Microbe. Interact.* 25: 1523–1530.

Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12: 59–60.

Cabanettes, F. and Klopp, C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6: e4958.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. and Li, H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18: 170–175.

Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89: 789–804.

Clarkson, J.J., Dodsworth, S. and Chase, M.W. (2017) Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (*Solanaceae*). *Plant Syst. Evol.* 303: 1001–1012.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., et al. (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356: 92–95.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., et al. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3: 99–101.

Edwards, K.D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A.D., Bombarely, A., et al. (2017) A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genom.* 18: 448.

Emms, D.M. and Kelly, S. (2018) (preprint) STAG: species tree inference from all genes. *bioRxiv* 267914. 10.1101/267914.

Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20: 238.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., et al. (2018) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47: D309–D314.

Ichihashi, Y., Fukushima, A., Shibata, A. and Shirasu, K. (2018) High impact gene discovery: simple strand-specific mRNA library construction and differential regulatory analysis based on gene co-expression network. *Methods Mol. Biol.* 1830: 163–189.

Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.* 12: 656–664.

Knapp, S., Chase, M.W. and Clarkson, J.J. (2004) Nomenclatural changes and a new sectional classification in *Nicotiana* (Solanaceae). *Taxon* 53: 73–82.

Kourelis, J., Kaschani, F., Grosse-Holz, F.M., Homma, F., Kaiser, M. and van der Hoorn, R.A.L. (2018) (preprint) Re-annotated *Nicotiana benthamiana* gene models for enhanced proteomics and reverse genetics. *bioRxiv* 373506. 10.1101/373506.

Kumar, S., Tamura, K. and Nei, M. (1994) MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput. Appl. Biosci.* 10: 189–191.

Kurotani, K., Huang, C., Okayasu, K., Suzuki, T., Ichihashi, Y., Shirasu, K., et al. (2022) Discovery of the interfamily grafting capacity of *Petunia*, a floricultural species. *Hort. Res.* 9: uhab056.

Kurotani, K. and Notaguchi, M. (2021) Cell-to-cell connection in plant grafting—molecular insights into symplasmic reconstruction. *Plant Cell Physiol.* 62: 1362–1371.

Leitch, I.J., Hanson, L., Lim, K.Y., Kovarik, A., Chase, M.W., Clarkson, J.J., et al. (2008) The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann. Bot.* 101: 805–814.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094–3100.

Mamedov, T., Yuksel, D., Ilgın, M., Gürbüzaslan, I., Gulec, B., Mammadova, G., et al. (2021) Production and characterization of nucleocapsid and RBD cocktail antigens of SARS-CoV-2 in *Nicotiana benthamiana* plant as a vaccine candidate against COVID-19. *Vaccines* 9: 1337.

Marçais, G. and Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764–770.

Naim, F., Nakasugi, K., Crowhurst, R.N., Hilario, E., Zwart, A.B., Hellens, R.P., et al. (2012) Advanced engineering of lipid metabolism in *Nicotiana benthamiana* using a draft genome and the V2 viral silencing-suppressor protein. *PLoS One* 7: e52717.

Nakasugi, K., Crowhurst, R., Bally, J. and Waterhouse, P. (2014) Combining transcriptome assemblies from multiple de novo assemblers in the allotetraploid plant *Nicotiana benthamiana*. *PLoS One* 9: e91776.

Narayan, R.K.J. (1987) Nuclear DNA changes, genome differentiation and evolution in *Nicotiana* (Solanaceae). *Plant Syst. Evol.* 157: 161–180.

Notaguchi, M., Kurotani, K., Sato, Y., Tabata, R., Kawakatsu, Y., Okayasu, K., et al. (2020) Cell–cell adhesion in plant grafting is facilitated by β-1,4-glucanases. *Science* 369: 698–702.

Pertea, M., Pertea, G., Antonescu, C., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33: 290–295.

Schiavinato, M., Marcet-Houben, M., Dohm, J.C., Gabaldón, T. and Himmelbauer, H. (2020) Parental origin of the allotetraploid tobacco *Nicotiana benthamiana*. *Plant J.* 102: 541–554.

Sierro, N., Battey, J.N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., et al. (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. commun.* 5: 1–9.

Sierro, N., Battey, J.N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., et al. (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* 14: 1–17.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212.

The Tomato Genome Consortium. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635–641.

Tong, Z., Xiao, B., Jiao, F., Fang, D., Zeng, J., Wu, X., et al. (2016) Large-scale development of SSR markers in tobacco and construction of a linkage map in flue-cured tobacco. *Breed. Sci.* 66: 381–390.

Townsley, B.T., Covington, M.F., Ichihashi, Y., Zumstein, K. and Sinha, N.R. (2015) BrAD-seq: breath adapter directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. *Front. Plant Sci.* 6: 366.

van Herpen, T.W.J.M., Cankar, K., Nogueira, M., Bosch, D., Bouwmeester, H.J. and Beekwilder, J. (2010) *Nicotiana benthamiana* as a production platform for artemisinin precursors. *PLoS One* 5: e14222.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40: e49.

Wheeler, H.-M. (1945) A contribution to the cytology of the Australian-South Pacific species of *Nicotiana*. *Proc. Natl. Acad. Sci. U. S. A.* 31: 177–185.

Yang, S.J., Carter, S.A., Cole, A.B., Cheng, N.H. and Nelson, R.S. (2004) A natural variant of a host RNA-dependent RNA polymerase is associated with increased susceptibility to viruses by *Nicotiana benthamiana*. *Proc. Nat. Acad. Sci. U. S. A.* 101: 6297–6302.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., et al. (2017) Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35: 31–34.