Tree Genetics & Genomes (2012) 8:841–854
DOI 10.1007/s11295-012-0469-8

ORIGINAL PAPER

Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis

Shingo Terakami · Yuichiro Matsumura · Kanako Kurita · Hiroyuki Kanamori · Yuichi Katayose · Toshiya Yamamoto · Hironori Katayama

Received: 15 June 2011 / Revised: 14 November 2011 / Accepted: 27 December 2011 / Published online: 8 February 2012
© Springer-Verlag 2012

Abstract The chloroplast genome of *Pyrus* was found to be 159,922 bp in length which included a pair of inverted repeats (IRs) of 26,392 bp, separated by a small single-copy region of 19,237 bp and a large single-copy region (LSC) of 87,901 bp. A total of 130 predicted genes (113 unique genes and 17 genes, which were duplicated in the IR) including 79 protein-coding genes, four ribosomal RNA genes and 30 tRNA genes were identified based on similarity to homologs from the chloroplast genome of *Nicotiana tabacum*. Genome organization was very similar to the inferred ancestral angiosperm chloroplast genome. Comparisons between *Pyrus*, *Malus*, and *Prunus* in Rosaceae revealed 220 indels (≥10 bp). Excluding *ycf1* and *ycf2*, which contained deletions in the coding region, all of these were detected in the spacer or intron regions. Three insertions and 13 deletions were detected in *Pyrus* compared to the same loci in *Malus* and *Prunus*. After comparing 89 noncoding chloroplast DNA regions in *Pyrus* and *Malus*, highly variable regions such as *ndhC-trnV* and *trnR-atpA* were identified. In *Pyrus* and *Malus*, the IR/LSC borders were 62 bp shorter than those of *Prunus*. In addition, there were length mutations at the IRa/LSC junction and in *trnH*. A total of 67 simple sequence repeats (more than 10 repeated motifs) were identified in the *Pyrus* chloroplast genome. The indels and simple sequence repeats will be useful evolutionary tools at both intra- and interspecific levels. Phylogenetic analysis demonstrated a close relationship between *Pyrus* and *Prunus* in the Rosaceae.

Keywords *Pyrus* · Pear · Chloroplast DNA · Complete sequence · Pyrosequencing · Rosaceae

Communicated by A. Abbott

S. Terakami and Y. Matsumura contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11295-012-0469-8) contains supplementary material, which is available to authorized users.

S. Terakami · T. Yamamoto
National Institute of Fruit Tree Science,
National Agricultural and Food Research Organization Tsukuba,
Ibaraki 305-8605, Japan

Y. Matsumura · H. Katayama (✉)
Food Resources Education and Research Center,
Kobe University,
Hyogo 675-2103, Japan
e-mail: hkata@kobe-u.ac.jp

K. Kurita · H. Kanamori
Institute of Society for Techno-innovation of Agriculture,
Forestry and Fisheries,
446-1 Kamiyokoba, Tsukuba,
Ibaraki 305-0854, Japan

Y. Katayose
National Institute of Agrobiological Sciences,
2-1-2 Kannondai, Tsukuba,
Ibaraki 305-8602, Japan

Introduction

Chloroplasts, plant cell organelles derived from independent living cyanobacteria (Keeling 2004) contain the entire enzymatic machinery for photosynthesis. In addition, several other biochemical pathways are present including those responsible for the biosynthesis of fatty acids, amino acids, pigments, and vitamins. Chloroplasts contain their own small genome and this generally has a highly conserved organization. In angiosperms, the genome is circular with a quadripartite structure that includes two copies of an inverted repeat (IR) that are usually 20–26 kb in length and that separate one large single copy (LSC) region and one small single-copy (SSC) region (Palmer et al. 1985).

Chloroplast genomes are 108–217 kb, with the vast majority in the 150–170 kb range. Most contain 110–130 distinct genes; the majority of these genes (about 80) code for proteins that are generally involved in photosynthesis or gene expression, with the remainder being transfer RNA (about 30) or ribosomal RNA (4) genes (Raubeson and Jansen 2005).

There has been a rapid increase in our understanding of chloroplast genome organization and evolution, owing to the availability of many new completely sequenced genomes.

Table 1 Genes locating on the *Pyrus* chloroplast genome

Category	Gene names
Ribosomal RNAs	*rrn16*, *rrn23*, *rrn4.5*, *rrn5*
Transfer RNAs	*trnA-UGC*, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC*, *trnG-UCC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnK-UUU*, *trnL-CAA*, *trnL-UAA*, *trnL-UAG*, *trnM-CAU*, *trnfM-CAU*, *trnN-GUU*, *trnP-UGG*, *trnQ-UUG*, *trnR-UCU*, *trnR-ACG*, *trnS-UGA*, *trnS-GUG*, *trnS-GGA*, *trnT-GGU*, *trnT*
Proteins of small ribosomal subunit	*rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, *rps16*, *rps18*, *rps19*
Proteins of large ribosomal subunit	*rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*
Subunits of RNA polymerase	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Subunits of NADH-dehydrogenase	*ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Subunits of Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
Subunits of Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*,
Large subunit of Rubisco	*rbcL*
Subunits of cytochrome b/f complex	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*
Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, atpH, atpI
Acetyl-CoA carboxylase	*accD*
Cytochrome c biogenesis	*ccsA*
Maturase	*matK*
Protease	*clpP*
Envelope membrane protein	*cemA*
Conserved hypothetical chloroplast reading frames	*ycf1*, *ycf2*, *ycf3*, *ycf4*

Since the first report on the complete chloroplast genome of Liverwort (Ohyama et al. 1986), more than 150 complete chloroplast genomes from plants and algae have been deposited in GenBank so far. Comparative studies indicate that the chloroplast genomes of land plants are highly conserved in both gene order and gene content. In several lineages of land plants, chloroplast DNAs have multiple rearrangements, including *Pinus* (Wakasugi et al. 1994) and the angiosperm families Campanulaceae (Cosner et al. 1997), Fabaceae (Milligan et al. 1989; Palmer et al. 1988), Geraniaceae (Palmer 1987), and Lobeliaceae (Knox and Palmer 1998).

The pace of chloroplast genome sequencing has increased markedly over the last 5 years (Jansen et al. 2005) driven largely by improvements in Sanger sequencing technology that have greatly reduced time and cost (Metzker 2005). In addition, new sequencing technologies by pyrosequencing platforms such as the Roche 454 system (Branford, CT, USA; available through Roche Diagnostics, Indianapolis, IN, USA) have been proposed in recent years that will further significantly reduce the time for and cost of obtaining whole chloroplast genome sequences (Huse et al. 2007; Moore et al. 2006; Tangphatsornruang et al. 2010; Yang et al. 2010). In a single run, the 454 system (Roche 454 GS FLX Titanium) generates up to 400 million high-quality bases in hundreds of thousands of short sequence reads called flowgrams, which are then assembled into genomic contigs. For relatively small genomes, the high number of reads results in a high average depth of sequence coverage, effectively overcoming many of the limitations of pyrosequencing, such as relatively short read length, and uncertainty in the length of homopolymer runs (Margulies et al. 2005; Ronaghi et al. 1998). Perhaps the greatest advantage of the 454 system is that it generates genome sequence much more rapidly and economically than traditional Sanger-based shotgun sequencing.

There are many advantages to using chloroplast DNA for taxonomy and evolutionary research: (1) it is of small size, has high copy number, and has a simple structure; (2) gene content and arrangement are more conserved than in mitochondrial and nuclear genomes making it easier to design primers; (3) it is maternally inherited and thus without the genetic re-assortment that interferes with molecular phylogenetic relationships. In addition, information about the chloroplast genome can be used for various research, chloroplast transformation (Maliga 2002), the development of crops with good agricultural traits (Bock and Khan 2004; Daniell et al. 2004).

Pear (*Pyrus*) is an important economic crop and a member of the Rosaceae family, that has been cultivated for more than 2,000 years and is among the most important fruits in all the temperate regions in about 50 countries of the world (Bell 1990). The classification of *Pyrus* is often very difficult due to natural or artificial interspecific hybrids, which arise easily because of self-incompatibility and the lack of distinguishable characters between species. To establish the

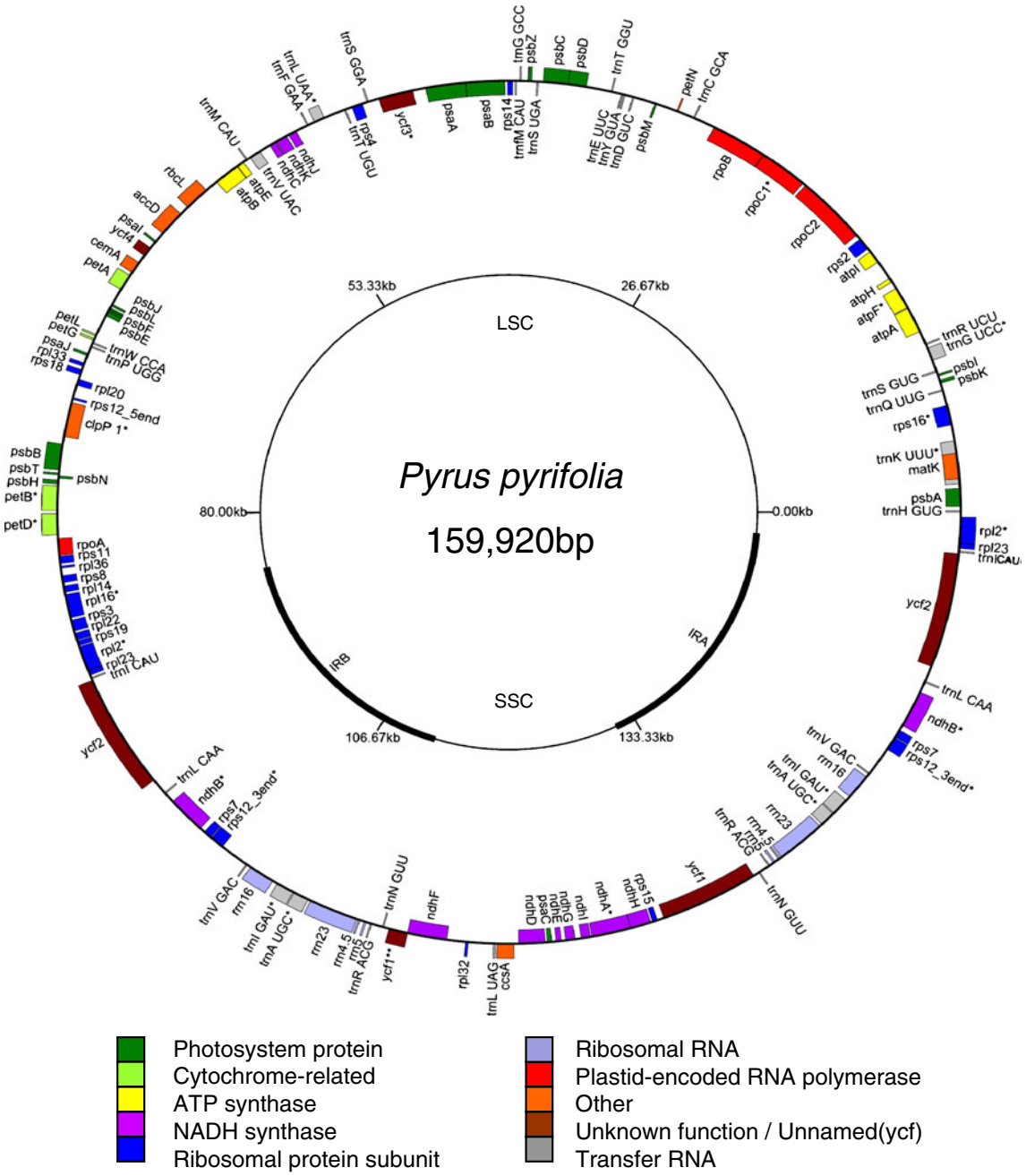

Fig. 1 Gene map of the *P. pyrifolia* chloroplast genome. The thick line indicates the extent of the IRs (IRa and IRb) which separate the genome into SSC and LSC regions. Genes on the outside of the map are transcribed in the clockwise direction and genes on the inside of the map are transcribed in the counterclockwise direction. Genes containing introns and pseudogenes are marked with * and **, respectively

phylogenetic classification of *Pyrus*, various comparisons have been tried using morphological characters, phenolic compounds, and DNA markers (Aldasoro et al. 1996; Bell 1990; Challice and Westwood 1973; Rehder 1940; Rubtsov 1944; Volk et al. 2006; Yamamoto et al. 2002). Until now, few chloroplast-derived markers have been used to study evolutionary relationships among *Pyrus* species (Iketani et al. 1998; Katayama and Uematsu. 2003; Kimura et al. 2003; Katayama et al. 2012). Additional markers are required to determine more detailed relationships among *Pyrus*, and for improved phylogenetic classification within the Rosaceae family in general which is currently controversial. We hope to add to the available information by assessing chloroplast genome variation using structural alterations such as indels and microsatellites. Two complete Rosaceae chloroplast genomes *Malus* × *domestica* (GDR/Genome Database for

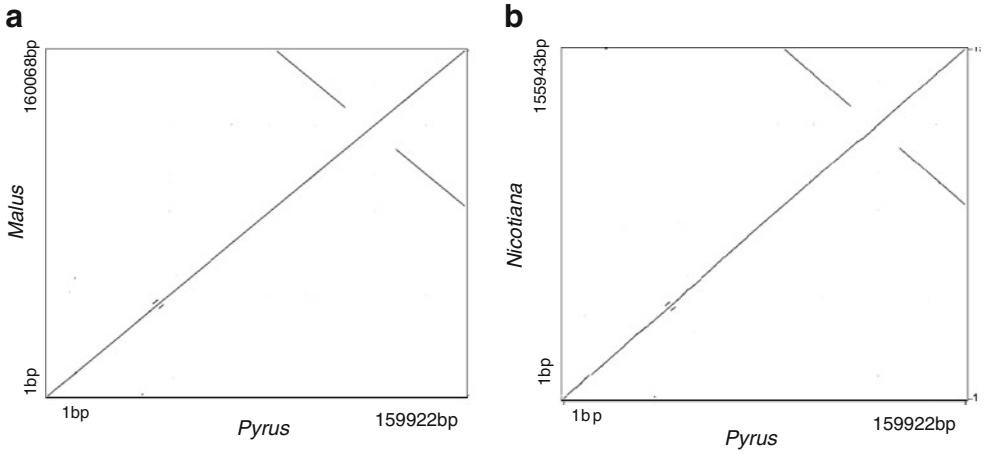

Fig. 2 Dot plot. **A** *P. pyrifolia* versus *Malus* x *domestica*. Numbers along the *x*-axis indicate the coordinates for *Pyrus* and along the *y*-axis for *Malus*. **B** *P. pyrifolia* versus *N. tabacum*

Rosaceae, http://www.rosaceae.org/projects/apple_genome) and *Prunus persica* (Jansen et al. 2011) were reported last year. However, detailed comparisons between these chloroplast genomes have not yet been achieved. Comparison of the three chloroplast genomes should reveal genome structures such as highly variable regions in the Rosaceae.

In this study, we present the complete sequence of the chloroplast genomes of pear (*Pyrus pyrifolia*) and carry out some comparative analyses with other known chloroplast genomes, especially in Rosaceae. We use one of the next-generation sequencing method—pyrosequencing (Roche 454 GS FLX Titanium) for complete chloroplast genome sequencing.

Materials and methods

Total DNA was isolated from fresh green leaves of a Japanese pear variety "Housui" (*P. pyrifolia*) using Genomic DNA Buffer set and QIAGEN Genomic-tip 20/G (QIAGEN). The DNA was sheared by nebulization, amplified by emulsion polymerase chain reaction (PCR), and sequencing was performed by the pyrosequencing method using the Genome Sequencer Roche 454 (GS)-FLX Titanium (Roche Diagnostics; Margulies et al. 2005). The raw sequences were trimmed with quality score less than 20. Adapter sequences were also trimmed, and cleaned sequences less than 150 bases in length were discarded using CLC genomics Workbench ver. 3.7.1 (CLC bio, Aarhus, Denmark). Mapping assembly was generated with the reference sequence, the complete chloroplast genome of *Nicotiana tabacum* (Z00044), using the CLC genomic workbench applying default parameters.

Misread and unread sequences were amplified by PCR and completed by Sanger sequencing. The primers used for sequencing are listed (Electronic supplementary material (ESM) Table 1). PCR was carried out by PrimeSTAR DNA polymerase (TAKARA BIO). Sequencing reactions were performed by Big Dye Terminator v3.1 (Applied Biosystems) and applied to ABI3100 DNA sequencer (Applied Biosystems).

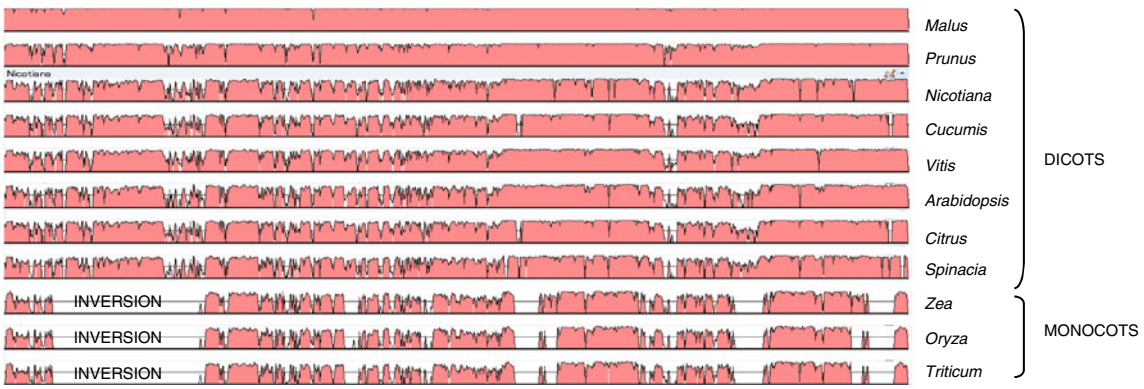

Fig. 3 Identities of several dicot and monocot species. VISTA plot (global alignment) comparison of pear chloroplast genome with 11 chloroplast genomes. Y-scale represents the percent identity ranging from 50% to 100%. Genomes are arranged according to the number of conserved bases with respect to *Pyrus*

Table 2 Loci of indels (≥10 bp) found in chloroplast genomes among three genera in Rosaceae

Locus	Sequence length (bp)			Identity(%) with *Pyrus*		Number of insertion and length with *Pyrus*		Number of deletion and length with *Pyrus*	
	Pyrus	*Malus*	*Prunus*	*Malus*	*Prunus*	*Malus*	*Prunus*	*Malus*	*Prunus*
psbA-trnK	264	236	258	82	87	0	0	1 (28)	0
trnK intron	2,409	2,514	2,538	99	93	1 (25)	0	0	0
trnK-rps16	1,697	1,716	1,710	95	85	1 (21)	3 (13,21,17)	0	4 (21, 33, 22, 21)
rps16 intron	905	862	871	95	87	0	0	2 (17a, 23b)	2 (17a, 23b)
rps16-trnQ	882	804	667	88	59	0	0	3 (28c, 30d, 23e)	4 (143, 28c, 37d, 22e)
psbK-psbI	256	257	230	99	87	0	0	0	1 (10)
trnS-trnG	650	658	403	99	49	0	0	0	1 (258)
trnG- intron	699	696	723	99	91	0	1 (20)	0	0
trnG-trnR	211	211	138	99	50	0	0	0	2 (50, 34)
trnR-atpA	784	842	456	85	44	3 (17, 20, 40)	0	3 (15, 11, 12f)	9 (24, 61, 39, 12f, 54, 12, 24, 87, 10)
atpF-atpH	450	449	436	99	90	0	0	0	1 (14)
atpH-atpI	1,173	1,175	1,210	97	73	0	2 (24, 15)	0	2 (22, 10)
rps2-rpoC2	298	294	262	97	79	0	0	0	1 (40)
rpoC1 intron	735	741	764	98	85	0	1 (21)	0	0
rpoB-trnC	1,215	1,210	1,016	98	75	0	2 (20, 11)	0	2 (47, 139)
trnC-petN	945	943	898	97	85	0	0	0	4 (14, 17, 11, 13)
petN-psbM	1,193	1,252	1,211	91	79	1 (62g)	2 (11, 71g)	0	4 (10, 24, 11, 13)
psbM-trnD	1,231	1,220	1,185	98	89	0	0	1 (14h)	2 (15h, 11)
trnD-trnY	443	444	439	99	85	0	1 (14)	0	1 (12)
trnE-trnT	523	524	453	99	80	0	0	0	3 (14, 28, 12)
trnT-psbD	1,433	1,486	1,219	95	76	3 (20, 16, 14)	1 (15)	0	3 (98, 65, 53)
psbC-trnS	255	249	264	98	89	0	1 (13)	0	0
trnS-psbZ	309	318	342	96	75	0	1 (50)	0	1 (24)
psbZ-trnG	408	518	273	81	67	5 (12, 54, 11, 12i,15)	1 (12i)	0	2 (112, 38)
psaA-ycf3	751	696	704	93	82	0	2 (15, 10)	1 (54j)	2 (57j, 11)
ycf3-trnS	610	608	622	99	83	0	2 (22, 17)	0	1 (16)
trnT-trnL	1,219	1,195	926	92	65	1 (20)	2 (11, 11)	3 (34k, 18l)	8 (34k, 11, 48, 29, 99, 36, 21, 18l)
trnL-trnF	395	403	384	98	82	0	0	0	1 (11)
trnF-ndhJ	698	705	733	98	85	0	2 (20, 17)	0	0
ndhJ-ndhK	137	137	127	99	89	0	0	0	1 (10)
ndhC-trnV	668	632	636	90	70	1 (12)	4 (27, 20, 10, 11)	2 (20, 16m)	4 (21, 12, 64, 16m)
trnV-trnM	167	185	206	91	82	1 (18)	1 (32)	0	0
trnM-atpE	233	233	247	99	64	0	2 (10, 45)	0	2 (20, 20)
rbcL-accD	568	578	613	98	81	0	1 (34)	0	2 (10, 20)

Table 2 (continued)

Locus	Sequence length (bp)			Identity(%) with *Pyrus*		Number of insertion and length with *Pyrus*		Number of deletion and length with *Pyrus*	
	Pyrus	*Malus*	*Prunus*	*Malus*	*Prunus*	*Malus*	*Prunus*	*Malus*	*Prunus*
accD-psaI	584	819	755	70	65	1 (229n)	1 (224n)	0	2 (19, 19)
ycf4-cemA	526	534	557	98	85	0	1 (15)	0	0
petA-psbJ	991	995	945	98	87	0	1 (13)	0	2 (19, 29)
psbE-petL	1,244	1,261	1,300	98	88	1 (18)	3 (11, 14, 12)	0	0
petL-petG	158	158	175	100	84	0	1 (12)	0	0
trnW-trnP	156	156	146	100	71	0	1 (16)	0	0
trnP-psaJ	407	407	385	99	83	0	0	0	1 (26)
rpl33-rps18	218	237	172	91	69	1 (17)	0	0	2 (16, 26)
rps18-rpl20	280	276	137	99	57	0	0	0	2 (129, 16)
rps12_5end-clP1	188	187	188	98	86	0	1 (10)	0	0
clpP intron	1,466	1,471	1,438	98	86	1 (12)	1 (11)	0	2 (27, 24)
petB intron	798	797	755	99	87	0	1 (20)	0	1 (30)
petD intron	798	797	755	99	95	0	1 (11)	0	0
petD-rpoA	223	223	236	99	85	0	0	0	2 (10, 12)
rps11-rpl36	168	176	147	97	82	0	0	0	1 (22)
rpl36-rps8	459	460	463	99	86	0	1 (11)	0	0
rps8-rpl14	206	204	175	98	68	0	1 (13)	0	2 (30, 14)
rpl16 intron	999	988	1,013	97	89	0	1 (18)	1 (20o)	2 (11, 20o)
ycf2	6,846	6,840	6,834	99	99	0	0	0	1 (18)
trnL-ndhB	623	604	619	97	95	0	1 (12)	1 (21)	1 (11)
trnR-trnN	612	601	590	98	95	0	0	0	1 (15)
ndhF-rpl32	1,061	1,046	902	97	84	0	2 (10,10)	1 (12)	1 (197)
rpl32-trnL	1,432	1,390	1,368	95	75	0	3 (26, 37, 14)	2 (17p, 22)	6 (26, 13, 17, 13, 17p, 42)
ccsA-ndhD	263	263	292	99	79	0	2 (14, 12)	0	0
psaC-ndhE	245	245	259	99	84	0	1 (13)	0	0
ndhE-ndhG	233	252	262	92	79	1 (19)	1 (26)	0	0
ndhG-ndhI	368	364	394	99	75	0	2 (33, 10)	0	1 (15)
ndhA intron	1,154	1,141	1,142	97	86	1 (10)	1 (25)	1 (23)	1 (42)
ycf1	5,640	5,640	5,616	99	92	0	1 (30)	0	4 (12, 12, 15, 21)
rpl2-trnH	294	220	340	73	65	0	2 (23, 20)	1 (75)	0

a–m The same chloroplasts adjunct

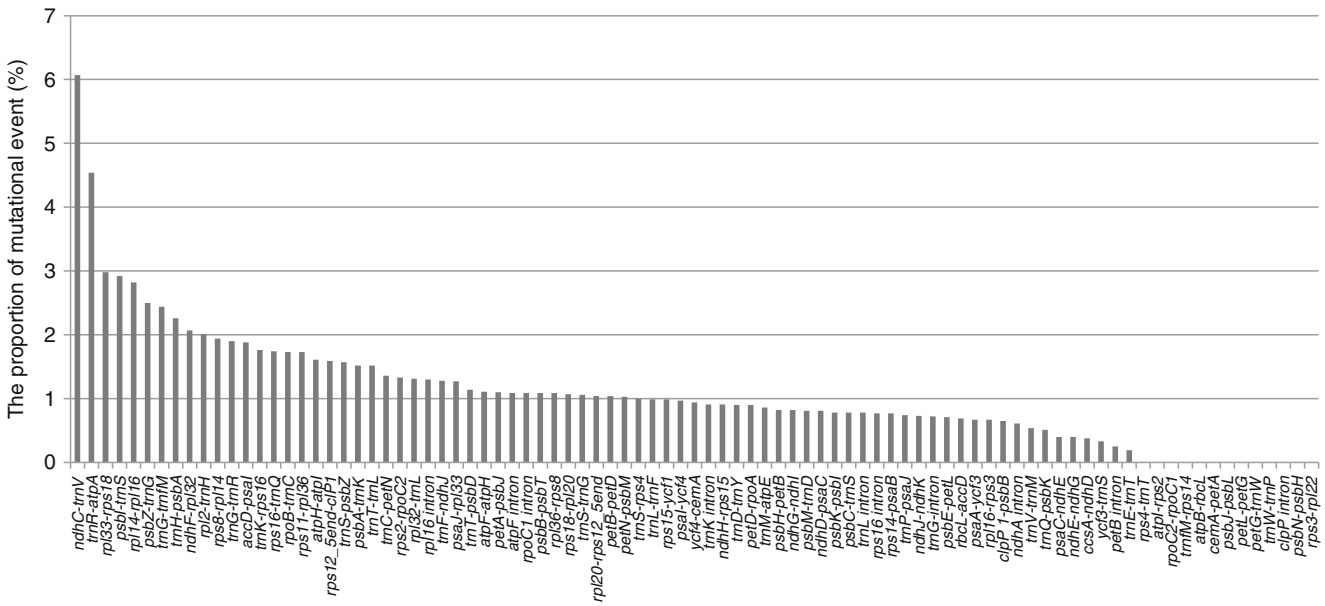

Fig. 4 The proportion of mutational events between *Pyrus* and *Malus* within 89 noncoding chloroplast DNA regions. The proportion of mutational events=(NS+ID+IV)/L×100. NS=the number of nucleotide substitutions, ID=the number of indels, IV=the number of inversions, and L=the aligned sequence length

The genomic sequence was annotated using the program Dual Organellar GenoMe Annotator (Wyman et al. 2004). Searches against a custom database of the previously published chloroplast genomic sequences using BLASTX were used to verify the predicted annotations (Altschul et al. 1990). Intron positions and rRNA genes were determined based on those of the *Prunus* chloroplast genome (Jansen et al. 2011). The tRNA genes were confirmed using tRNAscan-SE (ver. 1.23) (Lowe and Eddy 1997). The circular chloroplast genome map was drawn by the GenomeVx program (Conant and Wolfe 2008).

Translocation and inversion were examined by pair-wise comparisons between three genomes using PipMaker (Schwartz et al. 2000). Multiple alignments of the 11 complete chloroplast genomes *Malus* × *domestica* (GDR/Genome Database for Rosaceae; http://www.rosaceae.org/projects/apple_genome), *P. persica* (NC_014697; Jansen et al. 2011), *N. tabacum* (NC_001879; Shinozaki et al. 1986), *Cucumis sativus* (NC_007144; Kim et al. 2006), *Vitis vinifera* (NC_007957; Jansen et al. 2006), *Arabidopsis thaliana* (NC_000932; Sato et al. 1999), *Citrus sinensis* (NC_008334; Bausher et al. 2006), *Spinacia oleracea* (NC_002202; Schmitz-Linneweber et al. 2001), *Zea mays* (NC_001666; Maier et al. 1995), *Oryza sativa* (NC_001320; Hiratsuka et al. 1989), *Triticum aestivum* (NC_002762; Ogihara et al. 2000), and *P. pyrifolia* (present study) as the reference were performed using VISTA (Mayor et al. 2000). The number of nucleotide substitutions, indels, and inversions between *Pyrus* and *Malus* were tailed for 89 noncoding chloroplast DNA regions (≥100 bp). Indels, nucleotide substitutions, and inversions were scored as independent, single characters. The proportion of mutational events=(NS+ID+IV)/*L*×100, where NS=the number of nucleotide substitutions, ID=the number of indels, IV=the number of inversions, and *L*=the aligned sequence length was calculated as above.

Phylogenetic analyses were performed on an aligned data matrix that included 35 taxa of angiosperms and 81 protein coding genes (*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *ccsA*, *cemA*, *clpP*, *infA*, *matK*, *ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*, *petA*, *petB*, *petD*, *petG*, *petL*, *petN*, *psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*, *rrn16*, *rrn23*, *rrn45*, *rrn5*, *rpl22*, *rpl23*, *rps16*, *rbcL*, *rpl14*, *rpl16*, *rpl2*, *rpl20*, *rpl32*, *rpl33*, *rpl36*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps11*, *rps12*, *rps14*, *rps15*, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *ycf3*, *ycf2*, and *ycf4*). Amino acid sequences were aligned using Multiple Sequence Web viewer and Alignment Tool (http://mswat.ccbb.utexas.edu) and manually adjusted. The amino acid alignment was used to constrain the nucleotide alignment. Maximum parsimony (MP) analysis was performed using PAUP version 4.0b (Swofford 2003). MP analysis was performed with 100 random additional replicates and TBR branch swapping with the Multrees option. Nonparametric bootstrap analyses (Felsenstein 1985) were performed for 1,000 replicates with one random additional replicate and TBR branch swapping with the Multrees option.

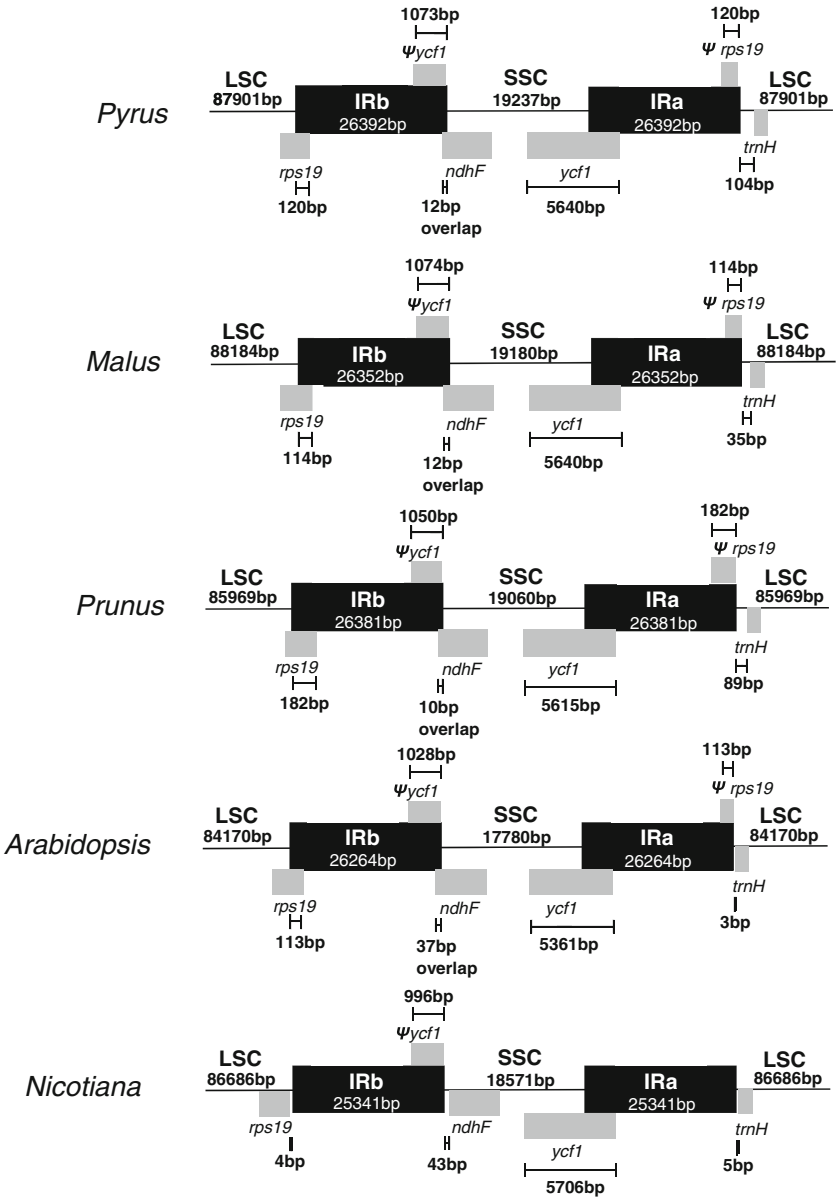

Fig. 5 Detailed view of the inverted repeat-single copy (IR/SC) border regions among five chloroplast genomes. Annotated genes or portions of genes are indicated by gray boxes above or below the genome

Results

Genome assembly and validation

Pyrus genomic DNA was sequenced using 454 Life Sciences technology on the GS FLX system. A total of 2,654,964 reads were generated with an average length of 375 bases that covered 995 Mb. These reads were cleaned and the remaining reads (2,254,379 reads with an average 410 bases) were assembled with reference to the *N. tabacum* chloroplast genome. The average genome sequencing depth of each nucleotide on the *Pyrus* chloroplast genome was 178×, and 60,313 chloroplast-related reads (2.68%) were collected with an average length of 426 bases that covered 25 Mb. There were 151 gap (mis-assembled) regions and 28 uncertain homopolymers in the assembled sequences.

Size, gene content, and organization of the *Pyrus* chloroplast genome

The complete chloroplast genome of *Pyrus* (GenBank/EMBL/DDBJ accession number: AP012207) was 159,922 bp in length (Fig. 1) and included a pair of IR of 26,392 bp separated by one small and one large single copy region (SSC and LSC) of 19,237 and 87,901 bp, respectively. The *Pyrus* chloroplast genome contained 113 unique genes, 17 of which were duplicated in the IR, giving a total of 130 genes (Table 1). There were four ribosomal and 30 tRNA genes, and seven tRNA genes and all the rRNA genes were duplicated in the IR. Eighteen genes contained one or two introns, and six of these were located in tRNAs. The numbers and kinds of tRNA genes from the *Pyrus* chloroplast genome were identical to that of well-characterized vascular plants. The genome consisted of

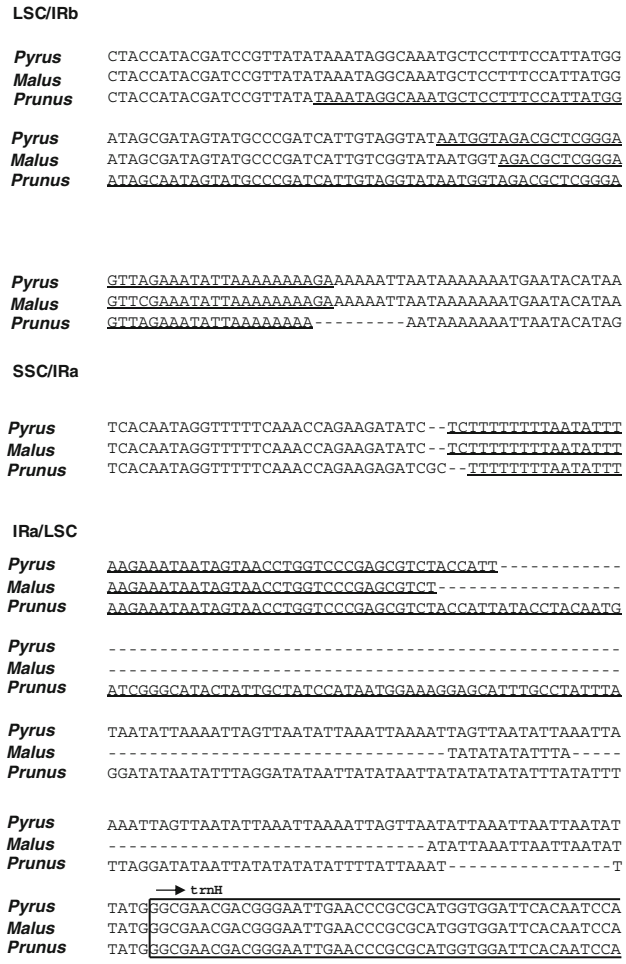

Fig. 6 Alignment of the IRb/SSC border region in the genus *Pyrus, Malus* and *Prunus*. Underlined sequences are within the IR

55.48% coding regions and 44.52% noncoding regions, including both intergenic spacers and introns. The overall GC and AT content of the *Pyrus* chloroplast genome was 36.58% and 63.42%, respectively.

Comparison of the whole chloroplast genome among angiosperms

Dot plot analysis showed that gene order and organization in *Pyrus* was similar to *Malus* and *Nicotiana* (Fig. 2). Neither translocation nor inversion were detected in the three plants species. *Malus* was shown to be the closest relative to *Pyrus* with global alignments using VISTA (Fig. 3). There were considerable differences between *Pyrus* and monocots such as large inversions with the identity plot patterns. Length mutations, such as indels more than 10 bp, were picked up by comparing the sequences of *Pyrus*, *Malus*, and *Prunus* (Table 2). With the exception of *ycf1* and *ycf2* which had deletions in the coding region, all the indels were detected in the spacer or intron regions. The indels of *ycf1*, *ycf2*, *trnL-ndhB*, and *trnR-trnN* were located within the IR region. In total, 46 indels of 23 insertions and 23 deletions were detected when comparing the chloroplast genomes of *Pyrus* and *Malus*. In the comparison between *Pyrus* and *Prunus*, there were a total of 174 indels (69 insertions and 105 deletions). Three insertions and 13 deletions were detected in *Pyrus* compared to the same locus in *Malus* and *Prunus*. A comparison of the mutational events within 89 noncoding chloroplast DNA regions in *Pyrus* and *Malus* revealed that intergenic spacer *ndhC-trnV* was most variable with a percentage variability of 6.07% (Fig. 4). The average percentage variability was 1.10%. In intergenic spacer *ndhC-trnV*, 41 mutational events were detected; 32 were nucleotide substitutions and nine were length mutations in which two indels had tandem repeats.

IR expansion/contraction

Figure 5 shows the detailed IR–SC border positions with respect to the adjacent genes in *Pyrus*, *Malus*, *Prunus*, *Arabidopsis*, and *Nicotiana*. In *Pyrus*, the IRa/SSC borders (position 133,432) were located in the 3′ region of the *ycf1* gene and created the *ycf1* pseudogenes of length 1,073 bp, at the IRb/SSC border (position 114,194). A similar structure was also observed in *Malus*, *Prunus*, *Arabidopsis*, and *Nicotiana*. In *Pyrus*, the IRa/LSC borders (position 159,823) were located downstream of the noncoding region of *trnH-GUG* while the IRb/LSC borders (position 87,803) were located within the coding region of *rps19*. Thus, a 3′-truncated *rps19* pseudogene was created at the other border (IRa/LSC) with a length of 120 bp (position 159, 703–159, 823). These features were similar in *Malus*, *Prunus*, and *Arabidopsis* but in *Nicotiana*, there was no *rps19* pseudogene at the other border (Fig. 5). Figure 6 compares the sequences of the IR/SC junction regions of *Pyrus*, *Malus*, and *Prunus*. In *Prunus*, the IR/LSC borders were 62 bp longer in the IR region than in *Pyrus* within the IR region. In addition, there were length mutations between the IRa/LSC junction and *trnH* but there were fewer length mutations in the IR/SSC borders. The IR/SSC junctions of *Pyrus* and *Malus* were only 2 bp longer than that of *Prunus* within the IR region.

Simple sequence repeats

Chloroplast simple sequence repeats (SSR), which align more than 10 repeated motifs, were investigated. Sixty-seven SSR loci from the *Pyrus* chloroplast genome were identified (Table 3). Thirty-one A stretches (10–22 bases), 34T stretches (10–24 bases), and 2C stretches (10 bases) were located within the genome but there were no G stretches. Forty-eight of the 67 SSR loci were detected in the intergenic spacers and were composed of A or T stretches. Fourteen of the 67 SSR loci were found within

Table 3 Distribution of simple sequence repeats (SSRs) loci in the *Pyrus* chloroplast genome

Length (bp)	Number and position of SSR			
	A stretch	C stretch	T stretch	G stretch
10	6 (52,309, 67,046, 68,686, 116,530, 117,842, 143,772)	2 (5,441, 5,637)	10 (6,536, 9,734, 27,455, 38,621, 57,783, 73,964, 87,074, 103,845, 123,749, 131,598)	0
11	6 (14,562, 28,431, 63,164, 74, 800, 81,437, 117,006)	0	9 (1,599, 2,745, 12,515, 13,235, 17,410, 19,659, 52,266, 71,780, 125,691)	0
12	4 (7,942, 8,081, 15,309, 125,556)	0	3 (9,371, 24,105, 28,843)	0
13	2 (47,551, 74,642)	0	2 (54,660, 68,391)	0
14	2 (32,217, 38,576)	0	0	0
15	2 (184, 70,641)	0	4 (166, 13,235, 60,492, 74,702)	0
16	4 (28,431, 39,192, 68,717, 132,226)	0	0	0
17	1 (4,748)	0	0	0
18	0	0	2 (72,499, 84,188)	0
19	2 (45,147, 50,289)	0	1 (83,025)	0
20	0	0	0	0
21	1 (6,998)	0	1 (15,564)	0
22	1 (116,439)	0	0	0
23	0	0	1 (117,289)	0
24	0	0	1 (66,621)	0
Total	31	2	34	0

The loci are indicated in parentheses

the intron region but only four were located in the gene coding regions.

Phylogenetic analysis

Phylogenetic analyses were performed on an aligned data matrix that included 35 taxa of angiosperms and 81 protein-coding genes. The total length of the aligned data set was 62,384 nucleotides. MP analyses resulted in a single, fully resolved tree with a length of 102,898, a consistency index of 0.48 and a retention index of 0.49 (Fig. 7). Bootstrap analyses indicated that 30 of the 32 nodes were supported by values ≥95%. The remaining two nodes had bootstrap values 90%. The MP trees were largely congruent with each other and with recent phylogenetic analyses based on complete chloroplast genomes (Jansen et al. 2007). MP analyses provided strong support for the monophyly of EurosidIand for the placement of *Pyrus* and *Prunus* in that clade as sister to *Morus*.

Discussion

The sequence of the complete chloroplast genome of *Pyrus* was determined using pyrosequencing (Roche 454 GS FLX Titanium) and assembled with the chloroplast genome of *N. tabacum* as reference sequence. Pyrosequencing has recently been used to determine the entire chloroplast sequence from, for example, *Phoenix* (Yang et al. 2010). When we started this research, the complete chloroplast genome sequence of the *Malus*×*domestica* and *P. persica* had not yet been released. As a result, there were 151 gap regions in the assembled sequences. Those regions were completed by Sanger sequencing using 99 primer pairs (ESM Table 1). If we assembled with the chloroplast genome of the other rosaceous, plants such as *Malus* or *Prunus*, number of the gaps might have been minimalized.

Homopolymers are stretches of the same nucleotide sequence and their contribution to technical sequencing errors when using 454 systems is well documented (Moore et al. 2006; Huse et al. 2007). The *Pyrus* chloroplast genome contained 28 uncertain homopolymers in the assembled sequence. This error is intrinsic to pyrosequencing and cannot be improved by increasing coverage.

The organization of the *Pyrus* chloroplast genome with two copies of an IR separating the SSC and LSC regions was identical to well-known angiosperm chloroplast genomes (Palmer 1991). The genome size, 159,919 bp, was also within the known size range for angiosperms. Gene order in the *Pyrus* genome was identical to that in *N. tabacum*. The gene content of *Pyrus* was also very similar to most other angiosperm chloroplast genomes (Raubeson and Jansen 2005). Chloroplast DNAs can be classified into three groups: chloroplast DNAs lacking IRs (group I), chloroplast DNAs containing IRs (group II), and chloroplast DNAs with tandem repeats (group III) (Sugiura 1992).

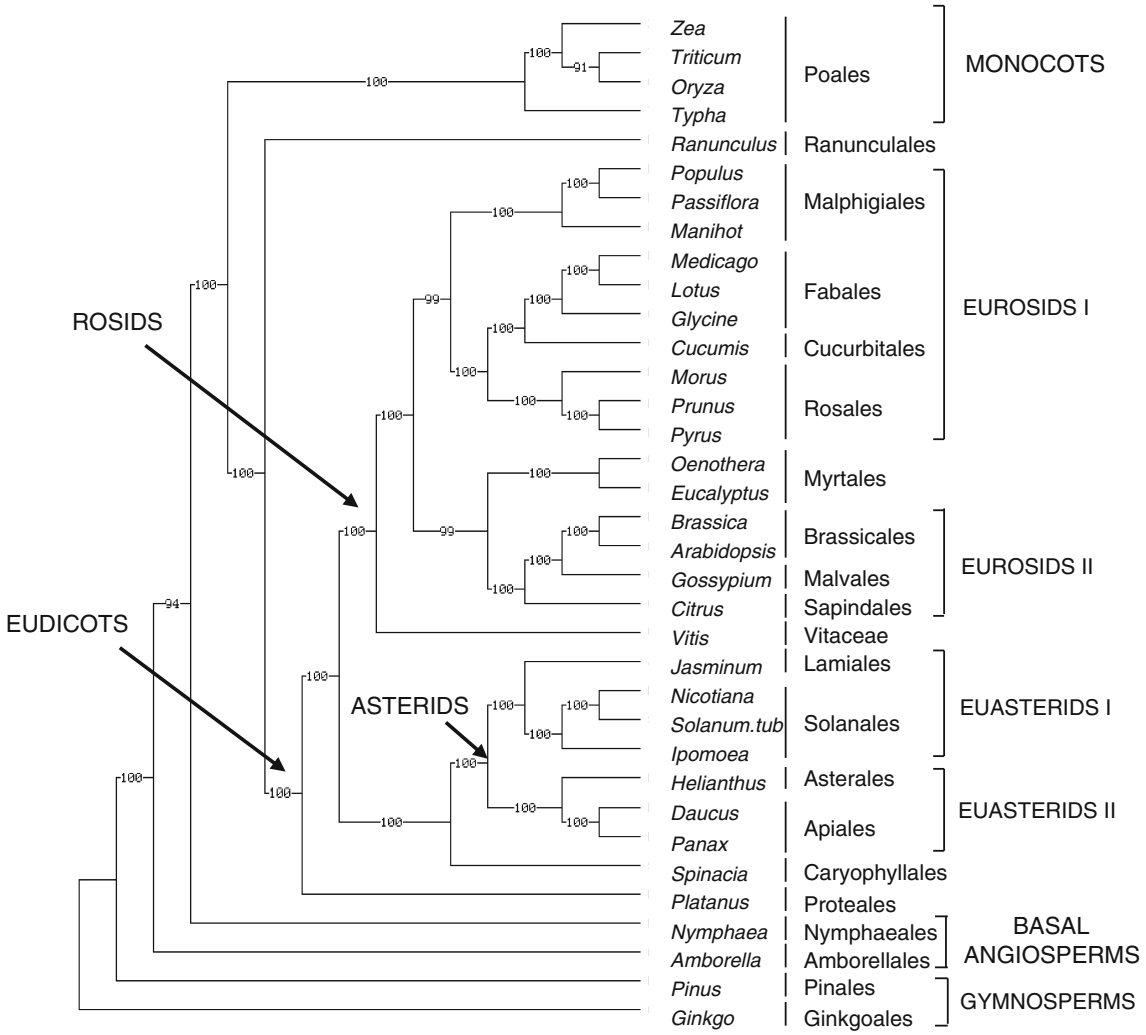

Fig. 7 Phylogenetic relationships of 35 taxa based on 81 chloroplast genes using maximum parsimony (MP). Bootstrap values (1,000 replications) are shown at the nodes

Almost all algae and higher plants belong to group II. In this group, *N. tabacum* is the most representative of land plants, which probably reflects the ancestral gene order among higher plants. Therefore, the *Pyrus* chloroplast genome can be considered as a standard type in higher plants.

A total of 220 indels of more than 10 bp were found in *Pyrus* by a comparison between *Pyrus*, *Malus*, and *Prunus* (Table 2). Most of these indels may have originated from slipped-strand mispairing of surrounding sequences (Levinson and Gutman 1987) or illegitimatic recombination events (Milligan et al. 1989; Ogihara et al. 1988; Shimada and Sugiura 1989). Only five indels were seen in the IR region. The sequences in the IR regions diverged at slower rates compared to the sequences of the LSC and SSC regions (Kim and Lee 2004). The results of this present study are consistent with this and support the idea that the stabilizing effect of the IR regions by genetic recombination is the main cause of their sequence conservation.

The largest indel in *Pyrus* was a 229 bp deletion (compared with *Malus)* between *accD-psaI* and was revealed by RFLP analysis based on physical mapping and DNA sequencing (Katayama and Uematsu 2003; Katayama et al. 2012). The intergenic region between *rbcL-cemA* which includes *accD-psaI* has previously been reported as a hypervariable region and might represent an intra molecular recombinational hot spot mediated by short direct repeats in the chloroplast DNA of Poaceae and Rosaceae (Ogihara et al. 1988; Shaw et al. 2007).

The indels that we have discovered might have numerous, important applications in systematics and evolutionary biology such as elucidating the origin of domesticated species (Wills and Burk 2006), tracing biogeographic movements (Ickert-Bond and Wen 2006; Schonswetter et al. 2006a, b), and clarifying complex relationships among species (Shaw and Small 2005). Sixteen unique indels to *Pyrus* were detected by comparing the chloroplast genomes of the

Rosaceae *Pyrus*, *Malus*, and *Prunus* (Table 2). It is highly likely that these indels will be very useful as intraspecific DNA markers in *Pyrus*.

By comparing 89 noncoding chloroplast DNA regions in *Pyrus* and *Malus*, we identified highly variable regions such as intergenic spacers of *ndhC-trnV* and *trnR-atpA* (Fig. 4). Previously, Shaw et al. (2007) reported 13 highly variable regions in the chloroplast genome of seven angiosperms including *Prunus*; intergenic spacer *ndhC-trnV* was one of these. However, highly variable region such as intergenic spacer *trnR-atpA* was not reported. The aligned sequence length of these regions was 674 and 887 bp, respectively (data not shown). The highly variable regions such as intergenic spacers of *ndhC-trnV and trnR-atpA* will be useful as intraspecific DNA markers. Katayama and Uematsu (2005) reported a hypervariable region was located within a region of about 9.1 kb which includes intergenic spacer *psbA-atpA* in the left border of LSC of *Prunus* chloroplast genome in the physical mapping. In the present study, the same trend was apparent within a region of about 10.1 kb which includes intergenic spacer *trnH-atpA* (position 77–10,979) in comparison of chloroplast genomes in *Pyrus* and *Malus* (Fig. 4). This highly variable region may be specific to Rosaceae.

The border between the two IR/LSC and the two IR/SSC regions usually varies between species as one of the highly variable regions in chloroplast genome, even between closely related genera of the same family (Kim and Lee 2004). Expansions and contractions of IR regions often create the large length variations between chloroplast genomes (Palmer et al. 1988; Raubeson and Jansen 2005; Wakasugi et al. 1994). The expansions/contractions of IR are probably mediated by gene conversion and recombinational repair of double-strand breaks (Goulding et al. 1996). In the present study, detailed comparisons of the IR boundaries in five genera highlighted the wide ranges of expansions and contractions possible in the IR (Fig. 5). In the comparison between *Pyrus*, *Malus*, and *Prunus*, the IR/LSC borders were more variable than the IR/SSC borders. In *Prunus*, the IR/LSC borders were 62 bp longer than those of *Pyrus*. In *Malus*, the length between the IRa/LSC junction and *trnH* was 69 bp shorter than in *Pyrus* (Fig. 6). It may be possible to use the differences directly as a DNA marker, which would be a useful evolutionary tool at both the intra- and interspecific levels.

A total of 67 SSR loci were identified in the *Pyrus* chloroplast genome. To date, chloroplast SSRs have been detected in *Pinus radiata* (Cato and Richardson 1996; Powell et al. 1995), *O. sativa* (Ishii et al. 2001), *Panax schinseng* (Kim and Lee 2004), *C. sativus* (Kim et al. 2006), and *Vigna radiate* (Tangphatsornruang et al. 2010). These chloroplast microsatellites can be useful in ecological and evolutionary studies because they are so variable at the inter- and intrapopulation levels. Therefore, *Pyrus* chloroplast microsatellites will be useful tools too.

There has been a rapid increase in the number of studies using DNA sequences from completely sequenced chloroplast genomes for estimating phylogenetic relationships among angiosperms (Bausher et al. 2006; Goremykin et al. 2005; Jansen et al. 2006, 2007; Leebens-Mack et al. 2005; Ravi et al. 2006; Ruhlman et al. 2006). The phylogenetic analyses reported here with expanded taxon sampling, were consistent with these earlier studies so our discussion will focus on relationships among Rosales (Fig. 7). The phylogenetic trees in this study indicate close relationships between *Pyrus* and *Prunus* with high bootstrap support (100%). These results agree with data confirmed by phylogenetic methods based on *matK* and *trnL-trnF* (Potter et al. 2002). Whole genome sequence of *Malus* in Pyreae could not be used for phylogenetic analysis in this study because updated online sequence of *Malus* was not annotated yet. Additional chloroplast genome sequences such as Spiraeae and Rosoideae are needed to generate a Rosaceae phylogeny based on whole genomes.

Acknowledgments Sincere appreciation is expressed to Dr. Anne Edwards, John Innes Centre, UK, for her English correction and useful suggestions. This work was partially supported by Grant-in-Aid (no. 17510196 and no. 19580031) for Scientific Research from the Ministry of Education, Science and Culture and by grant-in-aid (Genomics for Agricultural Innovation, DD-4040) for the Ministry of Agriculture, Forestry and Fisheries of Japan.

References

Aldasoro JJ, Aedo C, Garmendia FM (1996) The genus *Pyrus* L. (Rosaceae) in South-West Europe and North Africa. Bot J Linn Soc 121:143–158

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410

Bausher MG, Singh ND, Mozoru J, Lee SB, Jansen RK, Daniell H (2006) The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var. 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. BMC Plant Biol 6:21

Bell RL (1990) Pears (*Pyrus*). In: Moore JN, Ballington JR Jr (eds) Genetic resources of temperate fruit and nut crops I. International Society for Horticultural Science, Wageningen, pp 655–697

Bock R, Khan MS (2004) Taming plastids for a green future. Trend Biotechnol 22(6):311–318

Cato SA, Richardson TE (1996) Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don Theor Appl Genet 93:587–592

Challice JS, Westwood MN (1973) Numerical taxonomic studies of the genus *Pyrus* using both chemical and botanical characters. Bot J Linn Soc 67:121–148

Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24:861–862

Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr Genet 31:419–429

Daniell H, Ruiz ON, Dhingra A (2004) Chloroplast genetic engineering to improve agronomic traits. Methods Mol Biol 286:111–138

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evol 39:783–791

Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol Biol Evol 22:1813–1822

Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet 252:195–206

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Huse S, Huber J, Morrison H, Sogin M, Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143

Ickert-Bond SM, Wen J (2006) Phylogeny and biogeography of Altingiaceae: evidence from combined analysis of five non-coding chloroplast regions. Mol Phylogenet Evol 39:512–528

Iketani H, Manabe T, Matsuta N, Akihama T, Hayashi T (1998) Incongruence between RFLPs of chloroplast DNA and morphological classification in east Asia pear (*Pyrus* spp.). Genet Resour Crop Evol 45:533–539

Ishii T, Xu Y, McCouch SR (2001) Nuclear- and chloroplast microsatellite variation in A-genome species of rice. Genome 44:658–666

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leebens-Mack J, Cui L (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol 395:348–384

Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H (2006) Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol 6:32

Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA 104:19369–19374

Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H (2011) Complete plastid genome sequences of three rosids (*Castanea, Prunus, Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. Mol Biol Evol 28(1):835–847

Katayama H, Uematsu C (2003) Comparative analysis of chloroplast DNA in *Pyrus* species: physical map and gene localization. Theor Appl Genet 106:303–310

Katayama H, Uematsu C (2005) Structural analysis of chloroplast DNA in *Prunus* (Rosaceae): evolution, genetic diversity and unequal mutations. Theor Appl Genet 111:1430–1439

Katayama H, Tachibana M, Iketani H, Zhang S, Uematsu C (2012) Phylogenetic utility of structural alterations found in the chloroplast genome of pear: hypervariable regions in a highly conserved genome. Tree Genet Genomes. doi:10.1007/s11295-011-0442-y

Keeling PJ (2004) Diversity and evolutionary history of plastids and their hosts. Am J Bot 91:1481–1493

Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res 11:247–261

Kim JS, Jung JD, Lee JA, Park HW, Oh KH, Jeong WJ, Choi DW, Liu JR, Cho KY (2006) Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome. Plant Cell Rep 25:334–340

Kimura T, Iketani H, Kotobuki K, Matsuta N, Ban Y, Hayashi T, Yamamoto T (2003) Genetic characterization of pear varieties revealed by chloroplast DNA sequences. J Hort Sci & Biot 78:241–247

Knox EB, Palmer JD (1998) Chloroplast DNA evidence on the origin and radiation of the giant lobelias in eastern Africa. Syst Bot 23:109–149

Leebens-Mack J, Raubeson LA, Cui L, Kuehl J, Fourcade M, Chumley T, Boore JL, Jansen RK, dePamphilis CW (2005) Identifying the basal angiosperms in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol 22:1948–1963

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Lowe TM, Eddy SR (1997) tRNAscan-SE; a program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res 25:955–964

Maier RM, Neckermann K, Igloi GL, Kossel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628

Maliga P (2002) Engineering the plastid genome of higher plants. Curr Opin Plant Biol 5:164–172

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16:1046

Metzker ML (2005) Emerging technologies in DNA sequencing. Genome Res 15(12):1767–1776

Milligan BG, Hampton JN, Palmer JD (1989) Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol Biol Evol 6:355–368

Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, Soltis DE (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biol 6:17

Ogihara Y, Terachi T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc Natl Acad Sci USA 85:8573–8577

Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N, Takumi S, Ikeo K, Gojobori T, Murai R, Murai K, Matsuoka Y, Ohnishi Y, Tajiri H, Tsunewaki K (2000) Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. Plant Mol Biol Rep 18:243–253

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am Nat 130:S6

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (eds) The molecular biology of plastids. Academic, San Diego, pp 5–53

Palmer JD, Jorgensen RA, Thompson WF (1985) Chloroplast DNA variation and evolution in *Pisum*; patterns of change and phylogenetic analysis. Genetics 109:195–213

Palmer JD, Osorio B, Thompson WF (1988) Evolutionary significance of inversions in legume chloroplast DNAs. Curr Genet 14:65–74

Potter D, Gao F, Bortiri PE, Oh SH, Baggett S (2002) Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. Plant Syst Evol 23:77–89

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763

Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry R (ed) Diversity and evolution of plants—genotypic and phenotypic variation in higher plants. CABI, Wallingford, pp 45–68

Ravi V, Khurana JP, Tyagi AK, Khurana P (2006) The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. Tree Genet Genomes 3:49–59

Rehder A (1940) Manual of cultivated trees and shrubs, 2nd edn. Macmillan, New York, pp 426–451

Ronaghi M, Uhlen M, Nyren P (1998) A sequencing method based on real-time pyrophosphate. Science 281:363–365

Rubtsov GA (1944) Geographical distribution of the genus *Pyrus* and trends and factors in its evolution. Am Nat 78:358–366

Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell D (2006) Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. BMC Genomics 7:224

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 6:283–290

Schmitz-linneweber C, Maier RM, Alcaraz JP, Ccttet A, Herrmann RG, Mache R (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Mol Biol 45:307–315

Schonswetter P, Popp M, Brochmann C (2006a) Central Asian origin of and strong genetic differentiation among the populations of the rare and disjunct *Carex atrofusca* (Cyperaceae) in the Alps. J Biogeogr 33:948–956

Schonswetter P, Popp M, Brochmann C (2006b) Rare arctic-alpine plants of the European Alps have different immigration histories: the snow bed species *Minuartia biflora* and *Ranunculus pygmaeus*. Mol Ecol 15:709–720

Schwartz S, Zhang Z, Frazer K, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker: a web server for aligning two genomic DNA sequences. Genome Res 10:577–586

Shaw J, Small RL (2005) Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). Am J Bot 92:2011–2030

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparisons of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275–288

Shimada H, Sugiura M (1989) Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr Genet 16:293–301

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Sinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Swofford D (2003) PAUP: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland

Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N, Tragoonrung S (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17:11–22

Volk GM, Richards CM, Henk AD, Reilley AA, Bassil NV, Postman JD (2006) Diversity of wild *pyrus communis* based on microsatellite analysis. J Amer Soc Hort Sci 131:408–417

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91:9794–9798

Wills DM, Burk JM (2006) Chloroplast DNA variation confirms a single origin of domesticated sunflower (*Helianthus annuus* L.). J Hered 97:403–408

Wyman SK, Boore JL, Jansen RK (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Yamamoto T, Kimura T, Sawamura Y, Manabe T, Kotobuki K, Hayashi T, Ban Y, Matsuta N (2002) Simple sequence repeats for genetic analysis in pear. Euphytica 124:129–137

Yang M, Zang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssaiiem IS, Yu J (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS One 5(9):e12762