



Chloroplast DNA Underwent Independent Selection from Nuclear Genes during Soybean Domestication and Improvement

The chloroplast is one of the most important organs in plants because of its essential role in photosynthesis. Studies have shown that the chloroplast was once a free-living cyanobacteria and was integrated into the host species through endosymbiosis (Goksoyr, 1967), after which a large number of its genes had been donated to the host nuclear genome (Heins and Soll, 1998). Nonetheless most of the products of the transferred genes would be imported back to the organelle to participate in the photosynthesis processes (Keeling, 2010). On the other hand, plants also developed numbers of chloroplast subcellularly localized proteins encoded by nuclear genes to accomplish the biochemical processes in the chloroplast (Heins and Soll, 1998).

Domestication is one of the most significant activators in genome evolution (Diamond, 2002). With the rapid development of next generation sequencing technology, large numbers of genes that experienced selection during plant domestication have been identified from different species. However, all of these genome-wide analyses mainly focused on nuclear genes. As the chloroplast DNA is maternally inherited (Corriveau and Coleman, 1988), the chloroplast genomes principally would exhibit different patterns of genetic diversity compared to nuclear genomes (Birky et al., 1989). So far, the selection of chloroplast genes at a genome-wide level remains unclear. A systematic dissection of the chloroplast gene selection will determine the different evolutionary patterns of nuclear and chloroplast genomes, and will illuminate a clearer domestication history in a crop.

Soybean is an important crop that provides more than half of global oilseed production (Wilson, 2008). Cultivated soybean was suggested to be domesticated from wild soybean (*Glycine soja* Sieb. & Zucc.) approximately 5000 years ago and has been further developed in the last few decades (Carter et al., 2004). In this study, we investigated the evolutionary patterns of chloroplast genes during soybean domestication and improvement, and compared these patterns with those of nuclear genes.

To investigate the selection of different classes of chloroplast and nuclear genes, we firstly surveyed the

chloroplast, transferred, and nuclear genes in the soybean genome. The soybean chloroplast genome was reported to contain a total of 130 genes (Saski et al., 2005), including 85 coding genes, 37 tRNA genes and 8 rRNA genes (Fig. 1A). Among the predicated coding genes, *ycf15* in both the inverted repeat a (IRa) and IRb have evolved into pseudogenes, *rps19* and *ycf1* in IRa have been partially deleted, and *rps12* was annotated to cover more than half of the chloroplast genome. These four coding genes were excluded from this study.

To determine the genes transferred from endosymbiont to the soybean genome, all the soybean nuclear protein sequences (Schmutz et al., 2010) were BLAST searched against proteins from 39 cyanobacteria, 2 yeast and 15 bacterial species (Table S1) following a previously reported method (Martin et al., 2002). A total of 4608 soybean nuclear proteins were determined to be homologues to cyanobacteria proteins and were termed ‘transferred genes’ (Fig. 1B and Table S2). During evolution, some of the transferred genes had functionally diverged to be localized in other subcellular compartments rather than the chloroplast and may not participate in photosynthesis (Martin and Herrmann, 1998). Investigation of the predicted subcellular localizations of all the 4608 transferred genes revealed that 1720 proteins are subcellularly localized in the chloroplast (termed ‘TC genes’, transferred chloroplast subcellular localization genes), 2865 in non-chloroplast organelle (termed ‘TNC genes’, transferred non-chloroplast subcellular localization genes), and the localizations of 23 other proteins remained undefined (Fig. 1B). Subsequently, we investigated the predicted subcellular localizations of all the other proteins encoded by nuclear genes (the transferred genes were excluded and termed ‘nuclear genes’) in the soybean genome, and found that 7066 proteins are subcellularly localized in the chloroplast (termed ‘NC genes’, nuclear chloroplast subcellular localization genes), 37,204 proteins in non-chloroplast organelle (termed ‘NNC genes’, nuclear non-chloroplast subcellular localization genes), and the localizations of 5039 proteins were undefined (Fig. 1B). The 7066 proteins that are encoded by NC genes

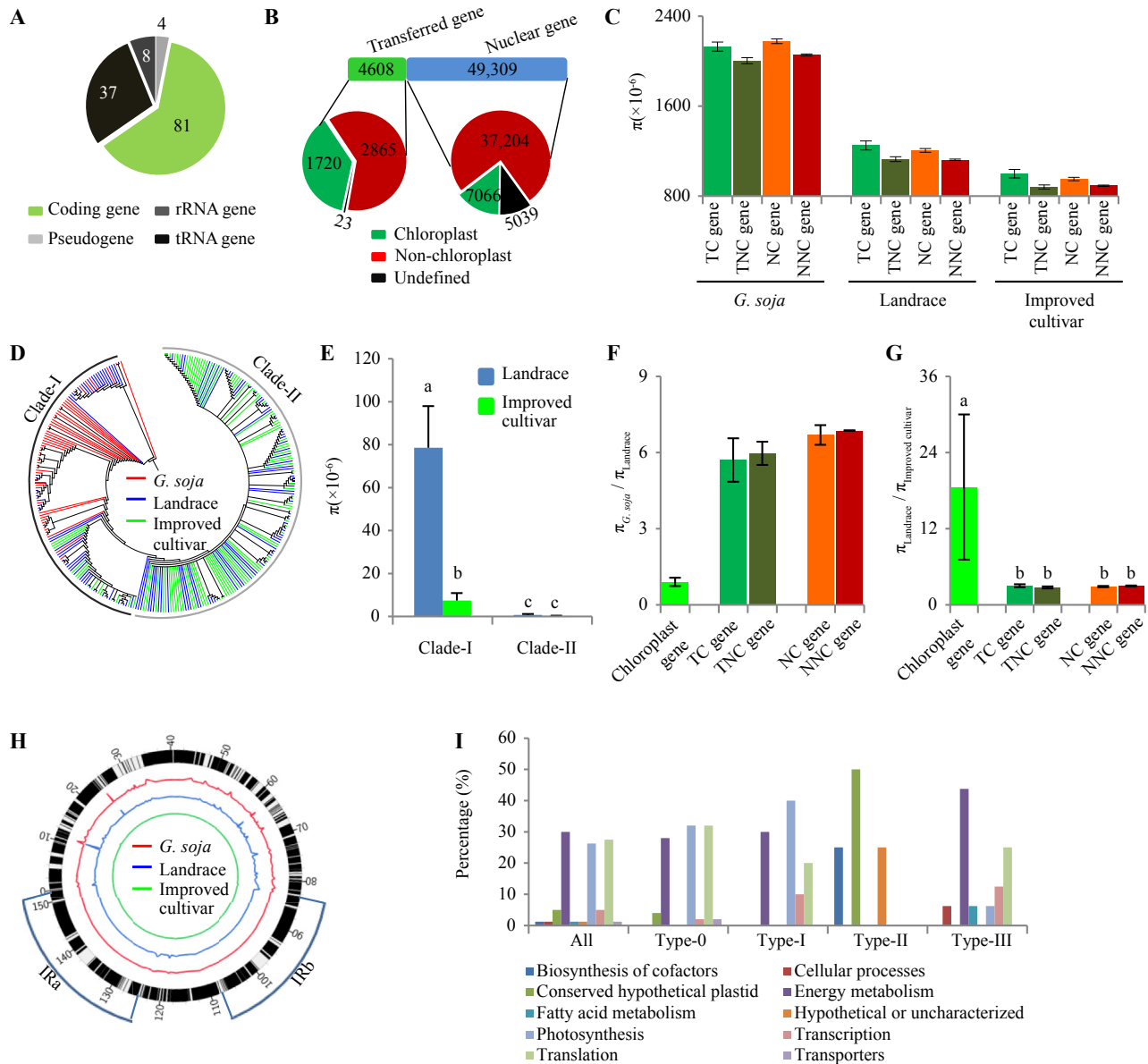


Fig. 1. Classification and genetic diversity of chloroplast, transferred, and nuclear genes during soybean domestication and improvement.

A: Classification of genes in the soybean chloroplast genome. **B:** Classification of transferred and nuclear genes in the soybean genome. **C:** Genetic diversity (π) of transferred and nuclear genes in wild soybeans, landraces, and improved cultivars. **D:** Phylogenetic relationship of wild and cultivated soybeans constructed from chloroplast gene SNPs. **E:** Comparison of the genetic diversity of landraces and improved cultivars belonging to clade-I and clade-II. **F:** Comparison of the genetic diversity ratio of different categories of chloroplast, transferred, and nuclear genes during soybean domestication. **G:** Comparison of the genetic diversity ratio of different categories of chloroplast, transferred, and nuclear genes during soybean improvement. **H:** Genetic diversity of individual chloroplast genes in wild soybeans, landraces and improved cultivars. The outer layer indicates the chloroplast genome, the black boxes indicate individual genes. The other three inner tracks (from the outside inward) indicate the genetic diversity of wild soybeans, landraces, and improved cultivars, respectively. **I:** Functional classification of the selected genes in each type. The bars in (C), (E), (F) and (G) indicate \pm standard error. Different lowercase letters show significant differences at $P < 0.05$.

could be newly evolved to involve in chloroplast biochemistry functions.

To explore the structural changes of different categories of chloroplast, transferred and nuclear genes, their exon length and intron number were investigated. It was reported that introns were rare in prokaryotic organisms, particularly for coding genes (Belfort et al., 1995). Consistently, we found that the chloroplast coding genes presented few introns. Whereas the transferred genes, either TC or TNC, exhibited

significantly more intron numbers than the chloroplast genes (Fig. S1A). The transferred genes also had longer exons than the chloroplast genes (Fig. S1B). We investigated the exon lengths of the genes from the 39 cyanobacteria genomes, and found the cyanobacteria genes that are homologues to the soybean transferred genes exhibited significantly longer exons than those of non-homologues (Fig. S1B), suggesting that the longer exon character of transferred genes might be inherited from the ancient endosymbiont. Another interesting finding

was that the genes subcellularly localized in chloroplast had higher exon numbers and shorter gene lengths than non-chloroplast subcellularly localized genes in the same category, such as in transferred and nuclear genes, respectively. The character differences among individual category genes confirmed that they may experience differentially selective regulatory processes in genome evolution after the cyanobacteria's endosymbiosis (Cullis et al., 2009).

Our previous phylogenetic analysis of 302 soybean accessions, including 62 wild soybeans (*G. soja*), 130 landraces and 110 improved cultivars, using SNPs from whole genome nuclear DNA, illuminated that wild soybeans could be clearly separated from the domesticated accessions to form one clade, meanwhile landrace and improved soybeans were grouped into another two main clades with a few mixtures inside (Zhou et al., 2015). Phylogenetic analyses showed that all of the transferred and nuclear genes, either subcellularly localized in the chloroplast or non-chloroplast exhibited similar patterns to the whole genome SNPs (Fig. S2). We found that the chloroplast subcellular localization genes, either from transferred genes or nuclear genes, exhibited higher genetic diversity (π) than non-chloroplast subcellularly localized genes in the same category (Fig. 1C), suggesting they experienced differentially selective pressures in genome evolution.

After mapping against the soybean chloroplast reference genome (Saski et al., 2005), we identified 206 high confidence SNPs. When the SNPs from chloroplast genes were used to explore the phylogenetic relationships, we found that the accessions could be divided into two main clades (Fig. 1D). Wild soybeans were mixed with a number of landraces and several cultivars to form clade-I, and the other landraces and improved cultivars were evenly mixed to form clade-II. The differences between phylogenetic trees from chloroplast and nuclear genes suggested that chloroplast and nuclear DNA were independently selected during soybean breeding, consistent with the fact that chloroplast DNA is maternally inherited (Corriveau and Coleman, 1988). When the genetic diversity was investigated, we found clade-II, both landraces and improved cultivars, exhibited significantly lower π values than clade-I. The π value of improved cultivars in clade-I was significantly lower than that of landraces, but was higher than that of clade-II (Fig. 1E). We proposed that at an earlier soybean domestication stage, multiple maternal lines were selected, after which the chloroplast genomes evolved into two main haplotypes. Whereas at later soybean improvement stage, haplotype-I experienced weak selection and was randomly utilized in the development of a small number of cultivated soybeans, but haplotype-II was strongly selected from a limited number of maternal lines and was soon fixed in the clade-II population.

If our hypothesis is correct, the chloroplast genes would exhibit different selective patterns from those of the nuclear genes during soybean domestication and improvement. We measured the ratio of their genetic diversity during soybean domestication ($\pi_{G. soja}/\pi_{Landrace}$) and improvement ($\pi_{Landrace}/\pi_{Improved cultivar}$). Our results showed that the genetic diversity

ratio of chloroplast genes during domestication was lower than that of nuclear genes (Fig. 1F). Nevertheless, it was significantly higher than that of the nuclear genes during improvement (Fig. 1G). The results supported our hypothesis that the chloroplast genome was independently selected from nuclear genes and underwent differential selective pressure at different soybean breeding stages, which might be caused by the fact that nuclear genome is two-parentally inherited and the chloroplast genome is maternally inherited.

To clarify the selection of each chloroplast gene during soybean domestication and improvement processes, we investigated the genetic diversity of individual genes in each population (Fig. 1H and Table S3). The results revealed that 91 of the chloroplast genes (termed 'type-0') did not show genetic diversity among all populations (Table S3). Their conservation indicated that they might be essential to the plant; accordingly strong natural selection had eliminated mutations in these genes. Another possibility might be due to the lower mutation rates in chloroplast DNA (Wolfe et al., 1987).

The soybean chloroplast genome, like most land plants (Raubeson, 2005), has a conserved organization and includes two copies of an inverted repeat (IR) (Saski et al., 2005). We found the sequences of the two IRs (IRa and IRb) did not show any genetic diversity among all populations (Fig. 1H), which is consistent with the observation that the substitution rate in IR regions was greatly reduced in comparison to single-copy sequences (Wolfe et al., 1987), indicating that the duplication status may decrease the evolutionary rate. It was reported that the soybean had undergone at least two polyploidy or whole genome duplication (WGD) events within the last 60 million years, resulting in approximately 75% of the genes present in more than one copy (Schmutz et al., 2010). We found that the duplication genes, in each of the transferred and nuclear categories, exhibited lower genetic diversity than that of the singletons in all the wild soybeans, landraces and improved cultivars (Fig. S3). The results suggested that although chloroplast and nuclear genes experienced differential selection, they were affected by the duplication status in the same manner.

We found that the chloroplast genes with genetic diversity variations could be classified into three types (Fig. S4). The genetic diversities of type-I and type-II both decreased during soybean domestication, but they were different in that after domestication, the type-I genes were fixed in the landrace (Fig. S4A) but type-II genes further experienced selection (Fig. S4B). Different from the first and the second types, the type-III genes exhibited a pattern of lower genetic diversity in wild soybeans, but with increased genetic diversity in landraces and subsequently being strongly selected in improved cultivars (Fig. S4C).

To elucidate if the different selection patterns are related to their functional variations, we checked their functional classification based on information from CpBase (<http://chloroplast.ocean.washington.edu/cpbase>). We found that genes related to photosynthesis were enriched in the type-I. Biosynthesis of cofactors, conserved hypothetical plastid, and hypothetical or uncharacterized genes were enriched in

type-II, while genes related to energy metabolism were enriched in type-III (Fig. 1I). The chloroplast genome is a useful tool for evolutionary studies because of its non-recombinant and uniparentally inherited nature (Provan et al., 2001). The domestication of nuclear genes in soybeans was well-studied, but the domestication of maternal lines remains unclear (Gai and Zhao, 2004). These results suggested that unlike the nuclear genes that may only come from a single or a few domestication events (Guo et al., 2010), chloroplast genes were selected from multiple maternal lines at an earlier soybean domestication stage and formed into two major haplotypes. Whereas at the later soybean improvement stage, one of the haplotypes was strongly selected from a limited number of maternal lines and soon fixed in the modern populations. The selection of the genes might be different between earlier and later domestication stages. The genes related to photosynthesis activity were strongly selected during earlier domestication stage, while the genes related to energy metabolism were strongly selected during later improvement processes.

Taken together, through comprehensive investigation and comparison of gene structure and genetic diversity among different soybean populations, we clarified how selection differentially acted upon different categories of chloroplast, transferred, and nuclear genes. The results presented here shed light on the co-evolution of chloroplast, transferred, and nuclear genes during domestication and improvement in palaeopolyploid plants.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91531304 and 31525018), “Strategic Priority Research Program” of the Chinese Academy of Sciences (Grant No. XDA08020202), a special program from the State Key Laboratory of Plant Cell and Chromosome Engineering (PCCE-TD-2012-01), and National Key Basic Research Program (No. 2013CB835200).

SUPPLEMENTARY DATA

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jgg.2016.01.005>.

Chao Fang^{a,b}, Yanming Ma^{a,c}, Lichai Yuan^d, Zheng Wang^a,
Rui Yang^a, Zhengkui Zhou^a, Tengfei Liu^{a,b}, Zhixi Tian^{a,*}

^aState Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

^bUniversity of Chinese Academy of Sciences, Beijing 100049, China

^cBeijing University of Agriculture, Beijing 102206, China

^dThe Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing 100193, China

*Corresponding author. Tel: +86 10 6480 3672, fax: +86 10 6480 1362.
E-mail address: zxtian@genetics.ac.cn (Z. Tian)

Received 8 October 2015
Revised 6 January 2016
Accepted 8 January 2016
Available online 28 January 2016

REFERENCES

- Belfort, M., Reaban, M.E., Coetzee, T., Dalgaard, J.Z., 1995. Prokaryotic introns and inteins: a panoply of form and function. *J. Bacteriol.* 177, 3897–3903.
- Birky, C.W., Fuerst, P., Maruyama, T., 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121, 613–627.
- Carter, T.E.J., Nelson, R.L., Sneller, C.H., Cui, Z., 2004. Genetic diversity in soybean. In: Boerma, H.R., Specht, J.E. (Eds.), *Soybeans: Improvement, Production and Uses*. American Society of Agronomy, Madison, pp. 303–416.
- Corriveau, J.L., Coleman, A.W., 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* 75, 1443–1458.
- Cullis, C.A., Vorster, B.J., Van Der Vyver, C., Kunert, K.J., 2009. Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? *Ann. Bot.* 103, 625–633.
- Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. *Nature* 418, 700–707.
- Gai, J.Y., Zhao, T.J., 2004. On the origin and evolution of cultivated soybean (*Glycine max* (L.) Merr.): a review and discussion. VII World Soybean Research Conference – VI International Soybean Processing and Utilization Conference – III Congresso Brasileiro De Soja, Proceedings, 517–517.
- Goksoyr, J., 1967. Evolution of eucaryotic cells. *Nature* 214, 1161.
- Guo, J.A., Wang, Y.S., Song, C., Zhou, J.F., Qiu, L.J., Huang, H.W., Wang, Y., 2010. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Ann. Bot.* 106, 505–514.
- Heins, L., Soll, J., 1998. Chloroplast biogenesis: mixing the prokaryotic and the eukaryotic? *Curr. Biol.* 8, R215–R217.
- Keeling, P.J., 2010. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 729–748.
- Martin, W., Herrmann, R.G., 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* 118, 9–17.
- Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D., 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA* 99, 12246–12251.
- Provan, J., Powell, W., Hollingsworth, P.M., 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* 16, 142–147.
- Raubeson, L.A., Jansen, R.K., 2005. Chloroplast genomes of plants. In: Henry, R. (Ed.), *Diversity and Evolution of Plants-Genotypic and Phenotypic Variation in Higher Plants*. CABI Publishing, Wallingford, pp. 45–68.
- Saski, C., Lee, S.B., Daniell, H., Wood, T.C., Tomkins, J., Kim, H.G., Jansen, R.K., 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Mol. Biol.* 59, 309–322.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., Jackson, S.A., 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183.

- Wilson, R.F., 2008. Soybean: market driven research needs. In: Stacey, G. (Ed.), *Genetics and Genomics of Soybean*. Springer Science+Business Media, LLC, New York, pp. 3–16.
- Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* 84, 9054–9058.
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., Fang, C., Shen, Y., Liu, T., Li, C., Li, Q., Wu, M., Wang, M., Wu, Y., Dong, Y., Wan, W., Wang, X., Ding, Z., Gao, Y., Xiang, H., Zhu, B., Lee, S.H., Wang, W., Tian, Z., 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414.