Contribution of plant–bacteria interactions to horizontal gene transfer in plants

Hongsheng Cai, Yingying Liu and Changhong Guo

ABSTRACT
Horizontal gene transfer (HGT) is recognized as a major driver of adaptive evolution in prokaryotes. However, HGT seems impossible in eukaryotes, particularly as recipients; therefore, debate rages regarding whether HGT takes place in eukaryotes, in addition to its potential mechanism or frequency. Bacterial symbionts, whether mutualistic, commensalistic or parasitic, have been considered potential donors for eukaryotes. In this study, we used a bacterial–plant interaction system to systematically investigate HGT in plants. In total, 373 HGT events were identified based on a pipeline procedure, and 90 HGTs were confirmed as true events, with 27.27%–86.5% sequence similarities. We propose that both ancient transfer and recent specific transfer (e.g. Agrobacteria) occurred in the course of plant evolution. The most enriched functional categories of the HGTs were metabolism processes of amino acids, cofactors and vitamins, and carbohydrates, and genetic information processing. Donor bacterial genera were significantly enriched in plant-associated bacterial groups, which indicated that plant–bacterial interaction facilitates HGT in plants. No clade- or species-specific HGTs were detected, and all occurred anciently during the origin of angiosperm plants. In addition, we identified 309 ‘one-species’ HGT events, and as expected, all the events could be accounted for as sequence errors or inaccurate annotations.

Introduction
All the currently living species represent less than one percent of all species that have ever lived on Earth [1]. They were highly successful considering their harsh growth and development conditions, e.g. unstable environmental conditions, dangerous predators and diseases. Although adaptive evolution occurs mainly via mutation and natural selection, it is a slow process and is often costly for organisms. Horizontal gene transfer (HGT), the movement of genetic material between organisms other than via the transfer of DNA from parent to offspring, is a relatively easy and rapid route for organisms to acquire novel traits and facilitate adaptation to environmental conditions [2,3]. Over the past few decades, following advancements in genome sequencing technologies, numerous studies have investigated HGT events among species. However, data from different studies are seemingly inconsistent, particularly with regard to HGT existence and its frequency in eukaryotes [4–6].

In prokaryotes, natural mechanisms i.e. transformation, conjugation and transduction facilitate HGT [7,8]. In addition, they are single-celled, which facilitates the passing of the horizontally transferred genes to the subsequent generation. Indeed, currently, HGT is considered a major driver of genome evolution and ecological adaptation in bacteria and archaea [9,10]. In contrast, multicellular eukaryotes possess somatic and germ cells, and it seems impossible for foreign genes to be translocated across germ cells and enter nuclear genomes. To date, the most likely HGT routes in eukaryotes are either via gene transfer across organelles, i.e. between mitochondria and chloroplast, or between closely associating organisms, i.e. symbiosis [11].

A bacterial endosymbiont, Wolbachia pipientis, which is a parasite in a number of insects, has been demonstrated to transfer genes to its hosts [12]. Some studies suggest that even an entire endosymbiont genome has been transferred to the Drosophila genome [13]. Another well-defined case of HGT from bacteria to eukaryotes...
is where Agrobacterium cause tumors in plants using transferred genes, and the phenomenon has been exploited as a modern genetic engineering tool [14,15]. Indeed, plant-associated microbes are potential sources of gene transfer into associated hosts. Among such microbes, bacteria are the most commonly reported donors [16]. The most common interactions between bacteria and plants can be classified into pathogenic, nitrogen-fixing and beneficial interactions, and they exhibit overlapping strategies of invading plant hosts [17]. However, it remains largely unclear whether such microbes have the capacity to transfer genes to their hosts, and how frequently, if at all.

A notable example of HGT in plants is the transfer of dozens of genes from hosts to the parasite plants [18], which reinforces the idea that genes could be transferred to plants from other organisms via mechanisms other than vertical transfer. Yue et al. showed that a high number of genes was transferred from bacteria to the genome of a moss, Physcomitrella patens; however, the underlying mechanism of transfer has remained elusive [19]. Particularly, it is unclear if plant–bacteria interactions facilitate such transfers. Therefore, based on the fact that close association or contact is considered essential for gene transfer, we hypothesized that bacterial–plant association is a major source of HGT in plants. In the present study, we test the hypothesis that genes are horizontally transferred to plants from bacteria by exploring potential gene transfer events in plants from plant–bacterial symbionts, backed by increased availability of completed genome sequences. We particularly focus on recent gene transfer events, and to minimize false HGT events, we select well-annotated plant genomes carefully.

Materials and methods

Collection of genome sequences

All the genomic sequences from bacteria and plants were downloaded from the National Center for Biotechnology Information (NCBI) database. Complete annotation was a major criterion for genome selection. All the nucleotide and protein FASTA files, in addition to the GenBank files, are publicly available.

Construction of ortholog groups and phylogeny

Forty-three genomes from angiosperms and 151 bacteria associated with plants were selected to construct ortholog groups using OrthoFinder [20]. In total, 3406 ortholog groups containing proteins from both bacteria and plants were formatted, and each group was further checked based on conserved motifs, resulting in 2987 groups. Orthologous protein sequences were aligned using MUSCLE [21]. Phylogenetic trees for each alignment were constructed using RAxML based on the GTRGAMMA model [22].

Detection of HGT from bacteria to plants

The HGT screening methods were according to Yang et al. with minor modifications [23]. For each phylogenetic tree, the ancestral node was defined as the node containing genes from both bacteria and plant species. Subsequently, at least two internal nodes with a bootstrap support cutoff of 50 were required for the positive one. Tree incongruent screen and visualization were performed using python scripts and ETE 3 package [24].

Validation of HGT events

Each identified HGT event was verified by the addition of more taxa and searching against a local refseq protein database for homologous sequences, and confirmed using hidden Markov models [25]. Phylogenetic trees were constructed with all the additional taxa for further justification. Considering we only selected bacteria associated with plants in the dataset, the above procedure was also used for screening other potential donor bacteria. Other measures of alleviating false positives included protein domain check and sequence statistic information [26]. The identified HGTs were annotated functionally using the KEGG pathway database (http://www.kegg.jp/). Gene expression analyses were performed in GENEVESTIGATOR [27].

Results

Schema for detection of HGT from plant-associated bacteria to plants

In the present study, we attempted to identify HGT events in certain clades, i.e. Rosid, Asterid, eudicot, monocot, or genes that are present in at least two genomes (Figure 1), which is more likely to be due to a horizontal transfer event. We also explored HGT events that were observed in specific species. All the identified trees were processed further through careful examination, including sampling more species, and checking other molecular characteristics.

Genes horizontally transferred from bacteria

In total, 373 HGT events were identified based on the pipeline procedure (> 2 species), and 90 HGTs were confirmed as true events (Supplemental Table S1). All
the HGTs were distributed unevenly across the angiosperm group, with no clade-specific HGTs observed. To further assign the donor species of the HGTs, we sampled more species from the refseq protein database and conducted BLAST searches for the top hits. As illustrated in Figure 2, donor species fell under different bacterial groups. In Alpha-protobacteria, Beta-protobacteria and Firmicutes, which host numerous bacterial genera that interact with plants, the plant-associated bacterial genera were enriched significantly for HGT ($p < 0.01$). The results suggest the contribution of bacterial–plant interactions to HGT from bacteria to plants.

In total, 309 phylogenetic trees indicated HGT events in a single plant species. Most of the genes had similarity scores of >97% with those from bacterial species. After careful manual examination, all were confirmed to be artifacts of sequencing error or inaccurate annotation (Supplemental Table S2).

**Function of horizontal transferred genes in plants**

Numerous genes associated with plant metabolism and development were identified in the KEGG analysis (Figure 3). The most enriched categories were metabolism processes of amino acids, cofactors and vitamins, and carbohydrates. Another dominant category was

Figure 1. Schematic representation of horizontal gene transfer (HGT) determination. (A) Potential HGT events that occurred in specific angiosperm lineages. (B) Potential HGT events that occurred in more than two species. (C) Potential HGT events that occurred in specific single species. Arrow indicates gene transfer to a specific lineage or species.

Figure 2. Number of HGT ortholog groups from donor bacterial groups. The horizontal axis indicates the donor bacterial groups and the vertical axis shows the number of corresponding HGT ortholog groups.
gene-related genetic information processing, e.g. transcription, translation, and repair.

Nearly 83% of the transferred genes coded for enzymes that are involved in plant metabolism under different categories, e.g. oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. Other types of transferred genes included transporters, inhibitors, DNA-repair proteins and signaling proteins. Such proteins have been repeatedly reported as donor genes during HGTs events across bacteria.

An example of the transferred genes is Glyoxalase I (GlyI). The phylogenetic trees were constructed using both the maximum likelihood and Bayesian methods. For Maximum Likelihood trees, the best fit model was the GTR model and a substitution rate distributed according to a discrete gamma with four categories. A phylogenetic Bayesian analysis was also performed using MrBayes 3.2 [28] with a Markov Chain Monte Carlo (MCMC) process using the same models and parameters as those established for the Maximum Likelihood analysis. These results supported the transfer of the gene from bacteria to plants (Figure 4). It is one of the enzymes in the glyoxalase pathway that catalyzes the conversion of toxic methyglyoxal to non-toxic D-lactate. A horizontal gene transfer event of nickel (Ni)/cobalt (Co)-dependent GlyI from bacteria to plants occurred, leading to Ni/Co-dependent GlyI enzyme activity in plants. Afterward, two gene duplication events occurred, which led to two-domain enzyme and multiple Ni/Co-dependent GlyI in plants. HGT was further confirmed through the analysis of the GlyI expression profiles from publicly available data (Supplemental Figures S1–S3). Such a HGT event has also been hypothesized in a previous study [29].

**Discussion**

Close contact is a prerequisite for HGT to occur between organisms, in addition to the foreign gene uptake and integration requirements. Surrounding bacteria are potential sources of genes for host plants [30], and the acquired genes could end up either as novel traits preventing pathogen attack or mitigating potential gene erosion in subsequent evolutionary processes. Using bacterial–plant interaction systems, the current study investigated HGT events from bacteria to angiosperm plants. A number of transferred genes were identified, and most participated in metabolism and genetic information processing, symbolizing adaptive evolution.

Notably, most of the transferred genes coded for enzymes belonging to different categories, e.g. oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. Such enzymes are associated with regular growth and development activities, and participate in diverse metabolic activities in organisms in nature. Considering that plants are exposed to shifting
environmental conditions within their habitats, enzymes facilitate their survival and reproduction. At the metabolic level, enzymes could facilitate adaptation to changing chemical conditions. Another enriched HGT family was transporters and inhibitors. The proteins are recognized from their roles in plant colonization of land, and adaptation to surrounding environments [16]. By assimilating such enzymes and transporters, the recipient plants could acquire a greater capacity to exploit resources available for enhanced adaption to environmental stress.

The distributions of the transferred genes in plant species were uneven. Gene duplication or gene loss events could have occurred in the course of evolution since gene transfer. Although we anticipated to observe HGTs that were associated with certain related plant species, e.g. monocots or eudicots, no clade-specific HGT events were observed, and all that were observed occurred potentially in ancient times during the origin of angiosperm plants, when the foreign genes could be much more easily integrated into the plant genome, as is observed currently in bacteria. If that is the case, then the cumulative effect or 70% rule, which is a key argument against HGT events in eukaryotes [4], should be re-evaluated. In line with the hypothesis, massive HGT events were identified in ancestors of land plants in a previous study, and the authors proposed that the genes that facilitate adaptation to terrestrial stress in land plants originated from soil bacteria by HGT [31].

However, there are often exceptions in biology. Certain bacterial species, e.g. Agrobacteria and rhizobia, represent classical examples that have the potential to transfer genes to their plant hosts frequently [32]. Given that the genome of its natural host, *Nicotiana tabacum* (common tobacco), is not annotated fully, it was not possible to explore the interaction between *N.tabacum* and Agrobacteria, although numerous studies have reported the transfer of genes from Agrobacteria to tobacco. The 70% rule is not comprehensive. We observed two transferred genes that had sequence similarities higher than 70%. The cases probably represent relatively recent specific transformation.

Considering that at least two species sharing genes similar to the donor gene was one of the criteria for the identification of HGT events, pipelines for dissecting HGT usually discard ‘one-species’ HGT events, since they are potentially ‘false positives’. In the present study, we identified 309 such cases, with sequence similarities in the 40–100% range and most of these occurred in certain species. We comprehensively re-analyzed the genes, and, as expected, all the events were identified as sequence errors or inaccurate annotations. This has also been observed in numerous studies, which raises concerns over whether true HGT events are present in eukaryotic genomics [5, 33,34]. The genes are listed in Supplemental Table S2, and could facilitate precise manual annotation activities in the future.

Multicellular eukaryotes, e.g. flowering plants, have natural barriers to foreign gene transfer [4]. Microbial symbionts, either beneficial or harmful microbes, represent potential donors of HGT. According to the results of the present study, plant-associated bacteria contribute to HGTs in plants, both based on ancient transfers or recent specific transfers. For specific transformation, e.g. Agrobacteria, the mechanism of DNA transfer has been unveiled by researchers over the last few decades, and the natural process has been developed into a more efficient system with broad biotechnological applications. However, with regard to the ancient transfer over the long course of evolution, when and how genes were transferred into plants remain largely matters of speculation, even though HGT is considered to have been critical for eukaryote evolution. More comprehensive and precisely annotated genome sequences, in addition to comparative genomic approaches that include more lineages, are required to enhance our understanding of eukaryote HGT.

**Conclusions**

By using a comparative genomic approach, this study confirmed the contribution of bacterial–plant interactions to HGT from bacteria to plants. Both ancient transfer and recent specific transfer occurred in the course of plant evolution. These results can explain the discontinuity of HGT from prokaryotes to eukaryotes.

**Acknowledgements**

The authors gratefully acknowledge the use of High Performance Computing facilities provided by Harbin Normal University.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Compliance with ethical standards**

This article does not contain any human or animal experiments to be subjected with ethical standards.
**Data availability statement**

The sequence data that support the findings of this study are openly available in GenBank at https://www.ncbi.nlm.nih.gov/genbank/, and a list of plant and bacterial species used in this study can be found in the supporting material Table S3.

**Funding**

This study was supported by The National Key Research and Development Program of China (2017YFD0101303), Heilongjiang Postdoctoral Fund (LBH-Z20160), and the National Natural Science Foundation of China (31770575, 31972507).

**References**

1. Newman ME. A model of mass extinction. J Theor Biol. 1997;189(3):235–252.
2. Baltrus DA. Exploring the costs of horizontal gene transfer. Trends Ecol Evol. 2013;28(8):489–495.
3. Wickell DA, Li FW. On the evolutionary significance of horizontal gene transfers in plants. New Phytol. 2020;225(1):113–117.
4. Danchin EG. Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube? BMC Biol. 2016;14(1):101.
5. Ku C, Martin WF. A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. BMC Biol. 2016;14(1):89.
6. Rossoni AW, Price DC, Seger M, et al. The genomes of polyeukaryotic cyanobacteria contain 1% horizontally transferred genes with diverse adaptive functions. elife. 2019;8:e45017.
7. Blesa A, Berenguer J. Methods to identify and analyze vesicle-protected DNA Transfer. Methods Mol Biol. 2020;2075:209–221.
8. Blokesch M. Natural competence for transformation. Curr Biol. 2016;26(21):R1126–R1130.
9. Hall RJ, Whelan FJ, McInerney JO, et al. Horizontal gene transfer as a source of conflict and cooperation in prokaryotes. Front in Microbiology. 2020;11:1569.
10. Vos M, Hesselman MC, Te Beek TA, et al. Rates of lateral gene transfer in prokaryotes: High but why? Trends Microbiol. 2015;23(10):598–605.
11. Bock R. The give-and-take of DNA: horizontal gene transfer in plants. Trends Plant Sci. 2010;15(1):11–22.
12. Dunning Hotopp JC. Horizontal gene transfer between bacteria and animals. Trends Genet. 2011;27(4):157–163.
13. Dunning Hotopp JC, Clark ME, Oliveira DCSG, et al. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science (New York, NY). 2007;317(5845):1753–1756.
14. Knydt T, Quispe D, Zhai H, et al. The genome of cultivated sweet potato contains agrobacterium T-DNAs with expressed genes: an example of a naturally transgenic food crop. Proc Natl Acad Sci U S A. 2015;112(18):5844–5849.
15. Quispe-Huamanquispe DG, Gheysen G, Kreuze JF. Horizontal gene transfer contributes to plant evolution: the case of agrobacterium T-DNAs. Front Plant Sci. 2017;8:2015.
16. Husnik F, McCutcheon JP. Functional horizontal gene transfer from bacteria to eukaryotes. Nat Rev Microbiol. 2018;16(2):67–79.
17. Cai H, Bai Y, Guo C. Comparative genomics of 151 plant-associated bacteria reveal putative mechanisms underlying specific interactions between bacteria and plant hosts. Genes Genomics. 2018;40(8):857–864.
18. Davis CC, Wurdack KJ. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from malpighiaceae. Science. 2004;305(5684):676–678.
19. Yue J, Hu X, Sun H, et al. Widespread impact of horizontal gene transfer on plant colonization of land. Nat Commun. 2012;3:1152.
20. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.
21. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–1797.
22. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313.
23. Yang Z, Zhang Y, Wafula EK, et al. Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. Proc Natl Acad Sci U S A. 2016;113(45):E7010–E7019.
24. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 2016;33(6):1635–1638.
25. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7(10):e1002195.
26. Ravenhall M, Skunca N, Lassalle F, et al. Inferring horizontal gene transfer. PLoS Comput Biol. 2015;11(9):e1004095.
27. Hruz T, Laule O, Szabo G, et al. Genevestigator v3: a reference expression database for the Meta-analysis of transcriptomes. Adv Bioinformatics. 2008;2008:420747.
28. Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–542.
29. Kaur C, Vishnoi A, Aryadasa TU, et al. Episodes of horizontal gene-transfer and gene-fusion led to co-existence of different metal-ion specific glyoxalase I. Sci Rep. 2013;3:3076.
30. Gao JG. Tracking the evolutionary innovations of plant terrestrialization. Gene. 2021;769:1579.
31. Cheng S, Xian W, Fu Y, et al. Genomes of subaerial zygnematophyceae provide insights into land plant evolution. Cell. 2019;179(5):1057–1067 e1014.
32. Lacroix B, Citovsky V. Transfer of DNA from bacteria to eukaryotes. MBio. 2016;7(4):e00863–16.
33. Husnik F, Guzeeva EA, Mantelin S, et al. Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode Globodera pallida to feed on host-derived sucrose. Mol Biol Evol. 2016;33(6):1571–1579.
34. Kikuchi T, Eves-van den Akker S, Jones JT. Genome evolution of plant-parasitic nematodes. Annu Rev Phytopathol. 2017;55:333–354.