



An insight into the evolution of introns in the *gyrase A* gene of plants

Mrinalini Manna¹ · Dhirendra Fartyal^{1,2} · V. Mohan M. Achary¹ · Aakrati Agarwal^{1,3} · Malireddy K. Reddy¹

Received: 28 January 2017 / Accepted: 28 January 2018 / Published online: 19 February 2018
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

Abstract

DNA gyrase is a type II topoisomerase essential for replication and transcription in prokaryotes and eukaryotic cell organelles. The functional gyrase enzyme is an A₂B₂ tetramer encoded by the *gyrA* and *gyrB* genes. Most of the eukaryotic *gyrase A* genes possess introns while they are intron-less in prokaryotes. In the present study, we found out the evolutionary passage of intron development in *gyrase A* gene with the help of bioinformatics approaches. All the plant *gyrase A* genes studied by us were found to be a part of the nuclear genome, and their respective proteins were targeted to the organelles. Except the green alga *Bathycoccus prasinus*, these genes contained introns, and the positions of the homologous introns were found to be highly conserved in diverse plant lineages despite having variation in their nucleotide sequence compositions and lengths. However, in red, brown, and green algae: *Chlorella variabilis* and *Chlamydomonas reinhardtii*, homologous intron positions were not conserved, which might be due to the independent acquisition of introns. The study makes it amply evident that the introns appeared in the gene following endosymbiotic gene transfer of the *gyrase A* to the nuclear genome of an ancestral green plant. The land plants appear to have acquired intron-bearing *gyrase A* gene from a common ancestral green alga and subsequently lesser re-arrangement of introns at homologous positions resulted in their positional conservation. However, the introns which are known to be under lesser selection pressure evolved differently in various plant species in terms of base composition and lengths.

Keywords Endosymbiosis · Gyrase A · Introns · Introns early model · Introns late model

Handling Editor: Jim Leebens-Mack.

Mrinalini Manna, Dhirendra Fartyal, and V. Mohan M. Achary have equally contributed to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00606-018-1503-6>) contains supplementary material, which is available to authorized users.

✉ Mrinalini Manna
mrinaliniatbiotech@gmail.com

✉ Dhirendra Fartyal
dhirenfartyal@gmail.com

¹ Crop Improvement Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India

² Uttarakhand Technical University, Dehradun, India

³ Plant Molecular Biology Lab, Department of Botany, University of Delhi, New Delhi, India

Introduction

The replication and transcription processes in prokaryotes are dependent on DNA gyrase enzyme, a type II topoisomerase (type II indicates the ability of DNA gyrase to transiently break both strands of DNA) (Wang 1996; Champoux 2001). The enzyme, an A₂B₂ tetramer encoded by the *gyrA* and *gyrB* genes, catalyzes negative supercoils in DNA during replication and transcription, which is a critical process for survival of the cells (Cozzarelli 1980; Gellert 1981). Eukaryotic nuclear genomes do not require gyrase activity as the negative supercoiling in DNA is achieved by wrapping up of DNA molecules around the histone proteins, and the relaxation of DNA molecules is carried out by gyrase-like type I and type II topoisomerases. On the other hand, organellar genomes, i.e., of chloroplast and mitochondria do not possess histones and essentially require the presence of DNA gyrase enzyme to supercoil and relax their DNA. Previous studies and sequencing data available in databases confirm that DNA gyrase genes are a part of nuclear genome of eukaryotes and the proteins encoded by them are targeted

to respective organelles for carrying out negative supercoiling of chloroplastic and mitochondrial DNA. Following endosymbiotic organelle acquisitions, many genes were transferred from the organelles to nucleus which contributed to organelles becoming dependent on host eukaryotic cells (Martin and Harrmann 1998). Phylogenetic analysis of *AtGyrA* (*gyrase A* gene of *Arabidopsis thaliana*) suggested that the *gyrase A* genes were assimilated through the acquisition of endosymbiotic bacteria by the plant (Wall et al. 2004).

Most of the eukaryotic genes present in nuclear genomes, including *gyrase*, possess introns. These introns are invariably spliceosomal introns, which are the segments of non-coding sequences excised from pre-mRNA by spliceosomal complex (Gilbert 1978; Grabowski et al. 1985; Moore and Sharp 1993; Nilsen 2003; Rogozin et al. 2012). There are two alternate hypotheses explaining the origin of spliceosomal introns, viz. “introns early model” and “introns late model.” The “introns early” model postulates that introns had always been an integral part of the genes, i.e., genes originated as interrupted structures and those presently without introns, must have lost them in the course of evolution (Gilbert and Glynias 1993). On the other hand, the “introns late” model proposes that the ancestral protein-coding genes were uninterrupted and that the introns were subsequently inserted into them (Cech 1986). The rationality of both the hypotheses is supported with various arguments. The “introns early” model has been found advantageous to explain the evolution of novel ancient genes through the process of recombination of different polypeptide units. The early cell is believed to have contained certain different protein-coding genes, and the evolution of many more and novel ones became possible through reshuffling of different polypeptide units to construct the newer ones, and it has been proposed that this reshuffling via recombination of protein-coding gene fragments was made possible by the assistance of introns (Blake 1985; Holland and Blake 1987). In support of this hypothesis, there are evidences from different genes, which were found to contain related exons, suggesting a pattern of gene assembling by the process of exon shuffling (Patthy 1999; Kolkman and Stemmer 2001). It might have also been possible that exon(s) from one gene translocated in the intronic region of another gene, thereby resulting in evolution of a newer protein. Since in many organisms, the introns are often much larger in size compared to the exons, it is likely to suggest the insertion of an exon or a sequence including an exon within an intron and be flanked by functional 5' and 3' splice junctions (Krebs et al. 2013). Introns have been discovered in several chloroplast genes (Plant and Gray 1988; Sheveleva and Hallick 2004), including some that are homologous to bacterial genes, suggesting that endosymbiotic event occurred before introns were lost from the prokaryotic lineage (Krebs et al. 2013). Although these

examples emphasize that introns were part of genes since the onset of evolution, there are certain examples which contradict this view and rather support the “introns late” model. For instance, in case of tRNAs, all of which have same general conformation, it is unlikely that the two regions of the gene evolved independently because the two regions base pair to fold the molecule into the functional shape. In case of the intron-bearing tRNA genes of eukaryotes (Lowe and Chan 2011; Yoshihisa 2014), introns are thought to become inserted into a continuous gene (Krebs et al. 2013). Ahmedi-nejad et al. (2010) compared exon–intron structures of 64 mitochondria-derived genes and found that intron density of recently transferred genes was similar to that of genes originated by most ancient transfers, which led to the conclusion that spliceosomal introns evolved following endosymbiotic gene transfer. Further, it has been discovered that the genes of yeast and mammalian mitochondria encode virtually identical proteins. But, vertebrate mitochondrial genomes are very small and extremely compact (Wolstenholme 1992; Taanman 1999), whereas yeast mitochondrial genomes are much larger and have some complex interrupted genes (Langkjaer et al. 2003). It appears that, since yeast mitochondrial introns (and certain other introns) can be mobile, they might have arisen by insertions thus supporting the “introns late” hypothesis (Krebs et al. 2013).

To understand the origin of introns in *gyrase A* gene (which is thought to be the counterpart of bacterial *gyrase A* subunit) of plants, we, in the present study, analyzed the *gyrase A* gene sequences of different prokaryotes and plant species for intron characteristics and identified that they appeared in the *gyrase A* genes following the endosymbiotic event. We also found that the positions of the homologous introns are highly conserved in diverse green plant lineages.

Materials and methods

Sequence retrieval and determination of intron positions

The coding sequences of *gyrase A* belonging to various organisms (listed in Table 1) were retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/>) or Uniprot (<http://www.uniprot.org/>) database. These CDS sequences were either subjected to BLASTN (<https://blast.ncbi.nlm.nih.gov/>) against genomic DNA database to identify the corresponding genomic DNA sequences or the corresponding genomic DNA sequences were directly retrieved from NCBI/Uniprot database. Each CDS sequence was again subjected to BLASTN to identify the presence of possible paralogs of *gyrase A*. All organisms except *Brassica napus* were found to possess single copy of *gyrase A* genes in their genomes. NCBI database indicated that *Brassica napus*

Table 1 Gyrase A protein characteristics, gene length, and number of introns present in the *gyrase A* genes belonging to different organisms

<i>gyrase A</i> accession No.	Source organism	aa length (bp)	MW (kDa)	pI	Protein localization	Gene length (bp)	No of introns
P0AES4	<i>Escherichia coli</i>	875	96.96	5.08	Cytoplasm	2628	0
P9WG47	<i>Mycobacterium tuberculosis</i>	838	92.27	5.41	Cytoplasm	2517	0
C5C7X9	<i>Micrococcus luteus</i>	898	99.09	5.06	Cytoplasm	2697	0
P35885	<i>Streptomyces coelicolor</i>	857	94.58	5.38	Cytoplasm	2574	0
O67108	<i>Aquifex aeolicus</i>	769	86.99	9.00	Cytoplasm	2235	0
Q5SIL4	<i>Thermus thermophilus</i>	805	89.12	6.14	Cytoplasm	2418	0
P37411	<i>Salmonella typhimurium</i>	878	97.06	4.98	Cytoplasm	2637	0
X60178.1	<i>Haloferax sp.</i> (Archaea)	858	94.46	4.41	Cytoplasm	2577	0
D8LPZ0	<i>Ectocarpus siliculosus</i> (Brown alga)	1031	111.23	8.69	NA	3096	23
M2X5A8	<i>Galdieria sulphuraria</i> (Red alga)	911	103.34	9.19	NA	2736	3
XP_007509930.1	<i>Bathycoccus prasinos</i> (Green alga)	1057	116.43	5.59	Chloroplast	3174	0
XM_005843519.1	<i>Chlorella variabilis</i> (Green alga)	950	102.49	8.55	Chloroplast	4817	17
EDP04229	<i>Chlamydomonas reinhardtii</i> (Green alga)	315	33.73	5.84	NA	6711	12
GAQ84561	<i>Klebsormidium nitens</i> (Green alga)	1004	109.84	6.82	NA	10,309	19
DS545011.1	<i>Physcomitrella patens</i> (Bryophyte)	813 (some aa are missing at N terminus)	–	–	–	9064	24
D8SVR4	<i>Selaginella moellendorffii</i> (Lycopodiophyte)	880	96.53	8.56	NA	3995	26
XM_003589611.2	<i>Medicago truncatula</i>	940	103.87	7.22	Chloroplast	17,441	26
XM_004499691.2	<i>Cicer arietinum</i>	942	104.43	8.55	Chloroplast	17,126	26
AP009637.1	<i>Lotus japonica</i>	945	104.22	7.99	Chloroplast	24,777	26
XM_016091364.1	<i>Arachis duranensis</i>	926	101.98	8.61	NA	14,287	26
XM_008233680.1	<i>Prunus mume</i>	946	104.48	8.17	Chloroplast	16,737	26
XM_016020802.1	<i>Ziziphus jujuba</i>	950	105.39	8.41	Chloroplast	13,326	26
NM_111905.2	<i>Arabidopsis thaliana</i>	950	104.54	6.47	Chloroplast	6632	25
XM_013792701.1	<i>Brassica napus</i>	940	103.18	6.69	Chloroplast	5830	25
XM_012602962.1	<i>Gossypium raimondii</i>	960	105.24	7.27	Chloroplast	19,330	26
XM_011014703.1	<i>Populus euphratica</i>	948	104.44	6.44	Chloroplast	10,671	26
XM_011080671.1	<i>Sesamum indicum</i>	939	103.96	7.08	Chloroplast	17,550	27
XM_010033038.1	<i>Eucalyptus grandis</i>	953	104.81	8.53	Chloroplast	15,756	26
XM_004981246.2	<i>Setaria italica</i>	928	101.75	8.53	NA	11,136	26
XM_008646174.1	<i>Zea mays</i>	933	102.16	8.97	NA	15,447	26
AC137507.3	<i>Oryza sativa</i>	931	102.34	7.56	Mitochondria	13,409	26

aa amino acid, MW molecular weight, pI isoelectric point

genome had three paralogs of *gyrase A* (accession numbers: XM_013880667.1, XM_013792701.1, XM_013829267.1)

belonging to three different chromosomes; A3, A7, and C3, respectively. The CDS and corresponding genomic DNA

sequence of each *gyrase A* were then aligned by using CLUSTALW software available at <http://www.genome.jp/tools/cluster/>, and the gaps resulted in the CDS sequences were identified as introns. The genomic DNA clone of *gyrase A* gene of *Lotus japonica* was retrieved from the NCBI database (AP009637.1), and its corresponding CDS sequence of the gene was manually identified by employing MAC vector software. The CDS and the genomic DNA sequences were further aligned to find out the positions of the introns as described earlier. The gene lengths of each *gyrase A* sequences are depicted in Fig. 1a, relative positions of introns in the CDS of *gyrase A* are shown in Figs. 1b and 3, and the lengths of introns are also shown in Fig. 4.

Multiple sequence alignment and analysis of phylogeny

Multiple sequence alignment of Gyrase A protein sequences was carried out by employing T-COFFEE protein alignment server available at <http://tcoffee.org.cat/apps/tcoffee/do:expresso>, and Phylogeny software available at <http://www.phylogeny.fr/> was used to construct the phylogenetic tree (Fig. 2). The “Structural alignments (Expresso)” option of the T-COFFEE protein alignment server with default parameters was chosen to align the Gyrase A protein sequences. The “One Click” option with default parameters was chosen for construction of phylogenetic tree that employs MUSCLE (Multiple Sequence Comparison by Log-Expectation) for multiple alignment, Gblocks for automatic alignment curation, PhyML for tree building, and TreeDyn for tree drawing (Dereeper et al. 2008).

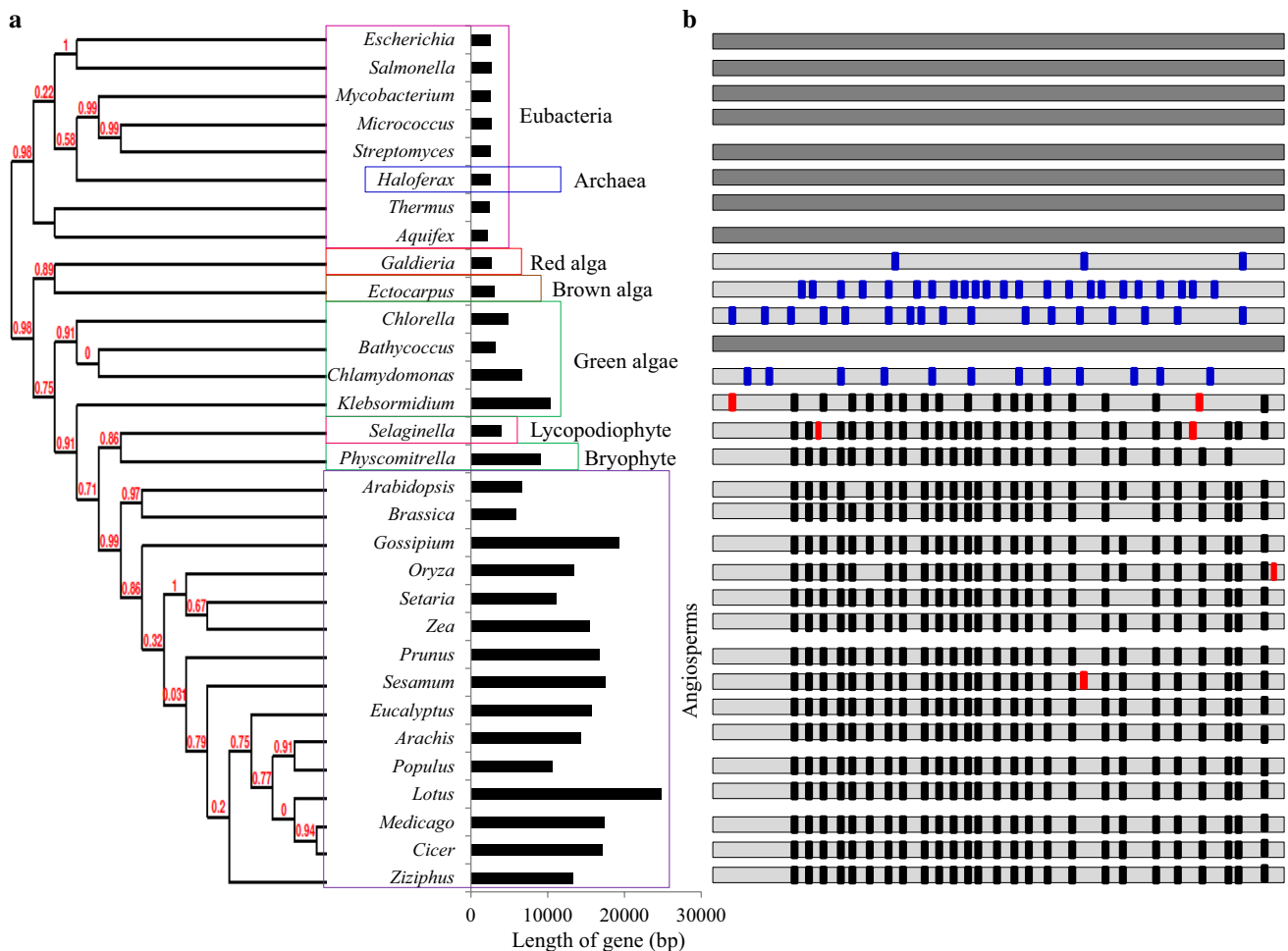


Fig. 1 a The *gyrase A* gene sizes belonging to different organisms. The lengths of intron-containing *gyrase A* genes are more as compared to intron-less *gyrase A* genes belonging to bacteria, archaea, and green alga *Bathycoccus*. The difference in sizes of intron-containing genes is due to varying lengths of introns possessed by them; **b**

schematic representation of distribution of introns in CDSs of *gyrase A* belonging to different organisms. Black bars represent the homologous introns, red bars indicate the extra introns that were missing in rest of the plant species, and blue bars represent nonconserved intron positions

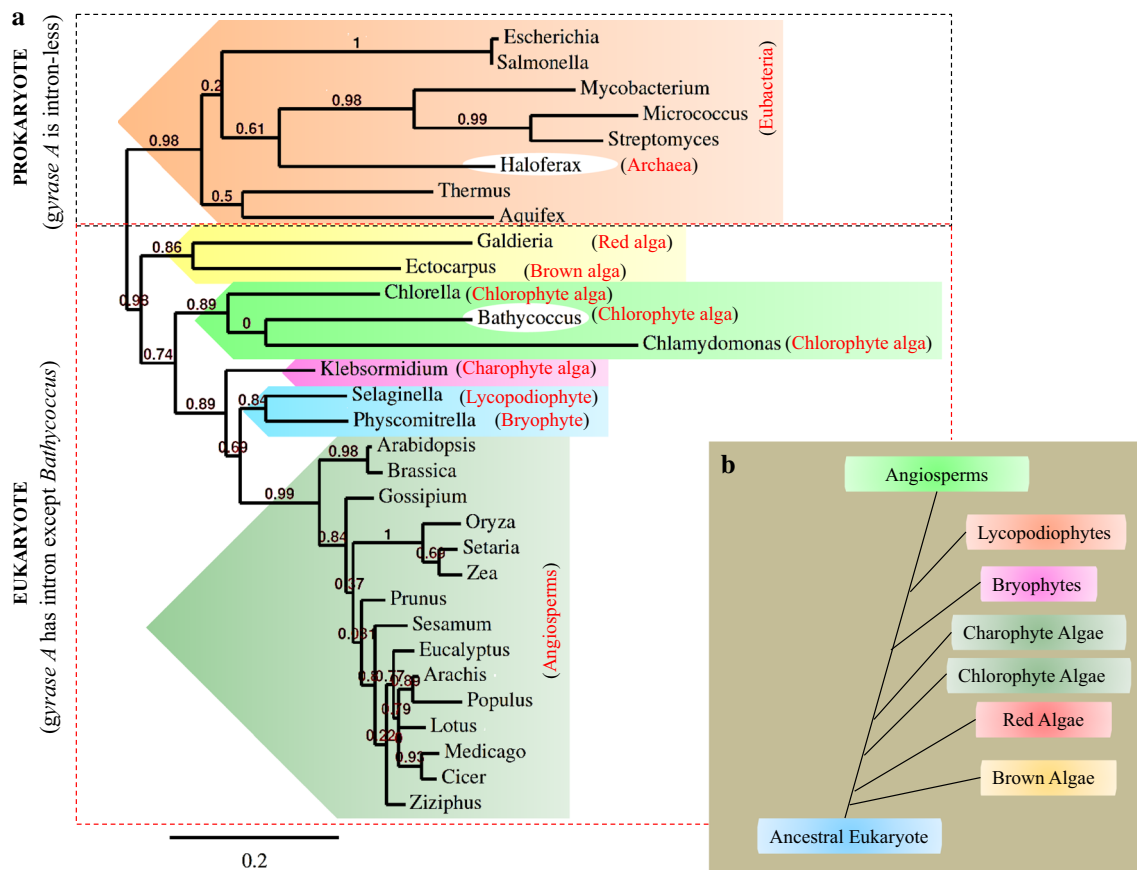


Fig. 2 a Evolutionary relationship between Gyrase A of different organisms. The phylogenetic tree (generated by Phylogeny software) showing grouping of Gyrase A proteins of various species based on

their evolutionary relationship; **b** simplified tree of life showing the algal/plant lineages and their evolutionary relationships (information obtained from Coates et al. 2015)

Identification of the positions of homologous introns

Identification of positions of homologous introns in the CDSs of intron-containing *gyrase A* genes of green alga *Klebsormidium nitens*, bryophyte *Physcomitrella patens*, lycopodiophyte *Selaginella moellendorffii*, and angiosperm species was carried out by aligning the sequences in CLUSTALW software (<http://www.genome.jp/tools/clustalw/>). About 10 bp upstream exon sequence of each intron (exon–intron junction sequences were found by carrying out pairwise alignment of CDS with genomic DNA sequences of respective plant species) was marked in the aligned sequence manually in all the *gyrase A* CDSs of the plant species. Schematic representation of distribution of introns in CDSs of *gyrase A* (Fig. 1b) was made by using same length bars as the aligned CDSs of various organisms, and the introns were represented by rectangular solid boxes. The approximate distances between each pair of introns were maintained by correlating the number of nucleotides with the lengths between them.

Construction of heat map to depict intron length variation

The heat map (Fig. 4) was constructed to depict intron length variation using the “Conditional Formatting” option of the “Home” section of Microsoft Excel (Windows 10). All the cells in Excel sheet containing the intron length values were selected first, and then the “Color Scales” option was chosen under the “Conditional Formatting” option, and further, under the “More Rules” option, the “2-Color Scale” format style option was chosen (one color representing the lowest value and another signifying the highest value) to generate the heat map.

Prediction of protein localization

Subcellular targeting of Gyrase A proteins was predicted by using the Web-based TargetP signal peptide predicting software programs (<http://www.cbs.dtu.dk/services/TargetP/>) (Emanuelsson et al. 2000). The “Plant” option under the “Organism group” section and “specificity > 0.95” option

under the “Cutoffs” section were chosen to predict the sub-cellular location of the Gyrase A proteins.

Protein molecular weight and isoelectric point determination

The molecular weight and isoelectric point (pI) of Gyrase A proteins were determined by using ExPASy Compute pI/Mw tool (http://web.expasy.org/compute_pi/) with default parameters.

Dot plot analysis

The dot matrices (Fig. 5) were prepared by employing multi-zPicture software available at <http://zpicture.dcode.org/> using all parameters by default (Ovcharenko et al. 2004). To construct the dot matrices, for each event, either two CDS sequences of *gyrase A* belonging to two different organisms or genomic DNA sequence of *gyrase A* of two different species was aligned by putting the two sequences under “SEQUENCE 1” and “SEQUENCE 2” columns, respectively.

Results

Sequence analyses of *gyrase A* genes from various organisms

A total of thirty-one different *gyrase A* gene sequences belonging to eubacteria (seven), archaea (one), red alga (one), brown alga (one), green algae (four), bryophyte (one), lycopodiophyte (one), and angiosperm (fifteen) (Table 1) were analyzed in the present study. We could not find any *gyrase A* gene sequences from gymnosperms in the available databases. The *gyrase A* genes belonging to red alga, brown alga, green algae (except *Bathycoccus prasinus*), bryophyte, lycopodiophyte, and angiosperms were found to possess introns while these were missing in the prokaryotes (bacteria and archaea) and the green alga *Bathycoccus prasinus*. Among all the intron-bearing *gyrase A* genes studied, only *Brassica napus* had three paralogs of *gyrase A*, while rest of the *gyrase A* genes were present as single copy in their respective plant genomes. The gene and protein sequences of two paralogs (XM_013792701.1 and XM_013880667.1) of *gyrase A* of *Brassica napus* were identical while the third paralog (XM_013829267.1) had slightly different gene and protein sequences (Online Resource 1). Introns are the characteristics of eukaryotic genes, which result in increased size of the genes. Due to intron accumulation, the size of the genes became bigger (Fig. 1a); however, their size varied

as they possessed introns of varying lengths (Fig. 4). In the present study, the protein sequences of different *gyrase A* gene were found to be highly conserved across these genera of prokaryotes and eukaryotes with high degree of similarity percentage (Online Resource 2). The phylogenetic studies established a close evolutionary relationship between the Gyrase A proteins of various organisms. Six distinct clades were observed among prokaryotic and eukaryotic Gyrase A (Fig. 2). The eubacterial and archaeal Gyrase A proteins grouped together in one clade, and those belonging to red alga (*Galdieria sulphuraria*) formed a clade with brown alga (*Ectocarpus siliculosus*). The Gyrase A from three species of green algae, namely *Chlorella variabilis*, *Bathycoccus prasinus*, and *Chlamydomonas reinhardtii*, formed another separate clade while the one from *Klebsormidium nitens* was the sole member of another clade in close association with Gyrase A of bryophyte, lycopodiophyte, and the angiosperms. The Gyrase A proteins from the bryophyte (*Physcomitrella patens*) formed a clade with lycopodiophyte (*Selaginella moellendorffii*), and those belonging to all the angiosperm species grouped together in separate clade (Fig. 2).

Protein localization of Gyrase A

The localization of Gyrase A was predicted for all the sequences taken under the study, and almost all the proteins were found targeted putatively to either chloroplast or mitochondria. Gyrase A proteins belonging to most of the angiosperms except *Oryza sativa* were found to localize in the chloroplast, and it was found to be localized in mitochondria in *Oryza sativa*. The organeller localization of Gyrase A proteins was not obtained for *Arachis duranensis*, *Setaria italica*, and *Zea mays* by the TargetP signal peptide predicting software at the desired level of specificity (indicated in the Materials and Methods section). Gyrase A belonging to green algae *Bathycoccus prasinus* and *Chlorella variabilis* were localized in chloroplast, and for rest of the green, brown, and red algae, bryophyte and lycopodiophyte, the protein localization targets were not found (Table 1) The localization of Gyrase A belonging to moss (bryophyte) *Physcomitrella patens* was intentionally not predicted since some amino acids from the N terminus were found to be missing from the protein (source: NCBI database).

Locating the position of homologous introns in *gyrase A* genes

The analysis of intron sequences of *gyrase A* genes revealed the presence of splicing junctions with the sequences GT–AG except some variations at a few places. The introns defined in this way start with dinucleotide GT and ends

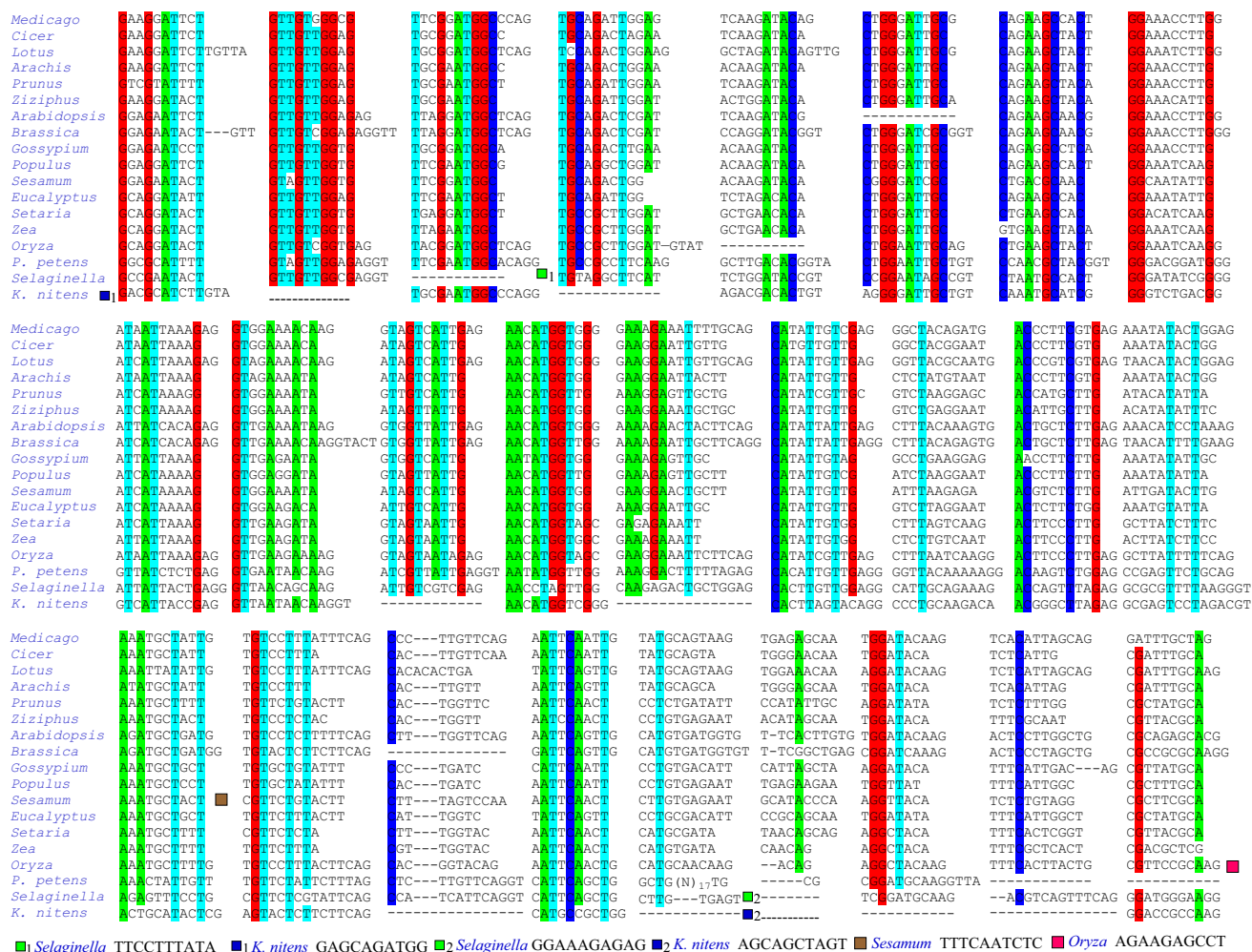


Fig. 3 Location of introns in the *gyrase A* genes. The figure depicts approximately 10 bp upstream sequences (exonic regions) of various homologous introns present in the *gyrase A* genes belonging to different plant species. The sequences represent snapshot view of

aligned coding sequences (CDSs) of the *gyrase A* genes. Extra exons belonging to *Selaginella*, *Klebsormidium*, *Oryza*, and *Sesamum* at the respective positions are marked with colored squares

with dinucleotide AG, following the GU–AG rule, which is a characteristic feature of most of the spliceosomal introns (Rogozin et al. 2012). The number of introns in *gyrase A* genes of various organisms is indicated in Table 1. The introns were invariably located at similar conserved positions in respective *gyrase A* genes of green alga *Klebsormidium nitens*, bryophyte, lycopodiophyte, and angiosperms (Figs. 1b and 3). Barring certain places (Fig. 3), the intron positions were well conserved in the gene sequences. While some homologous introns were found to be missing in *gyrase A* genes of *Klebsormidium nitens* (the green alga), *Physcomitrella patens* (the moss), *Selaginella moellendorffii* (the lycopodiophyte), *Arabidopsis thaliana*, *Brassica napus*, and *Oryza sativa* at various locations (Figs. 1b and 3), extra introns in nonhomologous positions were found

in *Klebsormidium nitens*, *Selaginella moellendorffii*, *Oryza sativa*, and *Sesamum indicum gyrase A* gene sequences (Figs. 1b and 3). Intron positions in three paralogs of *gyrase A* belonging to *Brassica napus* were found to be conserved (Online Resource 1).

Identifying the level of conservation of intron sequences

Although the intron positions were well conserved in *gyrase A* throughout the species taken under study, the introns at homologous locations had different lengths (Fig. 4). To

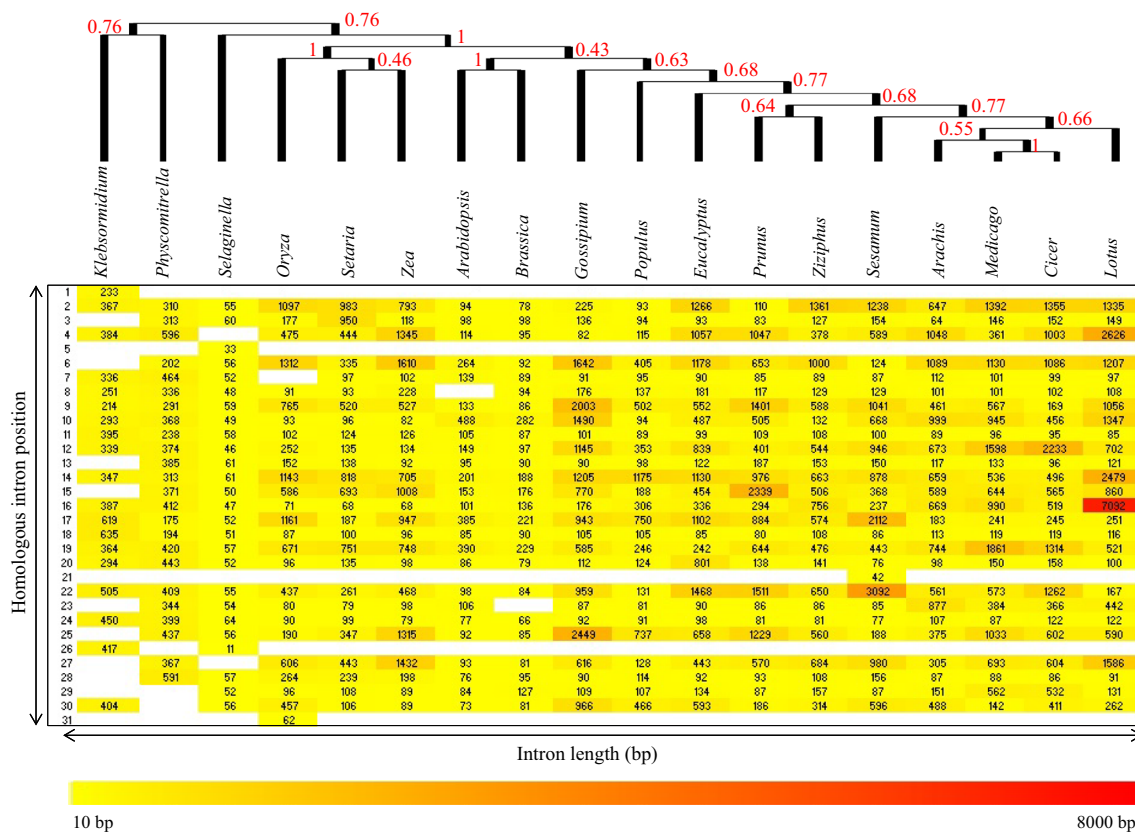


Fig. 4 Heat map and hierarchical cluster displaying the varying lengths of homologous introns belonging to *gyrase A* gene of different plant species. The lengths of homologous introns of *gyrase A* gene are not conserved throughout the evolution for being under lesser selection pressure

know the level of conservation in the sequence composition of these introns, dot matrices were plotted involving two *gyrase A* genomic DNA sequences from different plant species and the results were compared with other dot matrices which involved respective CDS sequences of both the species. The analysis showed a high degree of sequence conservation in CDSs while multiple gaps were present in the aligned dot matrices of their genomic DNA sequences (Fig. 5), wherein these gaps certainly arose due to the non-conservation of intron sequences.

Discussion

The DNA gyrase enzyme performs a crucial function of inducing negative supercoils in the DNA of bacterial and organellar genomes (Drlica 1978). The functional gyrase protein has two subunits: Gyrase A and Gyrase B encoded by *gyrase A* and *gyrase B* genes, respectively. The A and B

subunits together bind to DNA, hydrolyze ATP, and introduce negative supercoils. The A subunit carries out nicking of DNA, B subunit introduces negative supercoils, and finally, the A subunit reseals the nicked strands (Cozzarelli 1980; Gellert 1981). Any abrupt mutation in these genes would impair their functions and might lead to the death of organism due to a nonfunctional chloroplast or mitochondria. Considering above facts, it is extremely important for bacterial or plant species to conserve the sequence of this gene throughout the course of evolution, which made us curious to dissect evolutionary path of *gyrase A* gene. In the present study, a total of thirty-one *gyrase A* gene sequences from distantly related organisms were analyzed to evaluate the conservation of exon and intron sequence compositions and their respective positions. Out of thirty-one *gyrase A* gene sequences, nine have no introns, among which eight belong to prokaryotes and one to eukaryotic green alga *Bathycoccus prasinos*. In rest of the twenty-two *gyrase A* genes, eighteen gene sequences (one from green

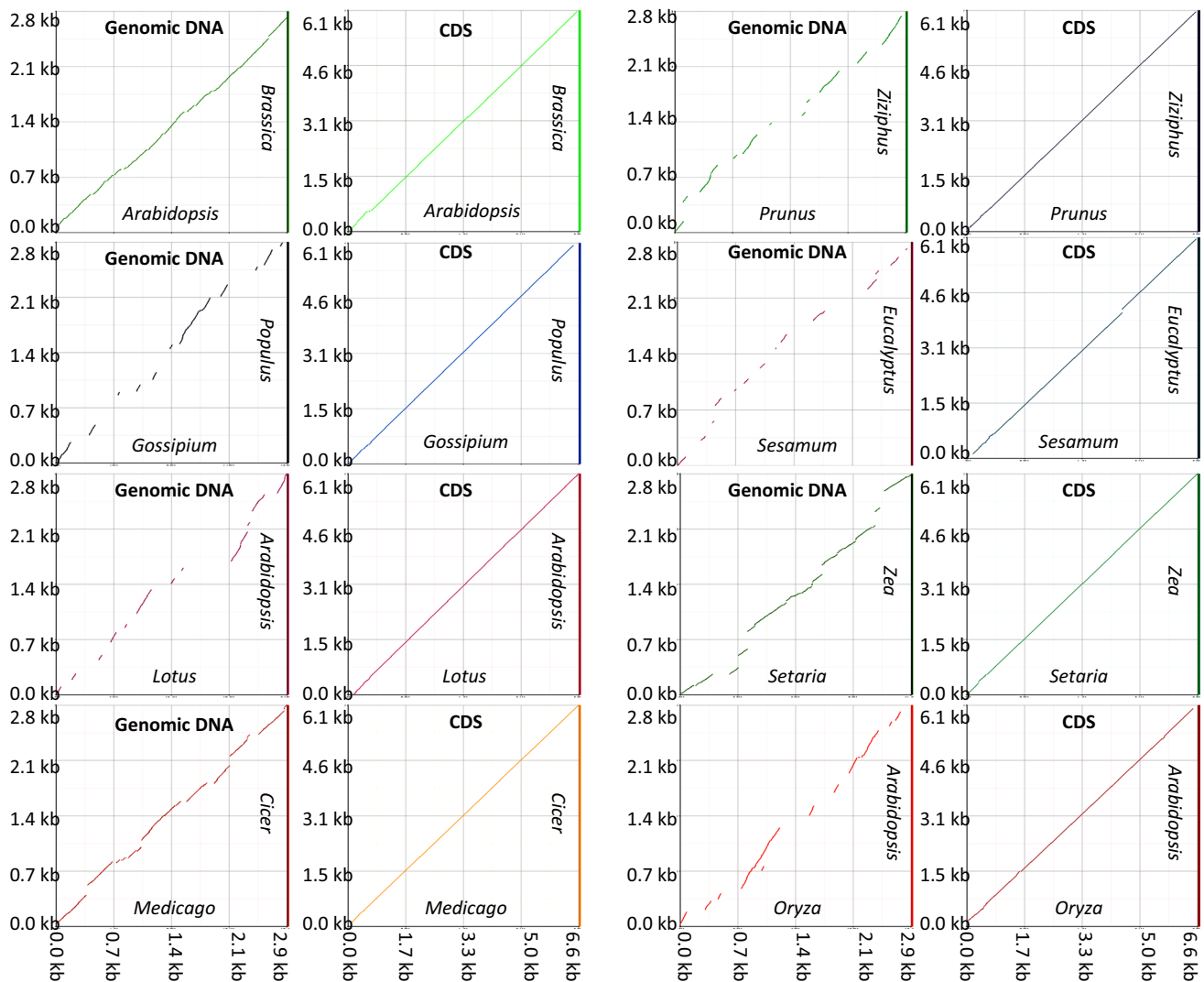


Fig. 5 Similarity level in coding and genomic DNA sequences of *gyrase A* gene. Dot matrices showing the levels of conservation in the paired genomic DNA or coding sequences (CDSs) of *gyrase A* genes. The dot matrices corresponding to genomic DNA sequences contain

gaps indicating nonconservation of intron sequences. On the contrary, CDSs are highly conserved as evident from near absence of gaps in the aligned CDSs

algae (*Klebsormidium nitens*), one from bryophyte (*Physcomitrella patens*), one from lycopodiophyte (*Selaginella moellendorffii*), and fifteen from various angiosperms have introns at the homologous positions while rest of the four *gyrase A* genes (from red, green, and brown algae) possess introns without any positional conservation. Previous studies also identified interruptions at homologous positions relative to the coding sequence of all active globin genes in mammals, birds, and frogs (Hardison 2012). In our study, we found a high level of positional conservation of homologous introns in *gyrase A* genes of distantly related plant species. There was a loss of introns in *gyrase A* genes of *Klebsormidium*, *Physcomitrella*, *Selaginella*, *Oryza*, *Arabidopsis*,

and *Brassica* and a gain of introns in *gyrase A* genes of *Klebsormidium*, *Selaginella*, *Oryza*, and *Sesamum* at various homologous positions (Figs. 1b and 4). The actual function of loss and gain of introns is not clear. In all *gyrase A* gene sequences, exons are well conserved, whereas introns are not only of varying lengths but also not conserved across the various plant species, which supports the view that introns evolve more rapidly than their neighboring exons for not being under higher selection pressure (Morello and Breviario 2008).

Earlier studies have confirmed DNA gyrase activity in the chloroplasts and mitochondria of plants. The inhibitors of bacterial DNA gyrase, novobiocin, and nalidixic acid were

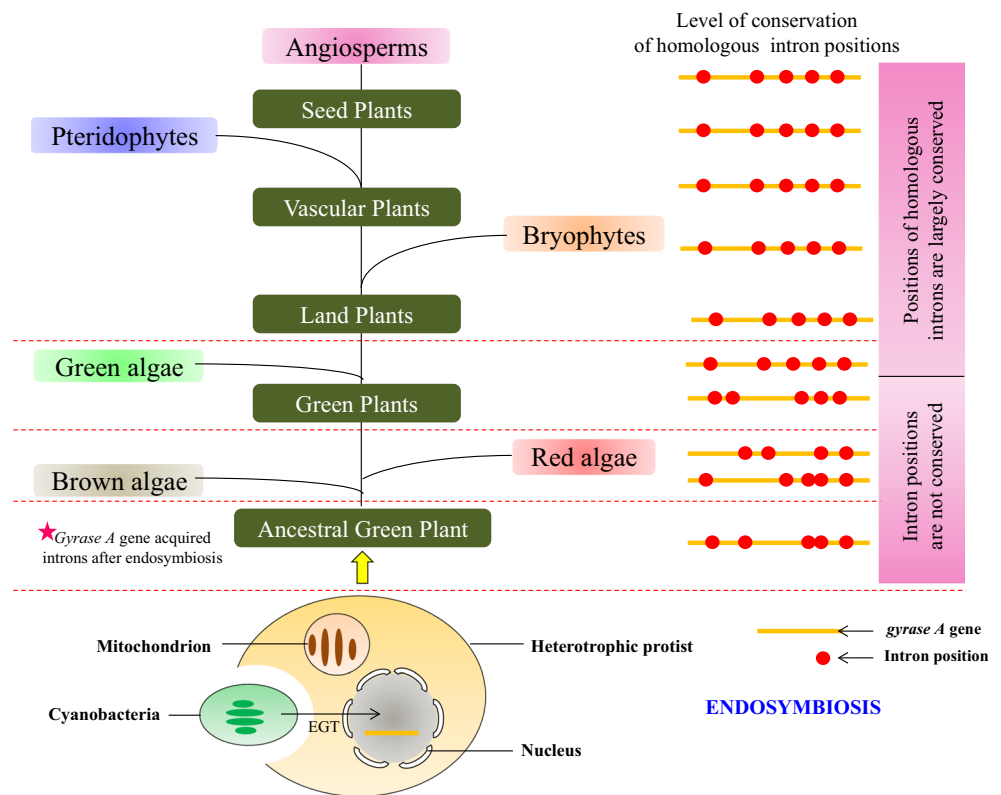
found to inhibit chloroplast DNA synthesis in higher plants (Heinhorst et al. 1985). Also, *Arabidopsis thaliana* genome analysis revealed that *Arabidopsis* DNA gyrase A and B subunits were targeted to organelles (Elo et al. 2003). In our study, all the eukaryotic Gyrase A proteins, except a few, were found to be localized in either chloroplast or mitochondria. It is worthwhile to mention here that the *gyrase A* genes of eukaryotes do not belong to the organellar genomes but are a part of the nuclear genomes, and most of these genes contain introns. Since Gyrase A is a prokaryotic protein, the crucial question arises whether it acquired introns in eukaryotes before the endosymbiotic gene transfer or the introns appeared in the gene after they migrated from organelles to the nucleus following the endosymbiotic event.

There are mainly two models to explain the process of evolution of introns in eukaryotic genes: “introns early” and “introns late” model. The “introns early” model proposes that introns have always been an integral part of the genes and intron-less genes evolved by losing introns during the course of evolution (Gilbert and Glynias 1993). It has been postulated that the introns (flanked by exons) had favored genetic recombination among the exons, leading to evolution of newer proteins in the ancient prokaryotic cell (Blake 1985; Holland and Blake 1987). But the “introns early” model fails to explain why introns are absent in prokaryotic genome. Although the “genome streamlining model” (according to this model, introns were eliminated by prokaryotic cells as the main pressure in the evolution of prokaryotes had been maximization of the replication rate) explains why introns are missing in the prokaryotic genome, it does not have any concrete evidence to support the validity of the hypothesis (Gilbert 1987; Roy 2003). The “introns late” model proposes that the ancestral genes were intron-less and those with introns gained them during the course of evolution (Cech 1986). The hypothesis emphasizes that the introns are a eukaryotic innovation and the intron gain has been a continuous process during the evolution of eukaryotes (Doolittle and Stoltzfus 1993; Mattick 1994; Stoltzfus et al. 1994). As opposed to the “intron early” model, the “introns late” model argues that exon shuffling through recombination is a recent phenomenon important for protein variability in complex eukaryotes only (Patthy 1991). We considered both “introns early” and “introns late” models and found that the “introns late” model suitably explains the evolution of introns in *gyrase A* gene of plants.

In the present study, conservation of position of homologous introns in most of the plant *gyrase A* genes makes the “introns early” model lucrative to explain the origin of introns in *gyrase A* because it might appear that the progenitor ancient *gyrase A* gene contained introns and the gene got passed onto higher order eukaryotic plant species resulting in positional conservation of homologous introns. However, this model fails to explain loss of introns in the green

alga *Bathycoccus prasinus*, where intron loss cannot be explained through “genome streamlining” hypothesis meant for prokaryotes. Also, nonconservation of intron positions in four algal *gyrase A* genes refutes the “introns early” model as there should have been no difference in intron positions in the genes if introns were the integral part of the gene since the time of its genesis. On the other hand, the “introns late” model successfully explained the evolution of introns based on acquired results. We propose that the ancient *gyrase A* gene was intron-less and acquired introns after endosymbiotic gene transfer and subsequently, the introns diverged with variations in both length and sequence compositions in different plant species. This seems a logical proposal as the prokaryotic bacteria and eukaryotic green alga *Bathycoccus prasinus* did not contain any intron in their *gyrase A* genes (*gyrase A* genes of algae are the part of nuclear genome), which indicates the absence of introns in ancestral *gyrase A* gene. The Gyrase A protein sequences belonging to seven different species of bacteria, namely *Escherichia coli*, *Mycobacterium tuberculosis*, *Micrococcus luteus*, *Streptomyces coelicolor*, *Aquifex aeolicus*, *Thermus thermophilus*, and *Salmonella typhimurium*, showed 86% similarity with the Gyrase A protein sequences from diversely related plant species (Online Resource 2) indicating a fairly good level of conservation in exonic regions. Hence, prokaryotic *gyrase A* gene should have been the progenitor of eukaryotic counterparts. Further, since prokaryotic *gyrase A* genes are intron-less, intron gain must have occurred after the endosymbiotic gene transfer. In our study, we found that in red, brown, and two green algae, the positions of homologous introns are not conserved. This might be due to independent acquisition of introns by various algal *gyrase A* genes. Since brown, red, and green algae evolved through separate lineages from a common eukaryotic ancestor (Coates et al. 2015), it might have led to independent acquisition of introns, which further resulted in nonconservation of intron positions in *gyrase A* genes of these algal species. We also believe that the *gyrase A* gene of green alga *Bathycoccus prasinus* did not acquire introns during the course of evolution. Interestingly, one species of green algae (*Klebsormidium nitens*) showed similarity in positions of homologous introns when compared to *gyrase A* genes of higher order plant species. Green algae are known to share significant structural and biochemical features with embryophytes (Shaw et al. 2011). Biochemical, molecular, ultrastructural, and phylogenetic studies support close relationship between charophycean algae and embryophytes (Turmel et al. 2006), and hence, green algae have been identified to be the ancestor of the bryophytes (Shaw et al. 2011). The green alga *Klebsormidium nitens* in the present study belongs to division charophyta, and the positions of homologous introns in its *gyrase A* gene are similar to bryophyte *Physcomitrella nitens*, which in turn depicts possibility of bryophyte to acquire *gyrase A* gene from an

Fig. 6 Evolution of introns in *gyrase A* gene of plants. Schematic diagram representing probable pathway for origin of introns in *gyrase A* genes of the plants. EGT denotes endosymbiotic gene transfer



ancestral species of green algae possibly belonging to division charophyta. Since the positions of the homologous introns remained more or less conserved without further rearrangement throughout the evolutionary timescale starting with the green algae, we hypothesize that the bryophytes acquired intron-bearing *gyrase A* from a common ancestral green algal species and thereafter, during speciation, the *gyrase A* gene containing introns moved to different plant species belonging to lycopodiophytes and angiosperms, and this consequently resulted in conservation of homologous intron positions throughout the course of evolution of green plant species. The conservation of intron positions in all the three paralogs of *gyrase A* gene of *Brassica napus* showed no effect of gene duplication on the positions of the homologous introns. It was not possible to analyze intron positions in *gyrase A* of gymnosperms due to the absence of sequence information in available databases. However, we believe that the homologous introns in those species will have the same conserved intron positions as in algae, bryophytes, lycopodiophytes, and angiosperms. Thus, on the basis of our observations, we propose the “introns late” model for the evolution of introns in *gyrase A* genes in plant species (Fig. 6).

Conclusions

To the best of our knowledge, this is the first report to discover that *gyrase A* genes in plants have well-conserved exons and possess introns at homologous positions with characteristic variations in their nucleotide sequence compositions and lengths. Further, we conclude that these spliceosomal introns evolved following the endosymbiotic gene transfer of the *gyrase A* to the nuclear genome of an ancestral green plant and the higher order green plants, i.e., bryophytes, lycopodiophytes, and angiosperms, acquired intron-bearing *gyrase A* gene from a common ancestral green algal species, which resulted in positional conservation of introns in the *gyrase A* genes. We believe that our approach can be suitably utilized in other genes too for studying intron and exon characteristics and their possible evolutionary pathway. Furthermore, more sequencing data of plant genomes will definitely be an important asset to analyze and consolidate our views regarding the “introns early” or “introns late” model.

Acknowledgements The authors are grateful to the International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, for providing infrastructure facilities. MM, DF, VMMA, and AA convey special thanks to MKR for conceiving the idea and providing necessary guidance. MM and AA were supported by Senior Research Fellowships of Council of Scientific and Industrial Research (CSIR), New Delhi, DF was supported by INSPIRE Fellowship of Department

of Science and Technology, India, and VMMA was supported by fellowship of Department of Biotechnology (DBT), New Delhi.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical statement Authors comply with all rules of the journal following the COPE guidelines; all authors have contributed to and approved the final manuscript.

Information on Electronic Supplementary Material

Online Resource 1. Multiple sequence alignments showing conservation level of CDS and protein sequences belonging to three *gyrase A* paralogs of *Brassica napus* and alignment showing conservation of positions of their homologous introns.

Online Resource 2. Multiple sequence alignment of Gyrase A proteins of different organisms.

References

- Ahmadinejad N, Dagan T, Gruenheit N, Martin W, Gabaldon T (2010) Evolution of spliceosomal introns following endosymbiotic gene transfer. *BMC Evol Biol* 10:57. <https://doi.org/10.1186/1471-2148-10-57>
- Blake CC (1985) Exons and the evolution of proteins. *Int Rev Cytol* 93:149–185
- Cech TR (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44:207–210. [https://doi.org/10.1016/0092-8674\(86\)90751-8](https://doi.org/10.1016/0092-8674(86)90751-8)
- Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. *Annual Rev Biochem* 70:369–413. <https://doi.org/10.1146/annurev.biochem.70.1.369>
- Coates JC, E-Aiman U, Charrier B (2015) Understanding “green” multicellularity: do seaweeds hold the key? *Frontiers Pl Sci* 5:737. <https://doi.org/10.3389/fpls.2014.00737>
- Cozzarelli NR (1980) DNA gyrase and the supercoiling of DNA. *Science* 207:953–960. <https://doi.org/10.1126/science.6243420>
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucl Acids Res* 36 (Web Server issue): W465–W469. <https://doi.org/10.1093/nar/gkn180>
- Doolittle WF, Stoltzfus A (1993) Molecular evolution. Genes-in-pieces revisited. *Nature* 361:403. <https://doi.org/10.1038/361403a0>
- Drlica K, Snyder M (1978) Superhelical *Escherichia coli* DNA: relaxation by coumermycin. *J Molec Biol* 120:145–154. [https://doi.org/10.1016/0022-2836\(78\)90061-X](https://doi.org/10.1016/0022-2836(78)90061-X)
- Elo A, Lyznik A, Gonzalez DO, Kachman SD, Mackenzie SA (2003) Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the Arabidopsis genome. *PL Cell* 15:1619–1631. <https://doi.org/10.1105/tpc.010009>
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Molec Biol* 300:1005–1016. <https://doi.org/10.1006/jmbi.2000.3903>
- Gellert M (1981) DNA topoisomerases. *Annual Rev Biochem* 50:879–910. <https://doi.org/10.1146/annurev.bi.50.070181.004311>
- Gilbert W (1978) Why genes in pieces? *Nature* 271:501. <https://doi.org/10.1038/271501a0>
- Gilbert W (1987) The exon theory of genes. *Cold Spring Harb Symp Quant Biol* 52:901–905
- Gilbert W, Glynias M (1993) On the ancient nature of introns. *Gene* 135:137–144
- Grabowski PJ, Seiler SR, Shrap PA (1985) A multicomponent complex is involved in splicing of messenger RNA precursors. *Cell* 42:345–353. [https://doi.org/10.1016/S0092-8674\(85\)80130-6](https://doi.org/10.1016/S0092-8674(85)80130-6)
- Hardison RC (2012) Evolution of haemoglobin and its genes. *Cold Spring Harb Prospect Med* 2:a011627. <https://doi.org/10.1101/cshperspect.a011627>
- Heinhorst S, Cannon GC, Weissbach A (1985) Chloroplast DNA synthesis during the cell cycle in cultured cells of *Nicotianatabacum*: inhibition by nalidixic acid and hydroxyurea. *Arch Biochem Biophys* 239:475–479
- Holland SK, Blake CC (1987) Proteins, exons and molecular evolution. *Biosystems* 20:181–206. [https://doi.org/10.1016/0303-2647\(87\)90044-X](https://doi.org/10.1016/0303-2647(87)90044-X)
- Kolkman JA, Stemmer WP (2001) Directed evolution of proteins by exon shuffling. *Nat Biotechnol* 19:423–428
- Krebs JE, Goldstein ES, Kilpatrick ST (2013) Lewin’s genes XI. Jones & Bartlett PUBL, Sudbury
- Langkjaer RB, Casaregola S, Ussery QW, Gaillardin C, Piskur J (2003) Sequence analysis of three mitochondrial DNA molecules reveals interesting differences among *Saccharomyces* yeasts. *Nucl Acids Res* 31:3081–3091
- Lowe T, Chan P (2011) Genomic tRNA database. Available at: <http://lowelab.ucsc.edu/tRNAscan-SE/>
- Martin W, Harrmann RG (1998) Gene transfer from Organelle to the Nucleus: how Much, What Happens, and Why? *PL Physiol* 118:9–17. <https://doi.org/10.1104/pp.118.1.9>
- Mattick JS (1994) Introns: evolution and function. *Curr Opin Genet Developm* 4:823–831
- Moore MJ, Sharp PA (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* 365:364–368. <https://doi.org/10.1038/365364a0>
- Morello L, Breviaro D (2008) Plant spliceosomal introns: not only cut and paste. *Curr Genomics* 9:227–238. <https://doi.org/10.2174/138920208784533629>
- Nilsen TW (2003) The spliceosome: the most complex macromolecular machine in the cell? *BioEssays* 25:1147–1149. <https://doi.org/10.1002/bies.10394>
- Ovcharenko I, Loots GG, Hardison RC, Miller W, Stubbs L (2004) zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res* 14: 472–477. <https://doi.org/10.1101/gr.2129504>
- Pathy L (1991) Exons—original building blocks of proteins? *BioEssays* 13:187–192
- Pathy L (1999) Genome evolution and the evolution of exon-shuffling—a review. *Gene* 238:103–144
- Plant AL, Gray JC (1988) Introns in chloroplast protein-coding genes of land plants. *Photosynthesis Res* 16:23–39. <https://doi.org/10.1007/BF00039484>
- Rogozin IB, Carmel L, Csuros M, Koonin EV (2012) Origin and evolution of spliceosomal introns. *Biol Direct* 7:11. <https://doi.org/10.1186/1745-6150-7-11>
- Roy SW (2003) Recent evidence for the exon theory of genes. *Genetica* 118:251–266
- Shaw AJ, Szövényi P, Shaw B (2011) Bryophyte diversity and evolution: windows into the early evolution of land plants. *Amer J Bot* 98:352–369. <https://doi.org/10.3732/ajb.1000316>
- Sheveleva EV, Hallick RB (2004) Recent horizontal intron transfer to a chloroplast genome. *Nucl Acids Res* 32:803–810. <https://doi.org/10.1093/nar/gkh225>
- Stoltzfus A, Spencer DF, Zuker M, Logsdon JMJ, Doolittle WF (1994) Testing the exon theory of genes: the evidence from protein structure. *Science* 265:202–207. <http://www.jstor.org/stable/2884169>

- Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1401:103–123. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3)
- Turmel M, Otis C, Lemieux C (2006) The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Molec Biol Evol* 23:1324–1338. <https://doi.org/10.1093/molbev/msk018>
- Wall MK, Mitchenall LA, Maxwell A (2004) *Arabidopsis thaliana* DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci USA* 101:7821–7826. <https://doi.org/10.1073/pnas.0400836101>
- Wang JC (1996) DNA topoisomerases. *Annual Rev Biochem* 65:635–692. <https://doi.org/10.1146/annurev.bi.65.070196.003223>
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 141:173–216. [https://doi.org/10.1016/S0074-7696\(08\)62066-5](https://doi.org/10.1016/S0074-7696(08)62066-5)
- Yoshihisa T (2014) Handling tRNA introns, archael way and eukaryotic way. *Frontiers Genet* 5:213. <https://doi.org/10.3389/fgene.2014.00213>