Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Original Article

The dynamics of genome size and GC contents evolution in genus *Nicotiana*

A dinâmica do tamanho do genoma e a evolução dos conteúdos de GC no gênero *Nicotiana*

Z. Hussain[a,b*], Y. Sun[a], S. H. Shah[c], H. Khan[d], S. Ali[b], A. Iqbal[b], M. A. Zia[e] and S. S. Ali[b]

[a]Chinese Academy of Agricultural Sciences, Qingdao, Shandong, China
[b]University of Swat, Centre for Biotechnology and Microbiology, Mingora, Swat, Khyber Pukhtunkhwa, Pakistan
[c]Allama Iqbal Open University, Faculty of Sciences, Department of Agricultural Sciences, Islamabad, Pakistan
[d]Quid-e-Azam University, Department of Biotechnology, Islamabad, Pakistan
[e]National Agricultural Research Centre – NARC, National Institute for Genomics and Advanced Biotechnology – NIGAB, Islamabad, Pakistan

Abstract

Hybridization and Polyploidization are most common of the phenomenon observed in plants, especially in the genus *Nicotiana* leading to the duplication of genome. Although genomic changes associated with these events has been studied at various levels but the genome size and GC content variation is less understood because of absence of sufficient genomic data. In this study the flow cytometry technique was used to uncover the genome size and GC contents of 46 *Nicotiana* species and we compared the genomic changes associated with the hybridization events along evolutionary time scale. The genome size among *Nicotiana* species varied between 3.28 pg and 11.88 pg whereas GC contents varied between 37.22% and 51.25%. The tetraploid species in genus *Nicotiana* including section *Polydiclae*, *Repandae*, *Nicotiana*, *Rustica* and *Sauveolentes* revealed both up and downsizing in their genome sizes when compared to the sum of genomes of their ancestral species. The genome sizes of three homoploid hybrids were found near their ancestral species. Loss of large genome sequence was observed in the evolutionary more aged species (>10 Myr) as compared to the recently evolved one's (<0.2 Myr). The GC contents were found homogenous with a mean difference of 2.46% among the *Nicotiana* species. It is concluded that genome size change appeared in either direction whereas the GC contents were found more homogenous in genus *Nicotiana*.

Keywords: *Nicotiana*, genome size, GC contents, flow cytometry (FCM), evolution.

Resumo

A hibridização e a poliploidização são os fenômenos mais comuns observados em plantas, principalmente no gênero Nicotiana, levando à duplicação do genoma. Embora as mudanças genômicas associadas a esses eventos tenham sido estudadas em vários níveis, o tamanho do genoma e a variação do conteúdo de GC são menos compreendidos devido à ausência de dados genômicos suficientes. Neste estudo, a técnica de citometria de fluxo foi usada para descobrir o tamanho do genoma e o conteúdo de GC de 46 espécies de Nicotiana, e comparamos as mudanças genômicas associadas aos eventos de hibridização ao longo da escala de tempo evolutiva. O tamanho do genoma entre as espécies de Nicotiana variou entre 3,28 pg e 11,88 pg, enquanto os conteúdos de GC variaram entre 37,22% e 51,25%. As espécies tetraploides do gênero Nicotiana, incluindo as seções Polydiclae, Repandae, Nicotiana, Rustica e Sauveolentes, revelaram aumento e redução do tamanho do genoma quando comparados à soma dos genomas de suas espécies ancestrais. Os tamanhos do genoma de três híbridos homoploides foram encontrados perto de suas espécies ancestrais. A perda da grande sequência do genoma foi observada nas espécies evolutivas mais velhas (> 10 Myr) em comparação com as que evoluíram recentemente (< 0,2 Myr). Os teores de GC foram homogêneos com diferença média de 2,46% entre as espécies de Nicotiana. Conclui-se que a mudança no tamanho do genoma apareceu em ambas as direções, enquanto os conteúdos de GC foram encontrados mais homogêneos no gênero Nicotiana.

Palavras-chave: *Nicotiana*, tamanho do genoma, conteúdo de GC, citometria de fluxo (FCM), evolução.

*e-mail: zahid@uswat.edu.pk
Received: November 7, 2020 – Accepted: May 11, 2021

1. Introduction

Polyploid and homoploid hybridization are two important evolutionary phenomena involved at species levels. These processes have regularly contributed in diversification of plant species. Evolutionary consequences associated with hybridization events have been studied at various levels such as chromosomal rearrangements, repetitive DNA sequence evolution, genome size change, and diploidization (Hegarty and Hiscock, 2008; Baack et al., 2005; Leitch et al., 2008; Renny-Byfield et al., 2011; Renny-Byfield et al., 2013). The genome size changes associated with hybridization and polyploidization and genomic GC contents variation has been the subject of immense interest. Recently, numerous studies have provided novel insights into the potential basis of genome size evolution in plants (Bennett and Leitch, 2011; Veselý et al., 2012). Similarly, the range of GC contents in major plant species studied is narrow except for grasses that exhibit a remarkable GC content heterogeneity (Barow and Meister, 2003; Šmarda et al., 2012). While it has also been shown that dynamics and magnitude of GC base composition is persistently lacking in plants (Tatarinova et al., 2010; Serres-Giardi et al., 2012).

Polyploidization can induce rapid genomic changes, including the gain or loss of DNA, but the magnitude and timing of such changes are not well understood (Baack et al., 2005). In this regard, the *Nicotiana* genus is more suitable candidate as this genus consists of several sections of allotetraploids formed at different times from their diploid ancestors and the estimates of ages of each section is also well studied (Clarkson et al., 2005; Leitch et al., 2008). The genus *Nicotiana* also offers the section *Suaveolentes* where multiple chromosome fusions resulted in chromosome number reduction (Chase et al., 2003; Leitch et al., 2008). On the other hand, homoploid hybridization has also significant role in contributing to species diversity in plants. While the genomic changes associated with homoploid hybrid speciation has been previously reported in *Helianthus* and *Paeonia* (Rieseberg and Willis, 2007; Paun et al., 2009) but not in the homoploid hybrids species of genus *Nicotiana* (Clarkson et al., 2010; Kelly et al., 2010). Such study in plants will require that genome size, evolutionary origin and age of all polyploid groups must be known (Lim et al., 2007). While at the same time, a very scattered and dichotomist viewpoint has emerged on the pattern of GC contents evolution in plants through the studies of few representative species of monocots and dicots (Wong et al., 2002; Wang and Roossinck, 2006; Serres-Giardi et al., 2012). However, until now, the GC content has been reported for limited plant species (Veselý et al., 2012), especially in the lower taxonomic groups (Genus level). In this regard, flow cytometry offers a reliable method to estimate GC contents (Šmarda et al., 2012).

The genus *Nicotiana* comprises of 76 species among which 35 are allotetraploid and the rest of them are diploid species (including the recently identified 3 homoploid hybrids). The polyploid and homoploid hybrid species in the genus *Nicotiana* have been evolved through different polyploidization and interspecific hybridization events respectively, involving different diploid ancestral species (Knapp et al., 2004). *Nicotiana* is a model genus to understand polyploidization in plants and it has long been used to explore many of the evolutionary processes involved in allopolyploidization events (Clarkson et al., 2005; Leitch et al., 2008). So far, the parental lineage of almost all the *Nicotiana* allotetraploid (including section *Suaveolentes*) and homoploid hybrid species (*N. linearis*, *N. spengzii* and *N. glauca*) has been documented based on morphological (Goodspeed, 1954), cytological, plastid sequence data and the most recently evolved nuclear coding sequence data (Clarkson et al., 2010; Kelly et al., 2010; Kelly et al., 2013). The age of each tetraploid section in the genus *Nicotiana* has also been documented (Knapp et al., 2004; Clarkson et al., 2005; Clarkson et al., 2010). In addition, the number of genomic resources for *Nicotiana* species are increasing over the last few years, with the draft genome sequence available for the most important species of *Nicotiana* (*N. Banthamiana*, *N. tabacum*, *N. tomentosiformis* and *N. Sylvestris*), along with the increasing knowledge on diversity of repetitive DNA elements (Koukalova et al., 2010), karyotpic studies (Marks et al., 2011) and genome size evolution (Leitch et al., 2008, Renny-Byfield et al., 2011; Renny-Byfield et al., 2013).

Given the importance of this genus, with such opportunities not available in other angiosperms group, this study was carried out to estimate the genome size and GC contents of the different *Nicotiana* species by flow cytometry and study the overall extent of genome size (up- and downsizing) and GC content variation.

2. Materials and Methods

2.1. Plant material

The seeds of all *Nicotiana* species (Table 2) were provided by the germplasm bank of Tobacco research institute of Chinese Academy of Agricultural Sciences, Beijing, China. The seeds of standard plants were obtained from Jaroslav Doležel, Experimental Institute of Botany, Czech Republic (Table 1). All the species of *Nicotiana* and standard plants were grown under controlled conditions in glass house and leaf samples were harvested for analysis.

Table 1. Plant standards for genome size estimation.

Standard species*	Genome size (pg)	GC content (%)
Glycine max Merr. 'Polanka' 46	2.50	63.6
Zea mays L. 'CE-777' 47	5.43	52.8
Pisum sativum L. 'Ctirad' 33	9.09	61.5

*The standard material were obtained from one source and the c-values assigned to the above plant standards were based on single primary internal reference standard.

Table 2. Genome size, nucleotides composition and average DNA contents per chromosome of 46 different Nicotiana species.

Species	Chromosome number (n)	Genome size (2C)±SE Mean	AT %	GC%	Average DNA contents per chromosome	Standard plant
Sect. Sualveolentes						
N. occidentalis	21	5.83±0.05	62.40	37.59	0.14	Pisum sativum
N. debneyi	24	9.15±0.04	59.24	40.76	0.20	Maize
N. exigua	16	6.95±0.02	59.33	40.67	0.22	Pisum sativum
N. goodspeedii	16	6.31±0.03	61.45	38.55	0.20	Glycine max
N. africana	23	9.66±0.04	51.40	48.60	0.21	Maize
N. gossei	18	6.89±0.05	60.21	39.79	0.19	Maize
N. suaveolens	16	11.88±0.02	59.79	40.21	0.37	Maize
N. rosulata	20	5.42±0.03	80.09	39.90	0.14	Pisum sativum
N. rotundifolia	16	5.44±0.02	82.19	37.22	0.17	Pisum sativum
N. benthamiana	19	6.92±0.02	58.67	41.33	0.18	Maize
N. Simulans	20	3.28±0.09	65.28	37.72	0.08	Glycine max
N. excelsior	19	6.65±0.07	58.72	41.28	0.17	Maize
N. rotundifolia	16	5.33±0.03	59.47	40.53	0.17	Pisum sativum
N. amplexicaulis	18	6.92±0.04	58.94	41.06	0.19	Maize
Sect. Repandae						
N. nudicaulis	24	7.05±0.07	60.01	39.99	0.15	Maize
N. repanda	24	9.98±0.01	58.35	41.65	0.21	Maize
N. nesophila	24	10.33±0.03	48.75	51.25	0.22	Maize
N. stocktonii	24	10.00±0.05	59.03	40.97	0.21	Maize
Sect. Polydicliae						
N. quadrivalvis	24	10.50±0.07	61.17	38.83	0.22	Maize
N. clevelandii	24	7.76±0.12	59.38	40.62	0.16	Maize
Sect. Sylvestris						
N. sylvestris	12	5.81±0.01	60.02	39.98	0.24	Glycine max
Sect. Tomentosae						
N. tomentosiformis	12	5.52±0.03	59.87	40.13	0.23	Glycine max
N. kawakamii	12	6.34±0.03	61.55	38.45	0.26	Pisum sativum,
N. otophora	12	5.99±0.01	59.69	40.31	0.25	Glycine max
Sect. Paniculatae						
N. benavidesii	12	6.11±0.04	61.34	38.66	0.25	Pisum sativum,
N. knightiana	12	6.57±0.01	61.00	38.99	0.27	Pisum sativum,
N. paniculata	12	6.40±0.05	75.79	37.50	0.27	Pisum sativum
Sect. Undulatae						
N. undulata	12	10.30±0.05	57.95	42.05	0.43	Maize
N. glutinosa	12	4.71±0.01	59.97	40.03	0.20	Glycine max
Sect. Petunioides						
N. miersii	12	5.82±0.04	60.31	39.69	0.24	Glycine max
N. attenuta	12	6.95±0.01	59.08	40.92	0.29	Maize
N. acuminata	12	5.45±0.02	60.41	39.59	0.23	Glycine max
N. linearis	12	6.50±0.05	59.90	40.10		Glycine max
N. spegazzinii	12	7.11±0.01	59.04	40.96		Maize
Sect. Alatae						
N. bonariensis	9	4.45±0.04	65.63	38.37	0.25	Glycine max
N. alata	9	4.53±0.02	59.45	40.54	0.25	Glycine max
N. alata (Red flowers)	9	5.49±0.02	79.89	39.12	0.30	Pisum sativum
N. langsdorfii	9	6.82±0.08	59.97	40.03	0.379	Maize
N. longiflora	10	5.74±0.04	58.63	41.37	0.29	Glycine max

Table 2. Continued...

Species	Chromosome number (n)	Genome size (2C)±SE Mean	AT %	GC%	Average DNA contents per chromosome	Standard plant
N. plumbaginifolia	10	5.46±0.02	61.17	38.82	0.27	Pisum sativum
Sect. Noctiflorae						
N. noctiflora	12	9.53±0.02	57.53	42.47	0.40	Maize
N. petunioides	12	5.30±0.01	61.80	38.20	0.22	Pisum sativum
N. acaulis	12	6.20±0.04	60.89	39.11	0.25	Pisum sativum
N. glauca	12	6.85±0.05	59.31	40.69	0.28	Pisum sativum
Sect. Rustica						
N. rustica	24	10.82	57.03	42.97	0.22	Maize
Sect. Nicotiana						
N. tabacum	24	9.77±0.04	57.63	42.37	0.20	Maize

2.2. Sample preparation

Fresh leaf sample from both standard and sample (50 mg) was co-chopped in plastic Petri dish by sharp razor blade in 500 µl of ice cold Otto-1 buffer supplemented with 2% mercaptoethanol. The suspension of nuclei was filtered through 30 µm of disposable filter (Partec) and stained with 2 ml of respective flourochrome buffer for 5 minutes in dark. Staining buffer for genome size estimation consisted of 1 ml Otto-II buffer supplemented with 50 µg propidium iodide and 50 ul Rnase I whereas for AT-specific staining 1 ml Otto-II buffer was supplemented with 5 µl of DAPI.

2.3. Genome size and nucleotide contents estimation

The nuclear suspensions stained with propidium iodide were subjected to flow Cytometer (Cube Partec, Germany). The channels were set into a proper position on the abscissa and different parameters like threshold level and gain value were adjusted with a flow speed of 0.5µl/sec and approximately 10, 000 nuclear particles were measured. The genome size was calculated by the method described (Doležel et al., 2007). The GC contents were calculated by the most widely accepted equations (eqns 7, 8) described by Barow & Meister (Barow and Meister, 2002). The calculations were performed with binding length of DAPI=4, as recommended by Barow & Meister. The average DNA contents per chromosome was calculated by dividing the genome size (2C) value by total number of chromosomes. The entire samples were analyzed in three replications with CV value of less than 5%.

2.4. Evaluation of genome size changes in tetraploid species

The expected genome size values of tetraploid species were calculated by the sum of genome sizes (1C flow cytometry) of their two-ancestral species that formed them whereas the observed values for the same tetraploid species were those obtained by flow cytometry. In all the polyploid cases, the extant diploid species are not what exactly formed the tetraploid species but these species are the closest living relatives to the diploids that formed them. The genome size changes of all the tetraploid section were also analyzed on evolutionary time scale, as the ages of all sections are known.

2.5. Statistical analysis

All samples were analyzed in three replications and the mean values along with the standard error were calculated. Boxplot distribution analyses of genome size and GC contents were performed on different polyploid sections based on the evolutionary age of each section (Figures 2 and 5). Scatter line plot were carried out on genome size vs GC contents (Figure 6). All the statistical analysis were carried out by MINITAB 16 statistical package. The graphs and figures were made by Origin 2015.

3. Results

3.1. Genome size data

The genome sizes, genomic base compositions (AT+GC) and average chromosome size of 46 different diploids and allotetraploid species are listed in Table 2.

3.2. Genome size changes along the evolutionary time scale in genus Nicotiana

The observed vs expected genome size values of the 8-tetraploid species were compared and both genome up and downsizing were observed in all tetraploid species (Figure 1). *N. tabacum*, *N. rustica*, *N. clevelandilii*, *N. nudicaulis* reveals genome downsizing whereas *N. quadrivalvis*, *N. repanda*, *N. nesophila*, *N. stocktoni* showed genome upsizing. The sum of 1C values of all the 14 species in section *Suaveolentes* (average) was compared with the sum of 1C values of their two-ancestral species (Figure 1). The observed vs expected genome size of this newly studied section *Suaveolentes* reveals a huge amount of genome downsizing. The section *Suaveolentes* originates through allopolyploidization that involves ancestral member of the section *Sylvestris* as paternal progenitor and a member of either section *Petunioides* or section *Noctiflorae* or a hypothetical hybrid species between these two sections as maternal progenitor (Kelly et al., 2013).

The three homoploid hybrid species (*N. linearis*, *N. spengzii* and *N. glauca*) and their possible ancestral species showed little differences among their genome sizes except for *N. noctiflora* (Figure 1). The genome size of these hybrid species ranges from 6.50 pg to 7.11 pg whereas

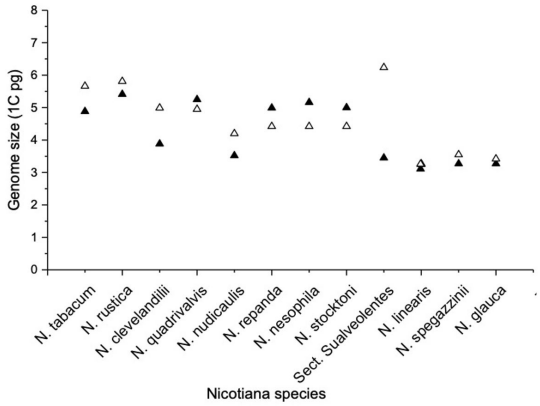

Figure 1. Observed (black triangles) and expected (white triangles) genome size values of the tetraploid and homoploid hybrid species (*N. linearis*, *N. spengzii* and *N. glauca*) evolved from their ancestral diploid species. The observed values are 1C genome estimated by flow cytometry and expected values are the sum of 1C parental genome.

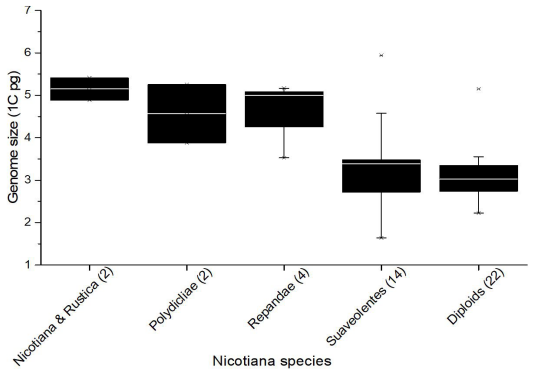

Figure 2. Boxplot plot distribution of genome size estimates (1C in pg) of different tetraploid sections over evolutionary time scale (time scale is represented from left to right on x-axis from most recently evolved species to the older one). Genome size estimates of each tetraploid section of genus Nicotiana are represented over evolutionary timescale along with average genome size of diploids progenitors.

approximately the same range of 5.30 pg to 6.95pg was observed in their ancestral species.

The loss in genome size in the five tetraploid sections was found directly proportional to the age of each section. The more recently evolved tetraploid sections i.e. *Nicotiana* and *Rustica*, revealed small amount of genome size loss whereas the section *Suaveolentes* showed large amount of genome size loss (Figure 2).

3.3. GC content variation in genus Nicotiana

The average GC contents of the 46 species in genus *Nicotiana* were estimated by flow cytometry (Table 2). Ascending pattern was observed in the GC contents from diploid to tetraploid species with a mean difference of 2.46% (Figure 5). The boxplot analysis of GC contents reveals

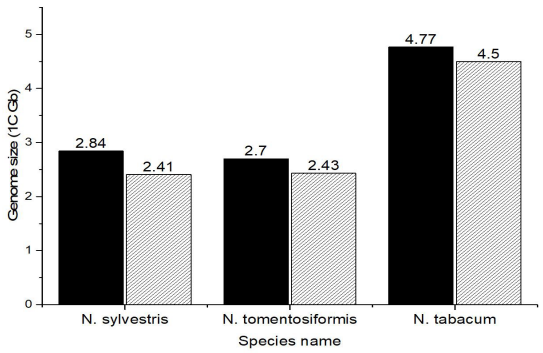

Figure 3. Comparison of genome size estimate with sequencing results. Black bars represent the genome size estimates by Flow cytometry whereas the line bars represent genome size estimates from 17-mer sequencing results (Unpublished data).

uniformity in the pattern of GC contents distribution but the more recently evolved species (*N. tabacum and N. rustica*) showed comparatively high GC contents (Figure 5). The mean difference in GC contents observed among all *Nicotiana* sections is 2.46% representing a more homogenous content within genus *Nicotiana*.

4. Discussion

4.1. Reliability of the genome size data

The genome size estimates of *N. sylvestris*, *N. tomentosiformis* and *N. tabacum* based on flow cytometry and 17-mer depth distributions of sequence data by Tobacco Research Institute (unpublished data) are diagrammatically represented (Figure 3). Our genome size estimates by flow cytometry were found 6-15% higher than 17-mer based sequencing results of the three species. For instance, the recently sequenced genomes of *N. sylvestris* and *N. tomentosiformis* were estimated as 2.41 Gb (2.63 pg) and 2.43 Gb (2.68 pg) respectively using a 17-mer distribution, smaller than expected 1C value estimated by flow cytometry (Sierro et al., 2013). Genome size estimates (FCM) of Arabidopsis (157 Mb) were found 25% larger than the Arabidopsis genome sequencing estimates of ~125 Mb. The discrepancy among genome size estimates might arise due to the un-sequenced gap in the heterochromatin region, telomere or nucleolar region (Bennett et al., 2003). Furthermore, the study of repetitive content in the 727 Mb potato genome assemblies reveals that much of the unassembled genome sequences were composed of repeats (Xu et al., 2011). Fortunately, considerable benefits can be achieved by bridging the genome size and sequence data, as uniformity exist between the two estimates.

Our study generated genome size values of 46 species of genus *Nicotiana* among which the values of 14 species were found in parallel with that of previous study (Leitch et al., 2008), with little differences observed among three species i.e. *N. tabacum*, *N. attenuata N. quadrivalvis and N. repanda* (Figure 4). Significant differences were observed

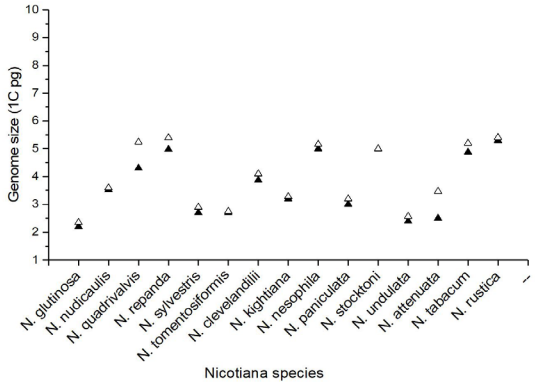

Figure 4. Comparison of genome size estimates between our study (black triangles) with 15 species estimated by Leitch et al., 2008 (white triangles).

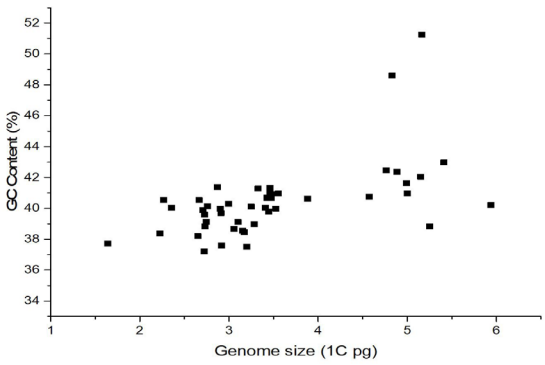

Figure 6. Scatter line plot of genome size vs GC contents in genus *Nicotiana* (correlation coefficient of 0.56).

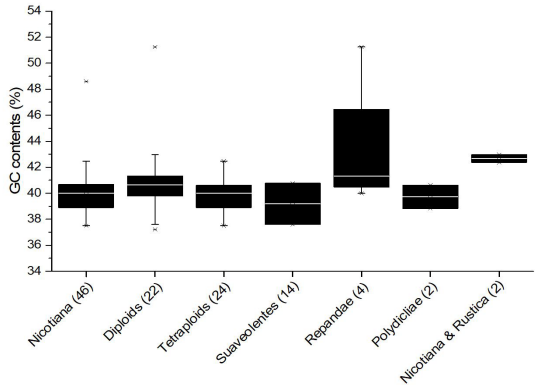

Figure 5. Boxplot plot distribution of GC (%) contents represents GC contents among 46 Nicotiana, 22 diploids, 24 tetraploids and each tetraploid sections of *Nicotiana.*

between our genome size estimates and 23 species listed by Narayan et al., 1987. However, the exact cause of such huge differences between the two studies might be methodological error as the previous study (Narayan, 1987) used Feulgen photometry for genome size estimation whereas flow cytometry has been emerged as a method of choice for genome size in the last decade (Doležel and Bartos, 2005). Furthermore, the genome size estimates for the standard plants used in the previous study (Narayan, 1987) were not accurate because sequenced genomes were not available at that time.

4.2. Genome size estimation and genomic changes along evolutionary time scale in the tetraploid species

Our study indicated differences in the extent of genome up- and downsizing with that of Leitch et al., 2008 but the direction of genome size change was found similar except for *N. clevandii.* Next generation sequencing data of the section *Repandae* also reveals both genome up and downsizing in the section *Repandae* (Renny-Byfield et al., 2013). Frequent loss of genomic sequences in polyploid species and genome contraction seems to be a general response to polyploidization (Leitch and Bennett, 2004; Renny-Byfield et al., 2011, Yang et al., 2011) whiles a few cases of genome size expansion has also been reported (Bennett and Leitch, 2005; Leitch et al., 2008). Nevertheless, some studies reported no DNA loss (Ozkan et al., 2006; Mestiri et al., 2010).

The observed vs expected genome size of the section *Suaveolentes* reveals a huge amount of genome downsizing. The exact cause of such a huge amount of genomic DNA loss is not clear until now but the dysploid reduction in chromosomes number, largely occur in section *Suaveolentes* due to the fusion of chromosome and might be one of the possible reason (Clarkson et al., 2004). The section *Suaveolentes* revealed a huge amount of genome size reduction among the polyploidy species because this is the oldest section among polyploid species in genus *Nicotiana* with an age of approximately 10 Myrs. The evolutionary age of each polyploid section of genus *Nicotiana* has been documented in the previous studies (Clarkson et al., 2005; Leitch et al., 2008). The extent of DNA sequence divergence encountered in polyploids is more dependent on the age of the species with genome turnover more evident in older species (Lim et al., 2007).

The homoploid hybrid species (*N. linearis*, *N. spengzii* and *N. glauca*) has been recently identified and evolved by hybridization of members from the section *Noctiflora* and *Petunoides* (Clarkson et al., 2010; Kelly et al., 2010). The genome size of *N. linearis*, *N. spengzii* and *N. glauca* were 6.50 pg, 7.11 pg and 6.85 pg respectively whereas approximately the same range of 5.30 pg to 6.95pg was observed in their ancestral species. As opposed to our findings, genome size expansion had been observed in three homoploid hybrid species in *Helianthus* with 50% more nuclear DNA than their parental species (Baack et al., 2005).

4.3. GC content variation in genus Nicotiana

Several studies in various organisms including plants have reported an increase in GC contents from diploid to tetraploid species. The more recent studies on seed plant reveal the GC poor and homogenous pattern of diploid species to a more heterogeneous and GC rich polyploid species (Serres-Giardi et al., 2012). The pattern of GC contents was tested on a narrower range in the genus *Nicotiana*. Our study indicates more homogenous pattern of GC contents among the diploid ancestors and

the polyploidy progenitors with median value of 39.97% and 41.28% respectively (Figure 5). The interquartile range of GC contents among the whole range of species were found 2.46%. The pattern of GC contents in the genus *Nicotiana* was found similar to that of the previous study (Serres-Giardi et al., 2012) but the magnitude of difference was different because their study includes wider range of species from eudicots to monocots. Positive correlation was found between genome size and GC contents with Pearson co-efficient of correlation value of 0.56 (Figure 6).

5. Conclusions

Our study provides a more comprehensive and recent review of genome size estimates of 46 different species of *Nicotiana* in both diploids and tetraploids. Altogether, our study reveals both genome up and downsizing along the evolutionary time scale in genus *Nicotiana*. Genome downsizing were observed in the large and newly studied section of *Suaveolentes* whereas genome size estimates of three homoploid hybrid species were found in similar range to their ancestral species. The genomic loss was found highly correlated to the age of each sections i.e. evolutionary older sections showed high amount of genomic sequence loss as compared to recently evolved sections of genus *Nicotiana*. The GC contents were found strongly correlated with genome size having correlation coefficient of 0.56. The GC contents were found more homogenous in this genus with a mean difference of 2.46%. The GC content also reveals moderate increase in the recently evolved species of section *Nicotiana* and *Rustica*. The sub-genomic processes and specific sequences that generate variation in genome size can only be examined in detail through large-scale comparisons of DNA sequences. Study of total DNA contents (C-value) and individual sequences can provide new spectrum to genome biology with sequence data providing novel insights into genome-size evolution, and with genome-size data being of both practical and theoretical significance for large-scale sequence analysis.

Acknowledgements

We would like to thank Jaroslav Doležel (Institute of experimental botany) for providing the seeds of standard plant materials. Our thanks to plant breeding department of TRIC for providing the seeds of all the Nicotiana species. The research project was funded by Tobacco Research Institute, Chinese Academy of Agricultural Sciences.

References

BAACK, E.J., WHITNEY, K.D. and RIESEBERG, L.H., 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in Helianthus homoploid hybrid species. *The New Phytologist*, vol. 167, no. 2, pp. 623-630. http://dx.doi.org/10.1111/j.1469-8137.2005.01433.x. PMid:15998412.

BAROW, M. and MEISTER, A., 2002. Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry*, vol. 47, no. 1, pp. 1-7. http://dx.doi.org/10.1002/cyto.10030. PMid:11774343.

BAROW, M. and MEISTER, A., 2003. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. *Plant, Cell & Environment*, vol. 26, no. 4, pp. 571-584. http://dx.doi.org/10.1046/j.1365-3040.2003.00988.x.

BENNETT, M.D. and LEITCH, I.J., 2005. Nuclear DNA amounts in angiosperms-progress, problems and prospects. *Annals of Botany*, vol. 95, no. 1, pp. 45-90. http://dx.doi.org/10.1093/aob/mci003. PMid:15596457.

BENNETT, M.D. and LEITCH, I.J., 2011. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany*, vol. 107, no. 3, pp. 467-590. http://dx.doi.org/10.1093/aob/mcq258. PMid:21257716.

BENNETT, M.D., LEITCH, I.J., PRICE, H.J. and JOHNSTON, J.S., 2003. Comparisons with Caenorhabditis (approximately 100 Mb) and Drosophila (approximately 175 Mb) using flow cytometry show genome size in Arabidopsis to be approximately 157 Mb and thus approximately 25% larger than the Arabidopsis genome initiative estimate of approximately 125 Mb. *Annals of Botany*, vol. 91, no. 5, pp. 547-557. http://dx.doi.org/10.1093/aob/mcg057. PMid:12646499.

CHASE, M.W., KNAPP, S., COX, A.V., CLARKSON, J.J., BUTSKO, Y., JOSEPH, J., SAVOLAINEN, V. and PAROKONNY, A.S., 2003. Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). *Annals of Botany*, vol. 92, no. 1, pp. 107–112.

CLARKSON, J.J., KELLY, L.J., LEITCH, A.R., KNAPP, S. and CHASE, M.W., 2010. Nuclear glutamine synthetase evolution in Nicotiana: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. *Molecular Phylogenetics and Evolution*, vol. 55, no. 1, pp. 99-112. http://dx.doi.org/10.1016/j.ympev.2009.10.003. PMid:19818862.

CLARKSON, J.J., KNAPP, S., GARCIA, V.F., OLMSTEAD, R.G., LEITCH, A.R. and CHASE, M.W., 2004. Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Molecular Phylogenetics and Evolution*, vol. 33, no. 1, pp. 75-90. http://dx.doi.org/10.1016/j.ympev.2004.05.002. PMid:15324840.

CLARKSON, J.J., LIM, K.Y., KOVARIK, A., CHASE, M.W., KNAPP, S. and LEITCH, A.R., 2005. Long-term genome diploidization in allopolyploid Nicotiana section Repandae (Solanaceae). *The New Phytologist*, vol. 168, no. 1, pp. 241-252. http://dx.doi.org/10.1111/j.1469-8137.2005.01480.x. PMid:16159337.

DOLEŽEL, J. and BARTOS, J., 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany*, vol. 95, no. 1, pp. 99-110. http://dx.doi.org/10.1093/aob/mci005. PMid:15596459.

DOLEŽEL, J., GREILHUBER, J. and SUDA, J., 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, vol. 2, no. 9, pp. 2233-2244. http://dx.doi.org/10.1038/nprot.2007.310. PMid:17853881.

GOODSPEED, T.H., 1954. *The genus Nicotiana*. Waltham: Cronica Botanica Company.

HEGARTY, M.J. and HISCOCK, S.J., 2008. Genomic clues to the evolutionary success of polyploid plants. *Current Biology*, vol. 18, no. 10, pp. 435-444. http://dx.doi.org/10.1016/j.cub.2008.03.043. PMid:18492478.

KELLY, L.J., LEITCH, A.L., CLARKSON, J.J., KNAPP, S. and CHASE, M.W., 2013. Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (Nicotiana section Suaveolentes). *Evolution; international journal of organic evolution*, vol. 67, no. 1,

pp. 80-94. http://dx.doi.org/10.1111/j.1558-5646.2012.01748.x. PMid:23289563.

KELLY, L.J., LEITCH, A.R., CLARKSON, J.J., HUNTER, R.B., KNAPP, S. and CHASE, M.W., 2010. Intragenic recombination events and evidence for hybrid speciation in Nicotiana (Solanaceae). *Molecular Biology and Evolution*, vol. 27, no. 4, pp. 781-799. http://dx.doi.org/10.1093/molbev/msp267. PMid:19897524.

KNAPP, S., CHASE, M.W. and CLARKSON, J.J., 2004. Nomenclatural changes and a new sectional classification in Nicotiana (Solanaceae). *Taxon*, vol. 53, no. 1, pp. 73-82. http://dx.doi.org/10.2307/4135490.

KOUKALOVA, B., MORAES, A.P., RENNY-BYFIELD, S., MATYASEK, R., LEITCH, A.R. and KOVARIK, A., 2010. Fall and rise of satellite repeats in allopolyploids of Nicotiana over c. 5 million years. *The New Phytologist*, vol. 186, no. 1, pp. 148-160. http://dx.doi.org/10.1111/j.1469-8137.2009.03101.x. PMid:19968801.

LEITCH, I.J. and BENNETT, M.D., 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society. Linnean Society of London*, vol. 82, no. 4, pp. 651-663. http://dx.doi.org/10.1111/j.1095-8312.2004.00349.x.

LEITCH, I.J., HANSON, L., LIM, K.Y., KOVARIK, A., CHASE, M.W., CLARKSON, J.J. and LEITCH, A.R., 2008. The ups and downs of genome size evolution in polyploid species of Nicotiana (Solanaceae). *Annals of Botany*, vol. 101, no. 6, pp. 805-814. http://dx.doi.org/10.1093/aob/mcm326. PMid:18222910.

LIM, K.Y., KOVARIK, A., MATYASEK, R., CHASE, M.W., CLARKSON, J., GRANDBASTIEN, M.A. and LEITCH, A.R., 2007. Sequence of events leading to near-complete genome turnover in allopolyploid Nicotiana within five million years. *The New Phytologist*, vol. 175, no. 4, pp. 756-763. http://dx.doi.org/10.1111/j.1469-8137.2007.02121.x. PMid:17688590.

MARKS, C.E., LADIGES, P.Y. and NEWBIGIN, E., 2011. Karyotypic variation in Nicotiana section Suaveolentes. *Genetic Resources and Crop Evolution*, vol. 58, no. 6, pp. 797-803. http://dx.doi.org/10.1007/s10722-011-9724-3.

MESTIRI, I., CHAGUE, V., TANGUY, A.M., HUNEAU, C., HUTEAU, V., BELCRAM, H., CORITON, O., CHALHOUB, B. and JAHIER, J., 2010. Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. *The New Phytologist*, vol. 186, no. 1, pp. 86-101. http://dx.doi.org/10.1111/j.1469-8137.2010.03186.x. PMid:20149116.

NARAYAN, R.K.J., 1987. Nuclear DNA changes, genome differentiation and evolution in Nicotiana (Solanaceae). *Plant Systematics and Evolution*, vol. 157, no. 3-4, pp. 161-180. http://dx.doi.org/10.1007/BF00936195.

OZKAN, H., TUNA, M. and GALBRAITH, D.W., 2006. No DNA loss in autotetraploids of Arabidopsis thaliana. *Plant Breeding*, vol. 125, no. 3, pp. 288-291. http://dx.doi.org/10.1111/j.1439-0523.2006.01211.x.

PAUN, O., FOREST, F., FAY, M.F. and CHASE, M.W., 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. *The New Phytologist*, vol. 182, no. 2, pp. 507-518. http://dx.doi.org/10.1111/j.1469-8137.2009.02767.x. PMid:19220761.

RENNY-BYFIELD, S., CHESTER, M., KOVARIK, A., LECOMBER, S.C., GRANDBASTIEN, M.A., DELOGER, M., NICHOLS, R.A., MACAS, J., NOVÁK, P., CHASE, M.W. and LEITCH, A.R., 2011. Next generation sequencing reveals genome downsizing in allotetraploid Nicotiana tabacum, predominantly through the elimination of paternally derived repetitive DNAs. *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2843-2854. http://dx.doi.org/10.1093/molbev/msr112. PMid:21512105.

RENNY-BYFIELD, S., KOVARIK, A., KELLY, L.J., MACAS, J., NOVAK, P., CHASE, M.W., NICHOLS, R.A., PANCHOLI, M.R., GRANDBASTIEN, M.A. and LEITCH, A.R., 2013. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *The Plant Journal*, vol. 74, no. 5, pp. 829-839. http://dx.doi.org/10.1111/tpj.12168. PMid:23517128.

RIESEBERG, L.H. and WILLIS, J.H., 2007. Plant speciation. *Science*, vol. 317, no. 5840, pp. 910-914. http://dx.doi.org/10.1126/science.1137729. PMid:17702935.

SERRES-GIARDI, L., BELKHIR, K., DAVID, J. and GLÉMIN, S., 2012. Patterns and evolution of nucleotide landscapes in seed. *The Plant Cell*, vol. 24, no. 4, pp. 1379-1397. http://dx.doi.org/10.1105/tpc.111.093674. PMid:22492812.

SIERRO, N., BATTEY, J.N., OUADI, S., BOVET, L., GOEPFERT, S., BAKAHER, N., PEITSCH, M.C. and IVANOV, N.V., 2013. Reference genomes and transcriptomes of Nicotiana sylvestris and Nicotiana tomentosiformis. *Genome Biology*, vol. 14, no. 6, p. R60. http://dx.doi.org/10.1186/gb-2013-14-6-r60. PMid:23773524.

ŠMARDA, P., BUREŠ, P., ŠMERDA, J. and HOROVÁ, L., 2012. Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. *The New Phytologist*, vol. 193, no. 2, pp. 513-521. http://dx.doi.org/10.1111/j.1469-8137.2011.03942.x. PMid:22050640.

TATARINOVA, T.V., ALEXANDROV, N.N., BOUCK, J.B. and FELDMANN, K.A., 2010. GC3 biology in corn, rice, sorghum and other grasses. *BMC Genomics*, vol. 11, p. 308.

VESELÝ, P., BURES, P., SMARDA, P. and PAVLICEK, P., 2012. Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? *Annals of Botany*, vol. 109, no. 1, pp. 65-75. http://dx.doi.org/10.1093/aob/mcr267. PMid:22021815.

WANG, L. and ROOSSINCK, M.J., 2006. Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. *Plant Molecular Biology*, vol. 61, no. 4-5, pp. 699-710. http://dx.doi.org/10.1007/s11103-006-0041-8. PMid:16897485.

WONG, G.K.S., WANG, J., TAO, L., TAN, J., ZHANG, J.G., PASSEY, D.A. and YU, J., 2002. Compositional gradients in Gramineae genes. *Genome Research*, vol. 12, no. 6, pp. 851-856. PMid:12045139.

XU, X., PAN, S., CHENG, S., ZHANG, B., MU, D., NI, P., ZHANG, G., YANG, S., LI, R., WANG, J., ORJEDA, G., GUZMAN, F., TORRES, M., LOZANO, R., PONCE, O., MARTINEZ, D., CRUZ, G., CHAKRABARTI, S.K., PATIL, V.U., SKRYABIN, K.G., KUZNETSOV, B.B., RAVIN, N.V., KOLGANOVA, T.V., BELETSKY, A.V., MARDANOV, A.V., DI GENOVA, A., BOLSER, D.M., MARTIN, D.M., LI, G., YANG, Y., KUANG, H., HU, Q., XIONG, X., BISHOP, G.J., SAGREDO, B., MEJÍA, N., ZAGORSKI, W., GROMADKA, R., GAWOR, J., SZCZESNY, P., HUANG, S., ZHANG, Z., LIANG, C., HE, J., LI, Y., HE, Y., XU, J., ZHANG, Y., XIE, B., DU, Y., QU, D., BONIERBALE, M., GHISLAIN, M., HERRERA, M.R., GIULIANO, G., PIETRELLA, M., PERROTTA, G., FACELLA, P., O'BRIEN, K., FEINGOLD, S.E., BARREIRO, L.E., MASSA, G.A., DIAMBRA, L., WHITTY, B.R., VAILLANCOURT, B., LIN, H., MASSA, A.N., GEOFFROY, M., LUNDBACK, S., DELLAPENNA, D., BUELL, C.R., SHARMA, S.K., MARSHALL, D.F., WAUGH, R., BRYAN, G.J., DESTEFANIS, M., NAGY, I., MILBOURNE, D., THOMSON, S.J., FIERS, M., JACOBS, J.M., NIELSEN, K.L., SØNDERKÆR, M., IOVENE, M., TORRES, G.A., JIANG, J., VEILLEUX, R.E., BACHEM, C.W., DE BOER, J., BORM, T., KLOOSTERMAN, B., VAN ECK, H., DATEMA, E., HEKKERT, B., GOVERSE, A., VAN HAM, R.C. and VISSER, R.G., 2011. Genome sequence and analysis of the tuber crop potato. *Nature*, vol. 475, no. 7355, pp. 189-195. http://dx.doi.org/10.1038/nature10158. PMid:21743474.

YANG, X., YE, C.Y., CHENG, Z.M., TSCHAPLINSKI, T., WULLSCHLEGER, S., YIN, W., XIA, X. and TUSKAN, G.A., 2011. Genomic aspects of research involving polyploid plants. *Plant Cell, Tissue and Organ Culture*, vol. 104, pp. 387-397.