

ORIGINAL ARTICLE

# Horizontal gene transfer from a flowering plant to the insular pine *Pinus canariensis* (Chr. Sm. Ex DC in Buch)

B Wang<sup>1</sup>, J Climent<sup>2</sup> and X-R Wang<sup>1</sup>

Horizontal gene transfer (HGT) is viewed as very common in the plant mitochondrial (mt) genome, but, to date, only one case of HGT has been found in gymnosperms. Here we report a new case of HGT, in which a mt *nad5-1* fragment was transferred from an angiosperm to *Pinus canariensis*. Quantitative assay and sequence analyses showed that the foreign *nad5-1* is located in the mt genome of *P. canariensis* and is nonfunctional. An extensive survey in the genus *Pinus* revealed that the angiosperm-derived *nad5-1* is restricted to *P. canariensis* and present across the species' range. Molecular dating based on chloroplast DNA suggested that the HGT event occurred in the late Miocene after *P. canariensis* split from its closest relatives, and that the foreign copy became fixed in *P. canariensis* owing to drift during its colonization of the Canary Islands. The mechanism of this HGT is unclear but it was probably achieved through either direct cell–cell contact or external vectors. Our discovery provides evidence for an important role of HGT in plant mt genome evolution.

*Heredity* advance online publication 21 January 2015; doi:10.1038/hdy.2014.118

## INTRODUCTION

Horizontal gene transfer (HGT), the transfer of genetic materials between sexually well-separated species, is a significant force in the evolution of eukaryotic genomes (Keeling and Palmer, 2008). Generally, HGT is less common in multicellular than in unicellular eukaryotes (Keeling and Palmer, 2008). In vascular plants, HGT appears to be rare in plastid and nuclear genomes but is frequent in the mitochondrial (mt) genome (Richardson and Palmer, 2007; Bock, 2010), and is facilitated by intimate physical associations between donor and recipient species through parasitism or epiphytism (Davis *et al.*, 2005; Mower *et al.*, 2010; Rice *et al.*, 2013; Xi *et al.*, 2013). The most extensive mt HGT events have been detected in the basal angiosperm *Amborella*, which acquired more than one megabase of foreign mtDNA, including four entire mt genomes from its epiphytes (Rice *et al.*, 2013). In cases where HGT has occurred between species without parasitic or epiphytic relationships, transfer of foreign DNA via vectors such as bacteria or fungi has been suggested as an alternative route (Bergthorsson *et al.*, 2003; Won and Renner, 2003; Bock, 2010). After gene transfer via HGT, foreign genes often undergo modifications such as gene conversion, differential gene loss and a change in mutation rate (Won and Renner, 2003; Hao *et al.*, 2010; Mower *et al.*, 2010). These processes are creative forces in the generation of mt genetic diversity and have an important role in plant mt genome evolution (Archibald and Richards, 2010; Hao *et al.*, 2010; Mower *et al.*, 2010).

To date, our knowledge and data about HGT in plants mainly originate from studies of angiosperms. Only one case of HGT has been reported in gymnosperms, in which an intron of a mt gene was transferred from an angiosperm to the ancestor of an Asian clade of *Gnetum* (Gnetales, gymnosperms) (Won and Renner, 2003). After the ancestral acquisition of the angiosperm-derived sequence, eight out of

20 species in this clade of *Gnetum* gradually lost the foreign copy, whereas all retained the native copy (Won and Renner, 2003). The causes and consequences of mt HGT remain largely unclear in gymnosperms owing to the limited number of cases discovered.

Here we report a new case of mt HGT from an angiosperm to a gymnosperm, the Canary Island pine *Pinus canariensis* (Pinaceae). *Pinus canariensis* originated from an ancient Mediterranean evolutionary center (Klaus and Ehrendorfer, 1989), and is currently restricted to the Canary Islands. We detected two copies of mt *nad5-1* (NADH dehydrogenase subunit 5 intron 1 plus its flanking exons 1 and 2) in *P. canariensis*. Phylogenetic analyses revealed that one of these was vertically inherited, whereas the other was horizontally transferred from an angiosperm species. To understand the specificity and history of the event, we surveyed sequence variation in the three genomes of the genus *Pinus* to establish the distribution of the angiosperm-derived *nad5-1* copy in the genus and the genomic location of this sequence; we estimated the timing of this HGT in *P. canariensis* and discussed its putative routes. Our discovery is important for understanding the role of HGT in the evolution of the mt genome in plants.

## MATERIALS AND METHODS

### Taxon sampling

To examine the presence of the angiosperm-type copy of *nad5-1* in the genus *Pinus*, we sampled 44 pine species to represent 9 of the 11 subsections of the genus *Pinus* recognized by Gernandt *et al.* (2005). Most species were represented by multiple accessions collected either from documented individuals grown by different institutions or from natural stands, and a total of 148 individuals were analyzed (Supplementary Table S1). For *P. canariensis*, 22 individuals were sampled, of which 16 were from four natural populations across the species' range, and the other six individuals were from a botanic garden (Supplementary Table S1). We retrieved the *nad5-1* sequences of 175

<sup>1</sup>Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden and <sup>2</sup>Department of Forest Ecology and Genetics, INIA-CIFOR, Madrid, Spain  
Correspondence: Dr B Wang, Department of Ecology and Environmental Science, Umeå University, Umeå 90187, Sweden.  
E-mail: baosheng.wang@emg.umu.se

Received 14 August 2014; revised 3 November 2014; accepted 10 November 2014

species from GenBank, spanning all major clades of land plants: angiosperms, gymnosperms, ferns and mosses (Supplementary Table S2).

### DNA isolation, PCR amplification and sequencing

Genomic DNA was extracted from needles or seedlings using a Plant Genomic DNA Kit (TIANGEN, Beijing, China) according to the manufacturer's instructions. The mtDNA *nad5-1* region in *Pinus* species was amplified using a pair of previously published primers (Wang *et al.*, 2000). This pair of primers generated two copies of *nad5-1* in *P. canariensis*. A BLAST search in the NCBI nucleotide database revealed that one of these was identical to the previously reported *nad5-1* sequence of *P. canariensis* (this copy is referred to hereafter as the gymnosperm-type), whereas the other was similar to those of angiosperm species (this copy is hereafter referred to as the angiosperm-type). To test the presence of angiosperm-type *nad5-1* in the genus *Pinus* further, we designed a pair of primers specific to this sequence, and surveyed 44 pine species (Supplementary Table S1). To test whether the remaining genomic regions surrounding the angiosperm-type *nad5-1* were also angiosperm-derived, we used primers specific to the neighboring *nad5* exons 1 and 2 of angiosperm species, and universal primers for *nad5* exon 2 of seed plants to amplify the regions connecting the *nad5-1* segment. The primers specific to *nad5* exons 1 and 2 of angiosperms were designed based on Genbank sequences, and the universal primers were from a previous study (Qiu *et al.*, 2006). The primer sequences, annealing temperature and sizes of each product are listed in Supplementary Table S3. The PCR products were subjected to electrophoresis using a 1.0% agarose gel, after which the desired band was cut from the gel and purified using a GFX PCR DNA and Gel Band Purification Kit (Amersham Pharmacia Biotech, Buckinghamshire, UK). The purified PCR products were sequenced directly on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). The sequences for each DNA region have been deposited in GenBank under accession numbers KP141747–KP141754 (Supplementary Table S2).

### Quantitative PCR

To determine the genomic location of the angiosperm-type *nad5-1* in *P. canariensis*, we applied a quantitative PCR (qPCR) strategy similar to that used by Mower *et al.* (2010). This approach is based on the fact that the copy numbers of the plastid (highest), mt (intermediate) and nuclear (lowest) genomes differ widely in a plant cell (Petit and Vendramin, 2007), and genes with a similar number of copies should be amplified at similar rates. Nine loci were selected for amplification in *P. canariensis*: two chloroplast (cp) loci (*rbcl* and *matK*), three native mt loci (*cox1*, *matR* and the gymnosperm-type *nad5-1*), three nuclear loci (*cco*, *cad* and *4CL*) and the angiosperm-type *nad5-1*. The qPCR primer sequences, annealing temperature and sizes of each product are listed in Supplementary Table S3. The qPCR assays were performed on a Bio-Rad CFX Real-time system (Bio-Rad, Hercules, CA) using iQ SYBR Green Supermix (Bio-Rad). Genomic DNA of three *P. canariensis* individuals was used as templates, a fivefold dilution series of each DNA was used in the amplification of the nine loci. The amplification efficiency for each locus was estimated using the Bio-Rad CFX manager version 3.1. We conducted a melt-curve analysis at the end of the PCR run to examine the specificity of the amplification. Each pair of primers demonstrated an efficiency >90% and amplified only the target region.

### Phylogenetic analyses

The phylogenetic analyses were performed using maximum parsimony (MP), maximum likelihood (ML) and Bayesian approaches. Gaps were treated as missing. MP analysis was implemented in PAUP version 4.0b10 (Swofford, 2002). Heuristic searches were performed using random sequence addition with 100 replicates, tree-bisection-reconnection branch swapping and the MulTrees option, and a maximum of 500 trees were saved per round. All character states were unordered and equally weighted. Bootstrap supports for the clades found in the most parsimonious trees were obtained from 1000 replicates using the same heuristic search settings. ML analysis was performed using RAxML version 7.3.0 (Stamatakis, 2006) under the GTRGAMMA model as suggested in the instructions. One thousand ML trees were generated to find the best-scoring ML tree, and topological robustness was investigated using 1000

nonparametric bootstrap replicates. Bayesian analyses were conducted using MrBayes version 3.2 (Ronquist *et al.*, 2012). The optimal model for the DNA sequence was determined using Modeltest 3.7 (Posada and Crandall, 1998). Four Markov Chain Monte Carlo runs were conducted, each comprising 10 million generations; trees were sampled every 100 generations, and the first 2.5 million generations of each run were discarded as the burn-in, to ensure that the chains were stationary. The remaining trees were used to calculate a strict consensus tree with the posterior probabilities for each split.

To assess the alternate topological placement of the angiosperm-type *nad5-1*, we performed the parsimony-based Templeton test (Templeton, 1983) and the likelihood-based Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999). For the Templeton test, we used one-tailed probabilities as suggested by Felsenstein (1985). For the Shimodaira–Hasegawa test, the significance levels of the differences in the likelihood scores for each comparison were checked for the null distribution obtained with the re-estimated log-likelihoods approximated with 1000 nonparametric bootstrap replicates. The Templeton and Shimodaira–Hasegawa tests were implemented in PAUP. Gene conversion between the angiosperm- and gymnosperm-type *nad5-1* was tested using RDP3.44 (Martin *et al.*, 2010). We set the highest acceptable *P*-value to 0.05 and default values for the remaining settings.

### Divergence time estimation

We extracted full chloroplast genome sequences of 38 species from Parks *et al.* (2012) to perform cpDNA-based divergence time estimation. These 38 species represent all 11 subsections of the genus *Pinus* recognized by Germandt *et al.* (2005), and three outgroups (*Picea sitchensis*, *Abies firma* and *Cathaya argyrophylla*) (Supplementary Figure S1). All seven species of the subsection *Pinaster*, including *P. canariensis*, were included in this analysis. Previous work shows that two loci (*ycf1* and *ycf2*) have very different mutation rates from other parts of the cp genome in pines (Parks *et al.*, 2009), thus these two regions were removed from this analysis. A likelihood ratio test in PAUP rejected a strict molecular clock for the cpDNA data ( $\chi^2 = 41445.87$ ,  $df = 36$ ,  $P < 0.001$ ), thus we used two relaxed molecular clock approaches to estimate the mutation rate and divergence time: one was a Bayesian method, with an assumption of independent lineage rate changes and was implemented in BEAST version 1.7.2 (Drummond *et al.*, 2012), and the other was a Bayesian estimation, with autocorrelated lineage rates and was implemented in Multidivtime (Kishino *et al.*, 2001). In each analysis, we set 85 million years ago (MYA) as a lower bound and 45 MYA as an upper bound for the divergence time between the two subgenera *Pinus* and *Strobus*, as proposed by Willyard *et al.* (2007).

For analysis in BEAST, we used a general time-reversible substitution model with Gamma distributed rate heterogeneity, an uncorrelated lognormal relaxed clock model and a Yule speciation model. We calibrated the tree by placing a uniform prior on the age of the split between the two subgenera, ranging from 45 to 85 million years (MY). The Markov chain Monte Carlo was run for 20 million generations, whereas trees and parameter estimates were sampled once every 2000 generations. The log file of the run was inspected using Tracer version 1.6 (<http://beast.bio.ed.ac.uk/Tracer>) to confirm that the chain was stationary and that the effective sample sizes were adequate (ESS > 200) for the estimated parameters. From the sampled posterior trees, the first 2500 (25%) were treated as the burn-in, and the remaining trees were used to summarize the node height and rate statistics using TreeAnnotator version 1.7.2 (part of the BEAST package).

For Multidivtime, the topology of the maximum clade credibility tree generated by BEAST was used as the target tree and reference topology. The maximum and the minimum constraints to the divergence between the subgenera were set to 85 and 45 MY, respectively. The prior for the mean and s.d. of the ingroup root age (rttm and rttmsd) were both set to 0.85; the mean and s.d. of the prior distribution for the rate of molecular evolution at the ingroup root node (rtrate and rtratesd) were both set to 0.02; and bigtime was set to 100. The Markov chain Monte Carlo was run for one million generations, and parameters were sampled every 100 generations after a burn-in of 0.1 million generations. Three independent runs were performed, and all generated similar results.

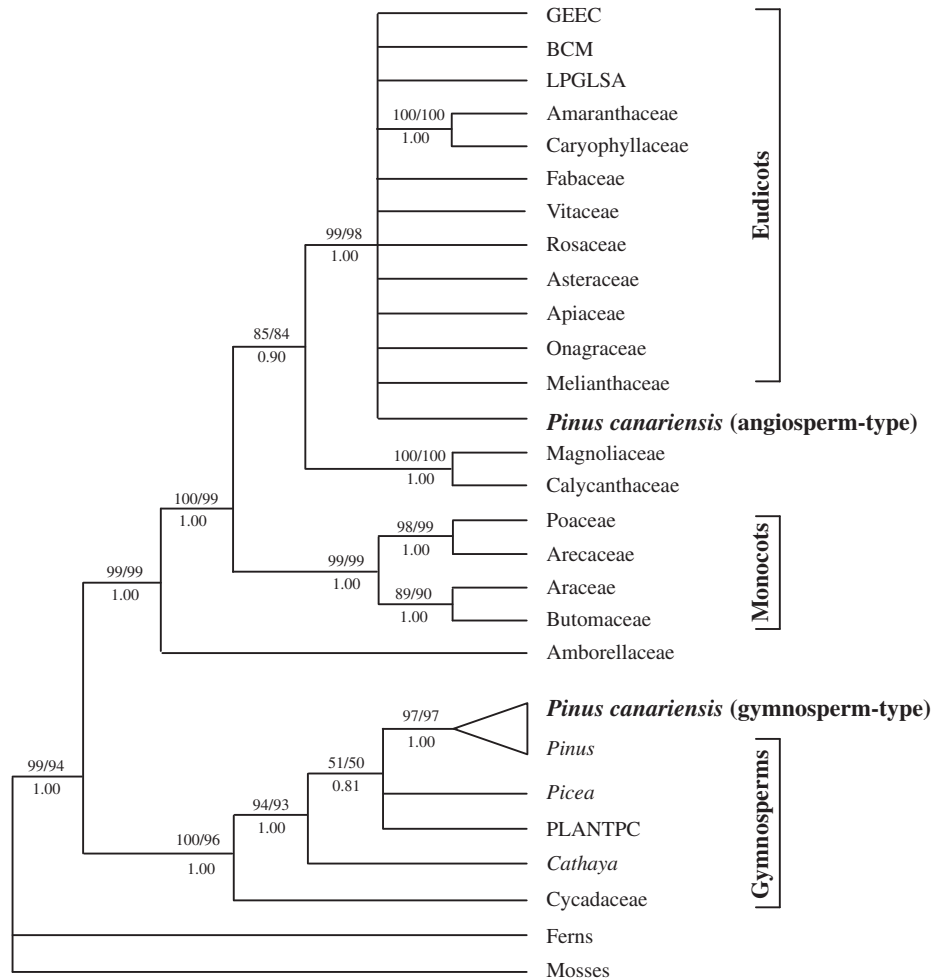
## RESULTS AND DISCUSSION

Using the primers previously developed for *nad5-1* in gymnosperms (Wang *et al.*, 2000), we co-amplified both angiosperm- and gymnosperm-type *nad5-1* in *P. canariensis* individuals. The angiosperm-type copy was 1084 bp and the gymnosperm-type was 1462 bp. These two copies of *nad5-1* sequences were highly divergent showing 66 substitutions, 24 indels ranging from 1 to 33 bp, and a 514 bp region that cannot be unambiguously aligned between them. The angiosperm-type *nad5-1* in *P. canariensis* consisted of a full-length *nad5* intron 1 and partial *nad5* exons 1 and 2. Using primers specific to *nad5* exons 1 and 2 of angiosperm species, we obtained an additional 197 bp sequence in exon 1, but failed to extend further into exon 2. We also used the universal primers for *nad5* exon 2 of seed plants on *P. canariensis*; these primers amplified only the gymnosperm-derived copy, not the angiosperm copy. Thus, a 1281 bp angiosperm-type *nad5-1* sequence (comprising the original 1084 bp plus the additional 197 bp from exon 1) was obtained from *P. canariensis* and used for subsequent analyses.

We performed phylogenetic analyses on the *nad5-1* sequences of *P. canariensis* and the other 175 land plant species

(Supplementary Table S2). The MP, ML and Bayesian analyses produced similar trees. The *P. canariensis* gymnosperm-type *nad5-1* was grouped with other subgenus *Pinus* species, whereas the angiosperm-type copy was clustered inside the eudicots of flowering plants (Figure 1; Supplementary Figure S2). The anomalous phylogenetic placement of the *P. canariensis* angiosperm-type *nad5-1* sequence was not caused by errors in phylogenetic reconstruction, because the topology was strongly supported by all three methods (bootstrap values=99 and 98% for MP and ML analyses, respectively, and posterior probability=1.00 for Bayesian analysis), and both the Templeton and Shimodaira–Hasegawa tests ( $P < 0.001$ ) suggested the placement of the angiosperm-type *nad5-1* in the angiosperm clade rather than in the subgenus *Pinus*. Apart from the peculiar placement of this copy of *P. canariensis*, the phylogenetic relationships throughout the rest of the tree were consistent with currently accepted seed plant phylogeny (Qiu *et al.*, 2006; APG, 2009).

The presence of the angiosperm-type copy of *nad5-1* in *P. canariensis* cannot be attributed to contamination of plant materials for the following reasons. First, genomic DNA of *P. canariensis* was



**Figure 1** The simplified strict consensus tree of land plants based on the *nad5-1* sequence (only families and genera in Pinaceae are shown). Bootstrap values for maximum parsimony (left) and maximum likelihood (right) analyses are presented above the branches, and Bayesian posterior probabilities are shown below the branches. The angiosperm- and gymnosperm-type *nad5-1* of *P. canariensis* are indicated in bold with an enlarged font. GEEC: the families Geraniaceae, Ericaceae, Euphorbiaceae and Cucurbitaceae; BCM: the families Brassicaceae, Caricaceae and Malvaceae; LPGLSA: the families Lamiaceae, Phrymaceae, Gesneriaceae, Lentibulariaceae, Solanaceae and Apocynaceae; PLANTPC: the genera *Pseudotsuga*, *Larix*, *Abies*, *Nothotsuga*, *Tsuga*, *Pseudolarix*, *Cedrus*. The details of this tree, including sampled taxa, are shown in Supplementary Figure S2.

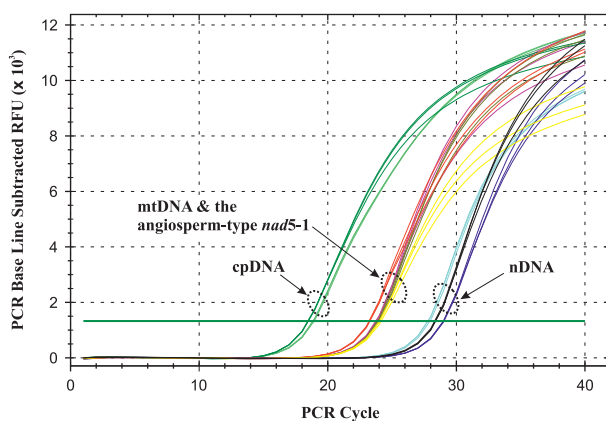
extracted independently in three different labs—Umeå University (Umeå, Sweden), Institute of Botany (Beijing, China) and Forestry and Forest Products Research Institute (Tsukuba, Japan)—from multiple individuals collected in different years, and all samples were amplified and sequenced at least twice. Second, the angiosperm-derived exons 1 and 2 were nonfunctional, owing to reading frame shifts that resulted in premature stop codons. Contamination of DNA from an extant plant would probably introduce functional copies of exons 1 and 2 rather than pseudogenes. Therefore, the most reasonable explanation for our results is that *P. canariensis* acquired the angiosperm-type *nad5-1* from an angiosperm via HGT.

For a HGT event to be successful, foreign DNA must be integrated into the recipient's genome. We adopted a qPCR strategy that is based on the differences in copy numbers among the three genomes in a plant cell. We found that genes from the same genome were amplified at similar rates. The signals of the two plastid genes (*matK* and *rbcl*) appeared first during qPCR amplification, the three mt genes (*cox1*, *matR* and the gymnosperm-type *nad5-1*) appeared next, and the three nuclear genes (*4CL*, *coo* and *cad*) appeared last (Figure 2), which is in accordance with copy number expectations for the three genomes in plants (Petit and Vendramin, 2007). The angiosperm-type *nad5-1* was amplified at a rate comparable to those of the gymnosperm-type *nad5-1* and two other mt genes (Figure 2), suggesting that the angiosperm-type copy was located in the mt genome. That the angiosperm-type *nad5-1* has been integrated into the mt genome is also supported by the low substitution rate of this region among individuals. Sequencing the angiosperm-type *nad5-1* in 22 individuals throughout the range of *P. canariensis* revealed no intraspecific variation. In contrast, genes transferred from mitochondria to nuclei often exhibit accelerated substitution rates (Palmer *et al.*, 2000). These results are consistent with previous observations that most transferred mt genes are integrated into the receptors' mt genomes (Bock, 2010). The presence of both angiosperm-derived exons and the intron in *P. canariensis* suggests that the exons and intron were transferred simultaneously as DNA. The angiosperm-derived exons 1 and 2 contained premature stop codons and thus were nonfunctional (see above), indicating pseudogenization of the foreign *nad5-1* in the host' genome after acquisition by HGT. Gene conversion between native and foreign gene

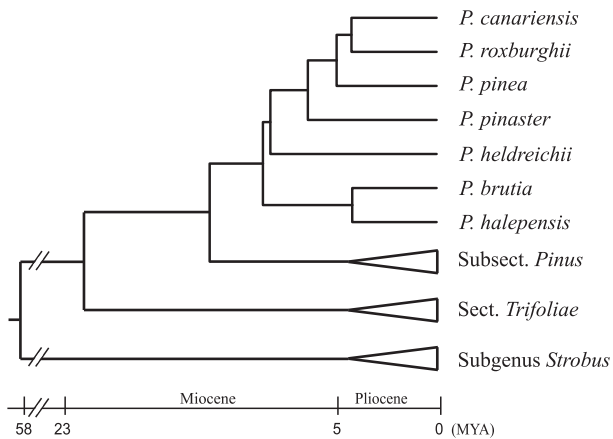
copies has been reported in previous studies of HGT (Hao *et al.*, 2010; Mower *et al.*, 2010), but this was not detected between the two *nad5-1* copies in *P. canariensis*. Future sequencing of the whole mt genome of *P. canariensis* is needed in order to clarify the accurate location of the two segments in the mt genome of *P. canariensis*, and the extent of mt HGT in this species.

Using primers specific to angiosperm-type *nad5-1*, we examined the presence of this sequence in 148 individuals of 44 *Pinus* species, including all six close relatives of *P. canariensis* in subsection *Pinaster* (Supplementary Table S2). We found that the angiosperm-type *nad5-1* was present in all 22 *P. canariensis* individuals but absent in all other 43 pine species investigated here. These results were further confirmed by the amplification of two copies of *nad5-1* in *P. canariensis* but only the gymnosperm-type copy *nad5-1* in all other pine species by the universal primers for seed plants' *nad5-1*. The presence of angiosperm-type *nad5-1* in all *P. canariensis* individuals but absence in all other pines indicated that the HGT probably occurred after *P. canariensis* diverged from its closest relatives. Thus, the divergence time between *P. canariensis* and its close sister species can be regarded as a lower bound for the HGT event. Previous studies suggest that *P. canariensis* is closely related to the Himalayan chir pine *Pinus roxburghii* (Klaus and Ehrendorfer, 1989; Parks *et al.*, 2012; Grivet *et al.*, 2013). On the basis of cpDNA dating, the divergence time between *P. canariensis* and *P. roxburghii* was 4.4 MYA (95% interval 7.3–1.6 MYA; Figure 3; Supplementary Figure S1 and Supplementary Table S4). A phylogenetic study based on the mt genome suggests that the mt genome of *P. roxburghii* was captured from an Asian hard pine (Wang and Wang, 2014). Thus, we cannot rule out the possibility that the HGT occurred in the common ancestor of *P. canariensis* and *P. roxburghii* and then the foreign copy of *nad5-1* was lost from *P. roxburghii* owing to the capture of another mtDNA. *Pinus canariensis* and *P. roxburghii* diverged from their closest relative *Pinus pinea* (Parks *et al.*, 2012; Grivet *et al.*, 2013) 5.1 MYA (8.4–2.0 MYA; Figure 3; Supplementary Figure S1 and Supplementary Table S4). This pushes the lower bound of the time of HGT slightly earlier. Taken together, the HGT event in *P. canariensis* probably occurred no earlier than the late Miocene. This result is consistent with the colonization history of *P. canariensis* in the Canary Islands (Navascues *et al.*, 2006). Island populations are typically established by only one or a few individual founders. When the founder population is small, rare alleles should theoretically have a higher probability of reaching high frequencies and eventually become fixed by drift in the newly colonized area (Klopfstein *et al.*, 2006). Severe bottlenecks promoting the fixation of rare mitotypes in newly occupied territories have been found in other pine colonization events (Wang *et al.*, 2011). In this context, the angiosperm-derived *nad5-1* could have been fixed by genetic drift during population expansion. The nonfunctional exons 1 and 2 of the angiosperm-type *nad5-1* also suggest that the fixation of this copy was most likely promoted by drift rather than because it confers any selective advantage.

There are two likely routes for the transfer of *nad5-1* into *P. canariensis*, either through direct cell–cell contact, or the foreign genetic material was bracketed and moved by an external vector, for example, fungi, bacterial pathogens or plant viruses. Many of the reported HGT events concern donor and recipient species that are engaged in parasitic or epiphytic interactions (Davis *et al.*, 2005; Mower *et al.*, 2010; Rice *et al.*, 2013; Xi *et al.*, 2013). They illustrate that physical association can facilitate HGT between species, and foreign DNA can be integrated into the recipient genome via recombination (Bock, 2010). This mechanism of HGT is supported by the observation that large pieces of DNA, and even entire



**Figure 2** Quantitative PCR assay of the angiosperm-type *nad5-1* (red), two chloroplast (cpDNA) (*matK*, light green; *rbcl*, dark green), three mitochondrial (mtDNA) (*cox1*, yellow; *matR*, pink; the gymnosperm-type *nad5-1*, golden) and three nuclear (nDNA) (*4CL*, dark blue; *coo*, light blue; *cad*, black) loci in *P. canariensis*. All reactions in the plot were run in three replicates using 6 ng genomic DNA as the template.



**Figure 3** Dated phylogeny for the genus *Pinus* based on full chloroplast genome sequences. Branch lengths are drawn to reflect BEAST divergence estimates by using 85 MYA (million years ago) as a lower bound and 45 MYA as an upper bound for the divergence time between the two subgenera *Pinus* and *Strobus*. Terminal branches within each group are collapsed for simplicity (See Supplementary Figure S1 for details), except those of subsection *Pinaster*.

organelles, can be transmitted between plant cells that come into contact with each other (Stegemann and Bock, 2009). Pines are hosts of mistletoe species (Santalales, eudicots), for example, European mistletoe (*Viscum album*) is known to parasitize pines in the Mediterranean region (Zuber and Widmer, 2009). The putative candidate angiosperm-type *nad5-1* donor in this case can be deduced from the relative position of the sequences in the phylogenetic tree. However, the low resolution of the *nad5-1* sequence tree does not allow for the identification of the candidate donor. Further studies including mistletoe species would help to test whether the angiosperm-type *nad5-1* in *P. canariensis* was from a mistletoes species via direct cell–cell contact. Vector-mediated HGT by fungi, bacterial pathogens or plant viruses has also been proposed in previous studies (Davis *et al.*, 2005; Mower *et al.*, 2010). Mycorrhizal fungi, that are essential for pines' survival and vitality, are notoriously nonspecific in their host selection and connect many distantly related plants in the same community (Simard *et al.*, 1997). The 'wood-wide web' (Simard *et al.*, 1997) could facilitate widespread exchange of DNA across phylogenetic distances spanning all green plants. Thus, it is also possible that the HGT event in *P. canariensis* was mediated through a fungal bridge between the donor and this species.

In summary, we found a new case of HGT in a gymnosperm: *P. canariensis* acquired an angiosperm-derived *nad5-1* sequence. The foreign copy of *nad5-1* has been integrated into the mt genome of *P. canariensis* and gone through pseudogenization. The HGT event was dated to the late Miocene after *P. canariensis* split from its closest relatives. Following the transfer, the foreign *nad5-1* copy was fixed by drift during the colonization of the Canary Islands by this species. The mechanism of this HGT is unclear but most likely occurred via either direct cell–cell contact or external vectors. Future investigation that includes candidate donor angiosperm species in the region would help to identify the origin of the foreign mtDNA segment in *P. canariensis* and the mechanism of the HGT. Overall, our findings provide new evidence for HGT as an additional source of mt genome variation in gymnosperms, a subject that has been little discussed thus far. Such findings together with those instances of HGT observed in angiosperms raise questions related to the evolutionary significance of HGT that need to be addressed, for example, the structural and functional

consequences of foreign gene integration, and the prevalence and characteristics of HGT in different groups of land plants. Our study provides a new natural system for exploring these questions further.

#### DATA ARCHIVING

DNA sequence: Genbank accessions KP141747–KP141754.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank Andreas D Drouzas and Grigorios Georgolopoulos for providing DNA samples of *Pinus heldreichii*. We are grateful to Fernando del Caño as curator of CIFOR's germoplasm bank. This work was supported by grants from Formas and Vetenskapsrådet, Sweden.

- APG (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* **161**: 105–121.
- Archibald JM, Richards TA (2010). Gene transfer: anything goes in plant mitochondria. *BMC Biol* **8**: 147.
- Berghorsson U, Adams KL, Thomason B, Palmer JD (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**: 197–201.
- Bock R (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci* **15**: 11–22.
- Davis CC, Anderson WR, Wurdack KJ (2005). Gene transfer from a parasitic flowering plant to a fern. *Proc R Soc B-Biol Sci* **272**: 2237–2242.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**: 1969–1973.
- Felsenstein J (1985). Confidence limits on phylogenies with a molecular clock. *Syst Zool* **34**: 152–161.
- Gernandt DS, Lopez GG, Garcia SO, Liston A (2005). Phylogeny and classification of *Pinus*. *Taxon* **54**: 29–42.
- Griivet D, Climent J, Zabal-Aguirre M, Neale DB, Vendramin GG, Gonzalez-Martinez SC (2013). Adaptive evolution of Mediterranean pines. *Mol Phylogenet Evol* **68**: 555–566.
- Hao WL, Richardson AO, Zheng YH, Palmer JD (2010). Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proc Natl Acad Sci USA* **107**: 21576–21581.
- Keeling PJ, Palmer JD (2008). Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**: 605–618.
- Kishino H, Thorne JL, Bruno WJ (2001). Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol* **18**: 352–361.
- Klaus W, Ehrendorfer F (1989). Mediterranean pines and their history. *Plant Syst Evol* **162**: 133–163.
- Klopfstein S, Currat M, Excoffier L (2006). The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**: 482–490.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuve P (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**: 2462–2463.
- Mower JP, Stefanovic S, Hao W, Gummow JS, Jain K, Ahmed D *et al.* (2010). Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol* **8**: 150.
- Navascues M, Vaxevidou Z, Gonzalez-Martinez SC, Climent J, Gil L, Emerson BC (2006). Chloroplast microsatellites reveal colonization and metapopulation dynamics in the Canary Island pine. *Mol Ecol* **15**: 2691–2698.
- Palmer JD, Adams KL, Cho YR, Parkinson CL, Qiu YL, Song KM (2000). Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA* **97**: 6960–6966.
- Parks M, Cronn R, Liston A (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* **7**: 84.
- Parks M, Cronn R, Liston A (2012). Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evol Biol* **12**: 100.
- Petit RJ, Vendramin GG (2007). Plant phylogeography based on organelle genes: an introduction. In: Weiss S, Ferrand N (eds) *Phylogeography of Southern European Refugia: Evolutionary Perspectives on the Origins and Conservation of European Biodiversity*. Springer: Amsterdam, pp 23–101.
- Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Qiu YL, Li L, Hendry TA, Li R, Taylor DW, Issa MJ *et al.* (2006). Reconstructing the basal angiosperm phylogeny: evaluating information content of mitochondrial genes. *Taxon* **55**: 837–856.
- Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J *et al.* (2013). Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**: 1468–1473.
- Richardson AO, Palmer JD (2007). Horizontal gene transfer in plants. *J Exp Bot* **58**: 1–9.

- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S *et al.* (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539–542.
- Shimodaira H, Hasegawa M (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**: 1114–1116.
- Simard SW, Perry DA, Jones MD, Myrold DD, Durall DM, Molina R (1997). Net transfer of carbon between ectomycorrhizal tree species in the field. *Nature* **388**: 579–582.
- Stamatakis A (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stegemann S, Bock R (2009). Exchange of genetic material between cells in plant tissue grafts. *Science* **324**: 649–651.
- Swofford DL. (2002). *PAUP\**. *Phylogenetic Analysis Using Parsimony (\*and other methods)*. Version 4.0b 10, Sinauer and Associates: Sunderland, MA, USA.
- Templeton AR (1983). Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**: 221–244.
- Wang B, Mao JF, Gao J, Zhao W, Wang XR (2011). Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Mol Ecol* **20**: 3796–3811.
- Wang B, Wang XR (2014). Mitochondrial DNA capture and divergence in *Pinus* provide new insights into the evolution of the genus. *Mol phylogenet Evol* **80**: 20–30.
- Wang XQ, Tank DC, Sang T (2000). Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Mol Biol Evol* **17**: 773–781.
- Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Mol Biol Evol* **24**: 90–101.
- Won H, Renner SS (2003). Horizontal gene transfer from flowering plants to *Gnetum*. *Proc Natl Acad Sci USA* **100**: 10824–10829.
- Xi ZX, Wang YG, Bradley RK, Sugumaran M, Marx CJ, Rest JS *et al.* (2013). Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genet* **9**: e1003265.
- Zuber D, Widmer A (2009). Phylogeography and host race differentiation in the European mistletoe (*Viscum album* L.). *Mol Ecol* **18**: 1946–1962.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)