Emilie Aubin [a], Christel Llauro [ab], Joseph Garrigue [c], Marie Mirouze [ad], Olivier Panaud [ae*], Moaine El Baidouri [ab*]

Characterization Of Interspecific Gene Flows at The Genome-wide Level in a Natural Ecosystem The Massane Forest Reveals New Insights Into Horizontal Transfer In plants.

a. Université de Perpignan Via Domitia, Laboratoire Génome et Développement des Plantes, UMR5096, 66860 Perpignan, France

b. Centre National de la Recherche Scientifique, Laboratoire Génome et Développement des Plantes, UMR5096, 66860 Perpignan, France

c. Réserve Naturelle Nationale de la forêt de la Massane, France

d. IRD, UMR232 DIADE, 34394 Montpellier, France

e. Institut Universitaire de France, 75231 Paris cedex 05, France

Corresponding authors :

Moaine El Baidouri (moaine.elbaidouri@univ-perp.fr)

Olivier Panaud (panaud@univ-perp.fr)

Abstract

Horizontal transfer (HT) refers to the exchange of genetic material between divergent species by mechanisms other than reproduction. In recent years, several studies have demonstrated HTs in eukaryotes, particularly in the context of parasitic relationships and in model species. However, very little is known about HT in natural ecosystems, especially those involving non-parasitic wild species, and the nature of the ecological relationships that promote these HTs. In this work, we conducted a pilot study investigating HTs by sequencing the genomes of 17 wild non-model species from a natural ecosystem, the Massane forest, located in southern France. To this end, we developed a new computational pipeline called INTERCHANGE that is able to characterize HTs at the whole genome level without prior annotation and directly in the raw sequencing reads. Using this pipeline, we identified 12 HT events, half of which occurred between lianas and trees. We found that only LTRs-retrotransposons and predominantly those from the Copia superfamily were transferred between these wild species. This study revealed a possible new route for HTs between non-

parasitic plants and provides new insights into the genomic characteristics of horizontally transferred DNA in plant genomes.

Introduction

Horizontal transfer (HT) is a process by which genetic material is exchanged between two distinct species without reproduction. HTs are well documented in prokaryotes and considered to play a major role in the adaptation and colonization of new ecological niches (Wiedenbeck and Cohan, 2011). The rapid spread of antibiotic resistance genes among bacteria is a good example of the adaptive role of HTs (Davies, 1994). Although HTs are thought to be less common in eukaryotes, numerous examples of HTs between multicellular eukaryotes such as plants, animals and insects have been reported in recent years (Keeling and Palmer, 2008; Boto, 2014; Soucy et al., 2015; Sibbald et al., 2020; Van Etten and Bhattacharya, 2020; Chen et al., 2021). Indeed, over the past decades, the number of sequenced and assembled genomes has steadily increased, facilitating the discovery of several horizontally transferred genes and transposable elements (TEs) between eukaryotes (El Baidouri et al., 2014; Peccoud et al., 2017; Dunning et al., 2019; Zhang et al., 2020). Some of these described HTs were adaptive (Christin et al., 2012; Graham et al., 2012; Yue et al., 2012; Li et al., 2014; Wang et al., 2020; Xia et al., 2021). For instance, in plants, there are several major cases of HTs leading to adaptive innovations such as the recent case of a detoxification gene transmitted horizontally from an endophytic fungus to a wild cereal (*Thinopyrum*) allowing the latter to become resistant to *Fusarium* (Wang et al., 2020). This naturally transferred gene was further introduced by breeders into wheat through wide hybridization, resulting in broad resistance to ear blight and crown rot. Another recent example is the acquisition of a detoxification gene through HT in the whitefly, a plant feeding insect, which enables it to overcome host plant defences (Xia et al., 2021).

Some HT mechanisms have been well characterized, such as in parasitic plants that are known to have experienced HT with their host plants.This is particularly true for the *Orobanchaceae* (Yoshida et al., 2010), *Striga* (Yang et al., 2016), *Cuscuta* (Yang et al., 2019) and *Rafflesia* (Davis, 2004; Xi et al., 2012). In animals, host-parasite relationships also appear to favor HT between multicellular eukaryotes, such as the HT of transposons between different mammalian species via a bloodsucking parasitic insect (Gilbert et al., 2010). Some other close biological relationships could facilitate HTs, such as natural grafting (Stegemann and Bock, 2009; Fuentes et al., 2014) or the presence of rhizomes in perennial plants (Hibdige et al., 2021). There is however also strong evidence that HTs could occur at a high rate between

species that do not share any host/parasite relationship (El Baidouri et al., 2014; Peccoud et al., 2017; Dunning et al., 2019; Zhang et al., 2020; Hibdige et al., 2021).

The vast majority of previous studies on HTs have been conducted using genomic data from public databases of model species for which sympatric relationships and biological interactions are not always known. This represents a major hurdle in attempts to understand the mechanisms and nature of biotic relationships that can promote HTs in natural ecosystems. Also, the prevalence and frequency of HTs in nature remains unknown because no previous studies have focused on characterizing eukaryotic HTs directly in a natural ecosystem. Furthermore, while previous reports have shown that both genes and TEs can be horizontally transferred between eukaryotes such as plants, it is not clear whether these two genomic components transfer at the same rate or whether certain types of genes or TEs are more frequently transferred than others. This is due to the fact that these studies have focused on a group of specific genes or TEs, mainly because the methodologies used so far require prior annotation of the sequences of interest (genes or TEs) limiting any investigation of HTs at the whole genome level.

To address these questions, we conducted a pilot study aiming to investigate HTs in a natural ecosystem, the Massane beech forest located in southern France, considered as one of the last relict forests of the Quaternary Period in Europe (Magri, 2008). We sampled 17 non-model species from this reservation, including trees, climbing plants, herbaceous species and fungi. Thanks to the *de novo* sequencing of the whole genome of these species using *Illumina* sequencing and the development of a new computational pipeline that we named INTERCHANGE (for horIzoNtal TransfER CHAracterization in Non-assembled Genome), we were able to characterize 12 cases of HTs involving 8 species in this ecosystem. These HTs are however ancient and only involve TEs, specifically LTR-retrotransposon *Copia* superfamily. Furthermore, we found that some climbing plants underwent multiple HT events with tree species which could constitute a new route of HT between non-parasitic plants.

Results

INTERCHANGE a new strategy for horizontal transfer identification at the whole-genome scale using unassembled reads is able to identify known and unknown horizontal transfer events in plant genomes

The inference of HTs is usually based on the use of three criteria (Wallau et al., 2012; Aubin et al., 2021): i) high sequence similarity between two distinct species (HS); ii) phylogenetic

incongruence between the evolutionary history of the species and that of the transferred sequence (PI); iii) patchy distribution of the transferred sequence in the phylogeny of the species (PD). Previous methods used for HT characterization based on one or two of the above criteria require the prior genes or TEs assembly and annotation. This makes the study of HTs between non-model species for which there is no reference genome or annotation available very challenging. We have therefore developed a tool to identify highly conserved regions that may arise from HT between two or several genomes of non-model species using raw short-read sequencing data. Briefly, this automatic pipeline that we called INTERCHANGE, (i) first identifies similar reads derived from conserved locus between the studied species using a K-mer approach (ii) assembles these reads into scaffolds (iii) annotate the scaffolds (iv) and test for high sequence similarity (HS) by comparing the sequence identity between conserved scaffolds with that of orthologous genes. Those HT candidates are then tested for the PI and PD criteria. The main steps of this pipeline are shown in Figure 1 and described in details in the Method section.

To validate our pipeline, we applied it to five distant plant genomes for which we previously reported HTs (El Baidouri et al., 2014): grapevine (*Vitis vinifera*), peach (*Prunus persica*), poplar (*Populus trichocarpa*), date palm (*Phoenix dactylifera*) and clementine (*Citrus clementina*) (Supplementary Table 1). These highly divergent species have experienced 6 HTs of LTR-retrotransposons named BO1, BO2, BO3, BO4, BO7 and BC1 (BO : HT between plant orders; BC : HT between plant classes) (El Baidouri et al., 2014). These transfers were taken as controls because (i) they correspond to HTs between highly divergent species (ii) they have been experimentally validated and (iii) some pairs of species have undergone multiple HTs of different LTR families. In our previous study, the identification of these HTs was done through a comparative genomic analysis using assembled and previously annotated genomes. In the present study, we retrieved unassembled raw *Illumina* genomic sequencing reads of these species and examined whether our new pipeline was able to characterize these HTs. A total of 10 whole genome comparisons were performed between the 5 species and 31 HTs candidates were identified using the INTERCHANGE pipeline, of which 30 correspond to LTR retrotransposons (29 *Copia* and 1 *Gypsy*) and one single gene (*Elongation factor 1*) (Supplementary table 2). Nine HT candidates meet both the HS, PI and PD criteria, while for the other 22, only one criterion was met leading to their rejection for further analysis. The PD criterion was tested only when the HS and PI criteria were met. Four among the six known HTs (BO1, BO3, BO4 and BO6) were identified by our new strategy (see Figure 2). BO2 and

BC1 did not pass our filter which requires that the minimum alignment length between homologous scaffolds should be greater than 1 kbp (see Figure 1-step 6 and Figure 2). Thus for BO2 (Grapevine/Clementine HT), INTERCHANGE was able to assemble 3.4 kbp of the element in grapevine but only 600 bp of its counterpart in clementine. Similarly, while our pipeline was able to assemble 1.9 kbp of the element in grapevine corresponding to the BC1 transfer (grapevine/palm), only 456 bp of this element was assembled in the palm tree. Nevertheless, this level of stringency was kept in order to limit the number of candidates to be processed in the following steps. As shown in Figure 2, we were able to recover an average of 63% of the total size of these LTR-retrotransposons compared to the elements extracted from the assembled reference genomes. We identified an additional five new HTs that were not previously identified by El Baidouri et al. (2014). This include one HT between grapevine and poplar (HT#07), two HTs between poplar and prunus (HT#20, HT#25) and two HTs between date palm and grapevine (HT#27, HT#28). Remarkably, as with the previously identified HTs, these new HTs correspond to LTRs-retrotransposons of the *Copia* superfamily. Moreover, the two HTs detected by our pipeline between date palm (monocot) and grapevine (dicot) are much more recent than BC1 (HT#30) which was the unique monocot/dicot transfer identified in our previous study. Indeed, these two retrotransposons show sequence identity at the nucleotide level of 91% and 95.5% respectively (see Supplementary Table 2), which is much higher than BC1 (86%). The reason why these elements were not previously identified is that these retrotransposons were annotated using LTRharvest program (Ellinghaus et al., 2008). It is important to note that all newly identified *Copia* elements using raw sequencing reads were also found in the reference genomes of these species. Together, these results show that the newly proposed method is an accurate approach for HT identification using unassembled raw sequencing Illumina reads.

Characterization of horizontal transfers in species from a natural ecosystem

In order to study HTs in a natural ecosystem, we chose the Massane forest, a protected reserve in southern France and a unique site in Europe designated as a UNESCO World Heritage Site in July 2021. It is mainly composed of beech trees (*Fagus sylvatica*) located between 600 and 1158 m of altitude and covering 336 hectares in total. We sampled 17 different species from this ecosystem including 14 plant species and 3 fungi (see Table 1). The plant species selected include 6 tree species, 4 lianas/climbing plants, 2 herbaceous, 1 bramble, and 1 shrub and are all non-parasitic species. The 3 fungi species were collected from tree trunks or dead wood and include 2 sparophitic and 1 parasitic species. The choice of these species was made on the

basis of several criteria, both biological and technical (i) biological interaction : some species have close biotic relationships (parasitism, close contact between liana and trees) and others have no known close physical contact (ii) genome size : to obtain a sufficient sequencing coverage for HT detection, we have chosen species with a genome size smaller than 3 Gbp (iii) species abundance : species that are common enough in the Massane forest in order to facilitate their collection (iv) taxonomy: the sample optimizes phylogenetic diversity.

The genomes of the selected species were sequenced using *Illumina* short-reads technology with 20X coverage (see Method and Supplementary Table 7). Using INTERCHANGE, we performed 136 whole-genome pairwise comparisons to identify highly similar regions between these species that may have originated from HTs. Following this analysis, we were able to identify 95 highly conserved genomic regions involving 8 out of the 17 studied species, including both genes and TEs (see Supplementary Table 3) and comprising 58 conserved TEs and 37 genes. In order to avoid redundancy of candidates due to the presence of multiple paralogs, clustering was performed using SiLiX (Miele et al., 2011), resulting in 73 different groups or families (80% identity over 80 length) (see Supplementary Table 3). All of the conserved genes that were identified came from the comparison between alder (*Alnus glutinosa*) and beech (*Fagus sylvatica*). These two species are in fact the closest relatives among all pairs of species examined, with a divergence time of 90 My (Timetree.org). Some of the conserved scaffolds between these two species were more similar than orthologous genes (HS criteria), such as a *Copia* LTR-retrotransposon (4 kbp) that has over 96% sequence identity at the nucleotide level between *Alnus* and *Fagus*, which is much higher than that of the most conserved orthologous gene between these two species. Phylogenetic analysis of this element, however, does not show strong and clear phylogenetic incongruence (PI criteria) and was therefore not considered as a candidate (see Supplementary Table 3, Cand9). In total, of the 47 non-redundant conserved scaffolds (22 genes and 25 TEs) that met the HS criteria (See Supplementary Figure 1 and 2), only 12 TEs also met the PI criteria. None of the 25 genes identified by our pipeline, despite their very high conservation (HS), show phylogenetic incongruence with species phylogeny. For the remaining 12 TEs meeting the HS and PI criteria, we further checked whether they meet the PD criterion by looking at their distribution in the phylogenetic tree of 400 plant species (See Supplementary Table 6). This analysis clearly showed that these elements have a patchy distribution, thus confirming the occurrence of HT (see Supplementary Figure 3, to 14).

***In silico* and wet-lab validation of the transferred TEs**

To rule out the possibility of contamination that could lead to the high sequence conservation between the studied species, despite all the precautions taken during sampling (see Methods), we first verified the presence of the transferred TEs in the genomes of the donor/recipient species for which a reference genome is available. These are *A. glutinosa*, *F. sylvatica*, *F. excelsior* and *P. avium*. All the elements transferred to or from these species that were identified by the INTERCHANGE pipeline were searched for homologs into their respective reference genomes by Blastn. This first analysis allowed us to find all the TEs that we identified in these genomes. In addition, we sampled and sequenced the genome of 3 additional individuals for the species involved in the HTs i.e. using long read Nanopore sequencing. These are *H. helix* (Ivy A : ~4Gb ; N50=14.4 kbp and Ivy B : ~4Gb ; N50=14.6 kbp) and *F. sylvatica* (7.4 Gb ; N50=20 kbp). We were also able to unambiguously identify the transferred TEs involving these two species in different Nanopore reads corresponding to different genomic regions (see Supplementary Figure 15 and 16). This analysis shows that the transferred TEs implicating the investigated species are not the result of contamination.

Species involved, direction and age of horizontal transfers in the Massane forest

Among the 12 HTs that we identified, none involved a saprophytic or parasitic fungi. As shown in Figure 3, these HTs occurred between 8 out of the 17 studied species. The species involved in these transfers are essentially trees and climbing plants. Indeed, 5 of the 6 analyzed tree species have experienced at least one HT event. These are, in decreasing order of HT frequency: *Fraxinus excelsior* (6 HTs), *Fagus sylvatica* (5 HTs), *Alnus glutinosa* (2 HTs), *Acer monspessulanum* (2 HTs), *Prunus avium* (1 HT). For the climbers, two species among the five analyzed have undergone HTs, namely *Dioscorea communis* (5 HTs) and *Hedera helix* (2 HTs). These HTs were identified between phylogenetically distant species that do not belong to the same plant class. In particular, the five HTs involving *D. communis* (Figure 3), which is a monocot species, occurred with dicot species that diverged over 150 million years ago. Interestingly, all HTs involving *D. communis* (4 HTs) and *H. helix* (2 HTs) occurred with tree species which may suggest that the close physical relationship between lianas and trees may be a facilitator of HTs between those plants. Additionally, we found that some species pairs underwent multiple independent HTs of different TE families such as the ones that occurred between *D. communis* and *F. excelsior* (2 HTs) and between *F. excelsior and F. sylvatica* (2 HTs). Based on the patchy distribution of the transferred LTR-retrotransposons, we could clearly identify the direction of HTs for 8 among the 12 HTs (Supplementary Table 4 and Supplementary Figure 2 to 13). While *F. excelsior* was both a

donor (1 HT) and a recipient (3 HTs) of HTs, all other species were exclusively donor or exclusively receiver, among the studied samples. For example, *F. sylvatica* was a donor of 3 HTs, *D. communi*s was a recipient of 4 HTs.

Sequence identity between the transferred TEs identified in the Massane forest varies from 89 to 97%, corresponding to an age of transfer between 3.8 and 1.15 million years (Mya) (using the molecular clock rate of Ma and Bennetzen, 2004). This indicates that these HTs occurred millions years ago. For the species for which the reference genome is available (*F. sylvatica, F. excelsior, P. avium* and *A. glutinosa*) we did find all the transferred TEs, which is an evidence for their ancient existence in the recipient genomes. This also indicates that the transfers that we identified in these wild species are ancient and probably fixed in natural populations. While we cannot determine whether these HTs occurred in the Massane forest despite its very ancient origin, it is important to note that all of the species involved in the HTs are native to European and Mediterranean regions and that their respective geographic distributions overlaps, indicating that they have been in contact for a long period of time, thus facilitating the occurrence of HTs.

Only LTRs-retrotransposons were horizontally transferred

Despite the fact that our approach of HT identification does not focus on specific types of sequences such as TEs or genes, unlike all other approaches, the HTs identified in the Massane forest involve only LTRs-retrotransposons. Of all the conserved genes that our INTERCHANGE pipeline has identified as highly conserved and potentially transferred, none meet both the HS and PI criteria. Further characterization of the protein-coding genes of these transferred LTR-retrotransposons shows that 11 among 12 belong to the *Copia* lineages (i.e. MaCo1 to 11) with the remaining one belonging to the *Gypsy* lineage (MaGy1). This result is consistent with what we observed in our previous work, where *Copia* were more frequently transferred than *Gypsy* (28 *Copia* vs 7 *Gypsy*) (El Baidouri et al., 2014). However, in order to ascertain this, it is essential to check whether this is not due to an overrepresentation of the *Copia* elements in the surveyed genomes. For this purpose, we estimated the relative frequencies of *Copia* and *Gypsy* in the 8 species involved in HTs by aligning their raw genomic reads to a collection of reference protein sequences (Neumann et al., 2019) using Diamond Blastx (Buchfink et al., 2015) (See Method). As shown in Figure 4-a, *Copia* elements were more abundant than *Gypsy* elements in 6 of the 8 species, equally abundant in *D. communis* and less abundant in *H. helix*. On average, *Copia* were 1.4 times more abundant

than *Gypsy*. This remains nevertheless insufficient to explain that 11 of the 12 identified transfers belong to the *Copia* clade.

Copia LTR-retrotransposons transferred in the Massane forest belong to Ale and Ivana lineage

We then investigated whether some *Copia* lineages have a greater propensity to transfer than others. To this end, we extracted from the Rexdb database the reverse transcriptase (RT) protein sequences of 17 *Copia* reference clades described in the literature (Neumann et al., 2019) as well as those of the 11 transferred *Copia* (Maco1 to Maco11) identified in this study (21/22 paralogs) (see Method). Furthermore, we also extracted the RT domain of the 28 *Copia* previously identified as horizontally transferred between several plant species (El Baidouri et al., 2014). The constructed phylogenetic tree shows that for the previously described HTs, 21 *Copia* (75%) belong to only two lineages Ale (13) and Ivana (8) where the 7 remaining HTs belong to different lineages such as Tork, TAR or Bianca (Figure 4-b). This trend was even more pronounced for the *Copia* elements identified in the Massane forest. In fact, all transferred elements belong only to these two lineages: Ivana (6 /11) and Ale (5/11) (Figure 4-b).

These results suggest that these two lineages are more prone to HTs compared to other *Copia* lineages in the analyzed species from the Massane ecosystem. In order to check whether this observation could be due to an overrepresentation of these two *Copia* lineages in the analyzed plant genomes, we estimated the frequency of all known *Copia* clades in the 8 plant species involved in the identified HTs (see Method). This analysis shows big disparities in *Copia* lineages frequencies in those species with no particular conserved trend. In five of the eight species involved in the transfers (*A. glutinosa, D. communis*, *F. sylvatica*, *P. avium*, *R. ulmifolius*) the Ale lineage was the predominant *Copia* lineage ranging from 8.6% in *A. monspessulanum* to 37.8% in *P. avium* (Figure 4-b). On average Ale accounted for 24% of the *Copia* elements in these genomes, followed by SIRE lineage (16.6%). Meanwhile, the percentage of the Ivana lineage varies from 3.3% in *A. monspessulanum* to 17% in *D. communis* with an average of around 6.5%. The fact that the transferred elements belong to the Ale and Ivana lineages can therefore not be explained then by the relative frequencies of these clades in their respective genomes.

The horizontally transferred *Copia* are active after their transfer but show low transpositional activity in both donor and receiver species

To better understand the dynamics of the transferred elements, we estimated their copy number in both donor and receiver species using unassembled raw genome sequencing reads. To this end, we used the coverage of single-copy BUSCO genes as a standard to normalize the observed coverage of each transferred element (see Methods). This analysis show that copy number of the transferred elements varies from single copy to 28 copies with an average of 4.3 copy per species (see Supplementary Table 4). For the HTs where we could identify the direction of HT, we can notice that some *Copia* did not transpose after their transfer (3/8) since they are single copies in the recipient species. The remaining elements show, on the contrary, a transpositional activity in the host recipient genomes that results in several copies corresponding to each transferred family. However, this post-transfer transpositional activity appears to be low, with only 2 to 5 copies observed for each family. It is interesting to note that this low transpositional activity is also observed in the donor species, suggesting that it is an inherent feature of the transferred *Copia*.

Discussion

In this study we investigated for the first time HTs between wild non-model species within a natural ecosystem. We sequenced the whole genome of 17 species including trees, climbing plants and fungi with *Illumina* technology. Because no bioinformatics tools that allow to perform direct comparison of whole genome sequencing reads from non-model species without any available reference genomes nor annotation, we have developed a new strategy for HT characterization.

INTERCHANGE a new pipeline for HT characterization at the whole-genome scale using raw sequencing reads

Our INTERCHANGE pipeline is able to perform all possible pairwise comparisons between a set of species of interest directly from raw whole genome sequencing reads. With this tool, we were able to report new HT events in previously studied species that had not been identified using previous methods (El Baidouri et al., 2014). However, as indicated by our evaluation of the control data, some of the known HT events (2/8) could not be identified or rather did not pass our filter, which requires a minimum alignment length between candidate sequences greater than 1 kbp. A decrease in stringency may allow for efficient characterization of these events, but this will be at the expense of specificity. Other parameters may also impact the specificity or sensitivity of INTERCHANGE. For example, the smaller the k-mer size, the greater the sensitivity and vice versa, but this will lead to an increase in the computational

time needed to perform all the possible pairwise comparisons. The different parameters of our tool can be modified by the user which allows great flexibility. However, we shall point out that INTERCHANGE can only detect relatively recent HTs because high sequence divergence between older transferred sequences will not satisfy the HS criteria. Despite these limitations, our results show that this pipeline is very efficient at detecting HT events at the whole genome scale using unassembled sequencing reads and is therefore a tool of choice for future studies of HT in natural ecosystems. INTERCHANGE could also be used to identify conserved sequences such as homologous genes, TEs or other types of sequences from unassembled genomes, which could be very useful for comparative genomics studies.

No plant-fungus horizontal transfer was identified at the Massane forest

Using the INTERCHANGE pipeline, we were able to identify 12 HTs implicating 8 plant species. We did not identify any transfer involving fungi even though the three studied species are saprophytic or parasitic and known to proliferate on tree trunks or dead wood. It is broadly accepted that close relationships such as endosymbiosis or parasitism are favorable for HTs in eukaryotes (Gilbert et al., 2010; Xi et al., 2013; Kim et al., 2014; Leclercq et al., 2016; Yang et al., 2016, 2019). The absence of plant-fungus HT in this study may suggest that such events are rare or too old to be detected (Richards et al., 2009).

Liana-tree interactions: a possible new route of horizontal transfer between non-parasitic plants?

Our results also show that the two climbing plants, the common ivy (*H. helix*) and black bryony (*D. communis*) have experienced several HTs events predominantly with trees. These findings are in agreement with our previous study that showed a higher frequency of HTs between grapevines and several tree species (El Baidouri et al., 2014). To date, no hypothesis has been put forward to explain this higher frequency of transfer in grapevines and whether it is due to an inherent genetic trait or to its particular ecological lifestyle. In fact, similarly to common ivy and black bryony, wild grapevine use trees as support for growth which could explain the high HT frequency observed in this species. A recent study on four different and closely related *Vitis* species seems to confirm this trend (Park et al., 2021). Using comparative genomics, the authors identified dozens of HTs between these four closely related vine species and mainly trees belonging to highly divergent taxa, although they did not highlight the greater frequency of HTs between grapevine and trees. In light of our findings, we hypothesize that liana-tree interactions may favor HTs between non-parasitic plants and could

be considered as a direct route by which HTs occur frequently in nature. For the three other climbing plant genomes analyzed in this study, we did not find any HT. Therefore, the question of whether some vine species are more prone to HTs than others remains unanswered at this point. This needs to be tested on a larger sampling of liana species. It is also possible that the tree-to-tree HTs that we identified between beech (*F. sylvatica*), ash (*F. excelsior*) or alder (*A. glutinosa*) could be mediated by other, yet not sequenced, climbing plant species.

Horizontal transfers in plants mainly involve low copy number LTR-retrotransposons belonging to the Ivana and Ale lineages of the *Copia* superfamily

Our study reveals that LTR-retrotransposons are the only genetic elements that experienced HTs in the studied plant species, which confirms earlier reports, but remains without mechanistic explanation. In fact, successful HT requires three key steps, namely the "excision" of genetic material in the form of DNA or RNA molecules from the donor genome, its transport to the recipient species and finally its integration into the target genome. Due to their transposition life style, LTR-retrotransposons are able to generate extrachromosomal double stranded DNA encapsidated in the VLP (Virus Like Particule) and accumulating in the cytoplasm of the cells (Lanciano et al., 2017; Lee et al., 2020). They also have the ability to integrate into the host genome using the integrase (IN) (Sabot and Schulman, 2006; Lee and Martienssen, 2021). LTR-retrotransposons may therefore be more likely to achieve successful HT, given their ability to generate double stranded DNA encapsidated in the VLP and because of their ability to integrate the host genome. Although both *Copia* and *Gypsy* elements can produce VLPs, *Copia* appears to be more prone to horizontal transfer than *Gypsy* elements. If *Copia* and *Gypsy* superfamilies differ mainly in the order of the IN and RT domains, there are some genomic and transpositional features specific to each of these superfamilies. For instance, *Copia* elements are abundant in gene-rich euchromatic regions while copia elements are mainly located in Heterochromatic and pericentromeric regions (Baduel et al., 2021). *Copia* are also generally activated in response to environmental stress as it has been shown for many plant species (Grandbastien, 2015). It is therefore possible that Copia elements, because of their presence in transcriptionally active regions of the genome and because of their responsiveness to environmental stresses could facilitate their HT.

Intriguingly, *Copia* elements that have been horizontally transferred between plant genomes belong mainly to the Ivana and Ale lineage and are low copy numbers. The reasons why Ale and Ivana clade appear to be more prone to HT compared to the other *Copia* lineages remain

unknown, as there are no well-known common specific features of these two clades that clearly differentiate them from others (Wicker and Keller, 2007). A recent population genomics study in *Arabidopsis arenosa* showed that *Copia* elements, particularly Ale and Ivana respond to temperature and irradiance (Wos et al., 2021). It is also interesting to note that the TEs shown to be currently active in *A. thaliana* namely *EVADE* (Mirouze et al., 2009) and *ONSEN* (Ito et al., 2011), also belong to the Ale and Ivana clade, respectively, and the latter is active in response to heat stress (Ito et al., 2011). As for the transferred *Copia*, these two families also have a low copy number: two copies for *EVADE* and eight copies for *ONSEN*. When a TE family reaches high copy numbers, it tends to be silenced by the production of small interfering RNAs and the epigenetic machinery depositing DNA methylation (Marí-Ordóñez et al., 2013). The silenced TEs would not be candidate for HT. If this holds true, the question of the presence and survival of Gypsy families in eukaryotic genomes remains to be explained by other mechanisms. Considering our study and previous ones on HT in plants, the propensity of *Copia* elements and in particular Ale and Ivana lineages to transfer horizontally can not be explained. Further studies are needed to elucidate the reasons for the remarkable ability of low copy number *Copia* to transfer horizontally in plants.

Conclusions

In this work, we conducted a pilot study on HTs *in natura* in a forest ecosystem. For this purpose, we implemented a new comparative genomics tool able to identify HTs at the whole genome level directly from raw sequencing reads. We characterized 12 HTs that all correspond to *Copia* LTRs retrotransposons and particularly those belonging to the Ale and Ivana lineages. Our study also shows that some lianas species have experienced recurrent horizontal transfers with trees that constitute their growth support in nature. This work sheds light on a new route of HTs between non-parasitic plant species and the type of genetic elements most likely to be transferred and integrated.

Materiel and methods

Sampling

The 18 species analyzed were sampled in the Massane Forest National Nature Reserve. After sampling the target tissues (leaf or sporophore), the samples were first washed with a detergent solution (Tween 80 at 0.1%) and then rinsed twice successively in a miliQ water

solution. The samples were then dried with absorbent paper and stored in liquid nitrogen and then at -80°C once in the laboratory before DNA extraction.

Illumina genome sequencing

The 17 species analyzed were sampled in the Massane Forest. In order to prevent contamination, the tissues (leaf for plants or sporophore for fungi) were washedwith a detergent solution (Tween 80 at 0.1%) and rinsed twice with miliQ water. The samples were then dried with paper towels and kept on ice for subsequent storage at -80°C. DNA from each sample was extracted using the CTAB2X (Debladis et al., 2017) and the quality of the DNA was estimated by Nanodrop (Thermo Scientific) and Qubit (Invitrogen) quantification. DNA libraries and sequencing was outsourced to Novogene company using the a Hiseq 2000 and Novaseq 6000 platforms. Briefly, a total amount of 1µg DNA per sample was used as input material for the DNA libraries. Sequencing libraries were generated using NEBNext® DNA Library Prep Kit following manufacturer's recommendations and indices were added to each sample. The genomic DNA was randomly fragmented to a size of 350bp by shearing, then DNA fragments were end polished, A-tailed, and ligated with the NEBNext adapter for *Illumina* sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and resulted libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. Paired-end sequencing was performed using a coverage of 20X and a read length of 150bp for each sample.

Nanopore genome sequencing

High-molecular-weight genomic DNA was extracted from 0.41 g to 0.5 g of frozen leaf tissue according to Pushkova et al., 2019 and the Oxford Nanopore Technologies protocol (February 2019). Briefly, after lysis of cell membranes with Carlson buffer, proteins were removed with chloroform. DNA was purified using Qiagen Genomic-tip 100 columns following the manufacturer's instructions. A selection of fragments > 10 Kb was performed using AMPure XP beads. DNA quantification was performed by Nanodrop (Thermo Scientific) and Qubit assays (Invitrogen) and the quality was assessed on a 0.8% agarose gel. We then followed the 1D genomic DNA protocol by ligation with the SQK-LSK109 kit to prepare the 3 libraries using 3 µg, 3.9 µg, and 4.1 µg of DNA (beech, ivy A, and ivy B), respectively. We successively loaded 1.7 µg of library onto a Flowcell R9, 2.6 µg and 2.7 µg of libraries onto two Flowcell R10. We produced 7.4 Gb, and 2 times 4 Gb of fastq pass reads with N50s of 20

kb, 14.4 kb and 14.6 kb, respectively. Bascalling was performed using guppy in the high accuracy (hac) mode (https://nanoporetech.com/nanopore-sequencing-data-analysis).

Automatic detection of Horizontal transfer using INTERCHANGE pipeline

High similarity criteria (Step1 to 8 using INTERCHANGE automatic pipeline)

Step 1 - Identification of homologous reads derived from conserved regions using a k-mers approach: *k*-mers indexes (k = 30) were generated using Tallymer mkindex option (Kurtz et al., 2008). with default parameters except for: -mersize 30; minocc 1. The search for identical *k*-mers between each species pair was performed using Tallymer search option with the following parameters: -output qseqnum qpos counts sequence.

Step 2 - Once identical k-mer have been identified between reads of two species, the overlapping k-mer are merged and the total similarity score is calculated for each pair of reads using the following formula: Read similarity = total length of identical non-overlapping k-mer / reads length. Reads with a similarity score greater than 50% are considered to originate from conserved homologous regions and are therefore kept for further analysis.

Step 3 - There are a significant number of identical k-mers that correspond to regions of simple repeats such as tandem repeats. Reads containing such repeats are removed using Prinseq-lite tool (Schmieder and Edwards, 2011) with the following parameters: out_format 1; -lc-method dust; -lc-thresholds 10.

Step 4 - The homologous reads that pass the similarity filter are then extracted and assembled separately for each species using SPAdes (Prjibelski et al., 2020) with the paird-end and only_assemble options. This step will result in the assembly in each species of scaffolds corresponding to highly conserved regions potentially derived from HTs.

Step 5 - The assembled scaffolds are then aligned using both Diamond blastp (Buchfink et al., 2015) and BLASTn against several databases with a minimum e-value de 1e-5 et 1e-20 respectively: CDDdelta, Repbase, mitochondrial, chloroplast, and ribosomal (TIGR) gene database (Ouyang, 2004).First, sequences that align to mitochondrial, chloroplastic and ribosomal genes are excluded. Indeed, these genes are generally highly conserved between distant species and therefore often meet the criterion of high similarity. When a scaffold aligns to several target sequences from multiple databases, only target sequences with the highest alignment score are considered as being homologous. At the end of this step, each

scaffold will be classified into one of these categories: genes, TEs, MCRs (mitochondrial, chloroplast or ribosomal genes)

Step 6 - Identification of homologous scaffolds: the objective of this step is to identify homologous scaffolds between each pair of compared species. For this purpose, a reciprocal Blastn is performed and homologous scaffolds are identified using the reciprocal best hit method (RBH). Only scaffolds with an alignment length greater than 1Kbp and with a minimum of 80% identity are retained for further analysis.

Step 7 - In order to distinguish, among the set of conserved scaffolds identified by INTERCHANGE, those that could originate from HT, it is necessary to first test the criterion of high similarity (HS). This means that the similarity of the transferred sequences between the donor and recipient species must be significantly higher than that of orthologous genes. Before assessing this criterion, it is therefore important to identify and assemble the conserved orthologous genes in the investigated species from unassembled short reads.

Characterization of orthologous BUSCO genes from unassembled reads: (i) Since the studied species from the Massane forest did not have any available gene annotation, we have assembled and annotated their BUSCO genes. These genes were used to test the HS criteria and to build the species phylogenetic tree. As a first step, the BUSCO genes of 400 publicly available assembled plant genomes (see Species list in Supplementary Table 6) were identified, resulting in a genomic database of ~169,000 BUSCO genes covering angiosperms, gymnosperms and basal plant species (this database has been deposited on the following link http://gamay.univ-perp.fr/~moaine/Database/). The genomic reads of each sequenced species from the Massane forest were mapped against this BUSCO database by minimap2 (Li, 2018). using default parameters. The mapped reads were extracted, merged and assembled by SPAdes (Prjibelski et al., 2020) using paired-end and -only_assembler options. The resulting scaffolds were then realigned by Blastn against the nucleic BUSCO database and assigned to their corresponding BUSCO genes.

Step 8 - Identification of high sequence similarity threshold based on the distribution of orthologous gene identities: In order to identify whether conserved scaffolds have higher sequence similarity compared to orthologous BUSCO genes, a high similarity threshold (HS) is determined based on the distribution of orthologous gene sequence identities according to the following formula: HS = Q3+(IQR/2); where Q3 is the third quartile, IQR is the interquartile range (Q3-Q1).

Step 9 - Phylogenetic incongruence criteria (PI)

Building the phylogenetic tree of the studied species: the phylogenetic tree of the studied species is built based on BUSCO genes previously identified in step 7. Multiple alignment of orthologous BUSCO genes of the studied species and the 400 plant genomes is performed using Mafft program (Katoh and Standley, 2013). The alignments are then cleaned with TrimAl (Capella-Gutierrez et al., 2009) and the trees constructed with FastTree (Price et al., 2009). A consensus tree is then obtained using Astral (Zhang et al., 2018) from the previously constructed trees.

Building the phylogenetic tree of the transferred sequence: To construct the phylogenetic tree of the transferred elements, we aligned each of these elements to the assembled genomes of 400 plant species using Blastn. Sequences with sequence identity greater than 80% and covering at least 60% of the element were considered homologous. We performed multiple alignments for each element and all its homologs using the Mafft program (Katoh and Standley, 2013). These alignments were then cleaned using trimAL (Capella-Gutierrez et al., 2009) with the following parameters: -cons 30; -gt 0.5. Finally, phylogenetic trees for each transferred element were inferred with FastTree (Price et al., 2009). The resulting trees were then manually compared to the species trees to check for the presence or absence of phylogenetic incongruencies. The trees were visualized using the Iroki Phylogenetic Tree Viewer (Moore et al., 2020).

The PI criterion is met when the phylogenetic tree of the HT candidate shows that the donor and recipient species are sister clades, unlike the species tree.

Step 10- Testing the Patchy distribution (PD) : Finally, to consider that there is an unequal distribution of this sequence in the tree of species, the candidate sequence must be found in species close to donor/recipient but missing in species closely related to partner implicated in the HT. Alternatively, the transferred sequence could also be found only in the two species involved in the transfer.

Candidates meeting the HS, the PI and the PD criteria are therefore considered as resulting from HTs.

Estimation of copy LTRs number in unassembled genomes

To estimate the copy number of each retrotransposon in the species involved in the transfer, we calculated the number of mapped reads on each transferred retrotransposon compared to

the numbers of mapped reads on single-copy genes. Total reads for each species were mapped onto the transferred LTRs-retrotransposons. For each LTR-retrotransposon, we calculated the coverage at each nucleotide of the element. The median coverage was taken as a proxy for the coverage of the element in the genome. The same strategy was adopted to estimate the coverage of the BUSCO genes of the studied species. We then used the following formula to estimate the total copy number of each transferred LTRs using *Illumina* reads: Copy number = (MCT / MCB), where MCT is the median coverage of LTRs and MCB: median coverage of BUSCO genes. To test whether this approach is an appropriate method to estimate copy number using genomic raw reads, we compared the copy number estimated from unassembled genomes with that obtained from assembled reference genomes in species for which the latter is available (Supplementary table 4). Copy numbers estimated from unassembled genomes and those obtained by Blastn against reference genomes are highly correlated, validating our approach (Pearson correlation; R = 0.982, p-value = 4.699E-10).

Phylogenetic tree of copia lineages

We extract the RT (reverse transcriptase) domain of the transferred *Copia* elements in both donor and receiver species (22 paralogs corresponding to the 11 *Copia* families). For 60% of the paralogs (13/22), the RT domain was assembled using our automatic INTERCHANGE pipeline. For the others, the RT domain was lacking. We then manually reassemble the lacking RT domains using raw *Illumina* reads of the corresponding species. For species for which the reference genome is available (*F. sylvatica*, *F. exclesior*, *P. avium* and *A. glutinosa*), we realigned the raw reads to the reference genomes and used the reference elements as a guide for manual assembly. Alternatively, we used homologs from other closely related plant species to guide manual assembly. Using this strategy, we obtained for most *Copia* paralogs involved in HTs nearly the complete elements with the corresponding RT domain (21/22).

Frequency estimates of the different *Copia* and *Gypsy* lineages

To estimate the relative frequency of *Copia* and *Gypsy* in the sequenced genomes, we aligned the raw genome reads of each species to a collection of protein sequences corresponding to the different known *Copia* and *Gypsy* lineages from the RexDB (Neumann et al., 2019) database by Diamond Blastx (evalue 1e-5) (Buchfink et al., 2015). The number of aligned reads on each superfamily and on each lineage was reported to the total number of aligned reads to estimate their relative frequency.

Data Availability

The data have been deposited in NCBI under BioProject accession number PRJNA788424 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA788424). INTERCHANGE is open source and available at https://github.com/emaubin/INTERCHANGE

Funding

M.E is supported by a grant from the Agence Nationale de la Recherche (ANR-21-CE02-0031-01). O.P is supported by a BQR grant of the University of Perpignan, the Institut Universitaire de France and Occitanie region (Biodivoc). E.A is supported by a PhD grant from Occitanie region. M.M is supported by a grant from the Agence Nationale de la Recherche (ANR-21-PRCI-CE02). This study is set within the framework of the "Laboratoire d'Excellence (LABEX)" TULIP (ANR-10-LABX-41) and of the "Ecole Universitaire de Recherche (EUR)" TULIP-GS (ANR-18-EURE-0019).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

M.E and O.P conceived and designed the study. J.G, E.A, M.E and O.P collected the samples. E.A and M.E performed the analysis. E.A and C.L extracted DNA for *Illumina* sequencing. C.L and M.M performed Nanopore sequencing. M.E wrote the manuscript with input from E.A. O.P and M.M provided critical feedback on the manuscript.

Acknowledgements

We are grateful to Marie-Chrsitine Carpentier, Joris Bertrand for their help for plant sampling at the Massane forest.

Figure legend and table

Common Name	Species	Type	Estimaed Genome size	Available reference genome
Beech	*Fagus sylvatica*	Tree	540 Mbp	yes
Ash	*Fraxinus excelsior*	Tree	840 Mbp	yes
Montpellier maple	*Acer monspessulanum*	Tree	730 Mbp	no
Wild cherry	*Prunus avium*	Tree	430 Mbp	yes
Alder	*Alnus glutinosa*	Tree	500 Mbp	yes
Whitebeam	*Sorbus aria*	Tree	1.03 Gbp	no
Hairy Greenweed	*Genista pilosa*	Shrub	1.04 Gbp	no
Honeysuckle	*Lonicera periclymenum*	Liana	2.81 Gbp	no
Ivy	*Hedera helix*	Liana	1.5 Gbp	no
Black bryony	*Dioscorea communis*	Liana	830 Mbp	no
Giant blackberry	*Rubus ulmifolius*	Bramble	450 Mbp	no
White Bryony	*Bryonia dioica*	Liana	1.6 Gbp	no
Sage	*Salvia sp*	Herbaceous	760 Mbp	-
Tinder Bracket	*Fomes fomentarius*	Fungi	50 Mbp	yes
Coral Tooth	*Hericium clathroides*	Fungi	40 Mbp	no
Oyster mushroom	*Pleurotus ostreatus*	Fungi	20/50Mbp	yes
Narrow-leaved Ragwort	*Senecio inaequidens*	Herbaceous	580 Mbp	no

Table 1 : Species sampled in the Massane forest and whose genome has been sequenced by *Illumina* short read sequencing.

Figure 1 : The different steps of the INTERCHANGE pipeline of horizontal transfer identification from unassembled and unannotated genomes. Steps 1 to 8 are completely automatic steps 9 and 10 are semi-automatic. Step1 : Identification of identical k-mers using Tallymer (Kurtz et al., 2008). Steps 2 & 3 Identification of reads derived from conserved regions & Removal of Tandem repeats using PRINSEQ lite (Schmieder and Edwards, 2011) . Reads sharing at least 50% of identical k-mers are considered as homologous reads. Step 4 :

homologous reads are extracted and assembled for each pair of species using SPAdes (Prjibelski et al., 2020). Step 5 : Scaffolds annotation using multiple protein and TEs database : CDDdelta, Repbase, mitochondrial, chloroplast, and ribosomal (TIGR) gene database. Step 6 : Identification of homologous scaffolds using reciprocal best hit (RBH). Step 7 : Identification of high sequence similarity threshold based on the distribution of orthologous BUSCO gene identities according to the following formula : high similarity threshold (HS) = Q3+(IQR/2); where Q3 is the third quartile, IQR is the interquartile range (Q3-Q1). Step 8 : Testing for HS criteria. Step 9 : Phylogenetic incongruence criteria. Step 10 : testing the Patchy distribution (PD) of transferred sequence. For details see Method section.

Figure 2 : HTs identified in 5 plant genomes known to have experienced several HTs. In green, HTs that were identified in our previous study (El Baidouri et al., 2014) and also detected using the INTERCHANGE pipeline based on unassembled sequencing reads. In gray, HT controls that did not pass all filters. In red, HTs only identified using the new INTERCHANGE pipeline. Y axis : sequence length comparaison of the HT candidates obtained using INTERCHANGE (Blue) and extracted from the reference genome of the corresponding species (orange). Kbp : kilo base pair.

Figure 3 : The phylogenetic tree of the 17 analyzed Massane species. The curves represent the identified HTs and link the involved species. Blue and red curves represent *Gypsy* and *Copia* HTs, respectively. The asterisks indicate multiple HTs. The horizontal scale represents the divergence time in million years (source: timetree.org). Correspondence of species names: Ace: *Acer monspessulanum*, Aln: *Alnus glutinosa*, Bry: *Bryonia dioica*, Dio: *Dioscorea communis*, Fag: *Fagus sylvatica*, Fra: *Fraxinus excelsior*, Fom: *Fomes fomentarius*, Gen: *Genista pilosa*, Hed: *Hedera helix*, Her: *Hericium clathroides*, Lon: *Lonicera periclymenum*, Ple: *Pleurotus ostreatus*, Pru: *Prunus avium*, Rub: *Rubus ulmifolius*, Sal: *Salvia sp*, Sen: *Senecio inaequendis*, Sor: *Sorbus aria*.

Figure 4 : The relative abundance of LTRs-retrotransposon superfamilies in species that have experienced HTs. a) Relative frequency of Copia and Gypsy in the studied species involved in HTs. In blue : *Copia* frequency, in yellow : *Gypsy* frequency b) Phylogenetic tree of transferred *Copia* detected in this study using the RT domain. In bold, the consensus sequence of the reference *Copia* lineages. Maco1 to 11 correspond to horizontally transferred elements identified between the plant species from the Massane. BO1 to BO8, BG1 to BG and BC1 correspond to *Copia* elements identified in our previous study. Correspondence of species names as in Fig. 3. c) Copia lineages relative frequencies in species involved in HT.

Supplementary Figure 1 : Sequence identity distribution of the assembled BUSCO genes in the studied species. These sequence identities were obtained by Blastn alignment. N: corresponds to the total number of BUSCO genes that can be aligned at the nucleic level between each pair of species. n: the number of genes above our HS high similarity threshold represented by a vertical dashed bar. HS = (Q3+IQR/2), the inter-quartile range IQR = Q3-Q1 (Q1 and Q3 correspond to the first and third quartile respectively). The age in millions of years (Mya) represents the divergence time between species according to Timetree.org. Correspondence of species names: Ace: *Acer monspessulanum*, Aln: *Alnus glutinosa*, Bry: *Bryonia dioica*, Dio: *Dioscorea communis*, Fag: *Fagus sylvatica*, Fra: *Fraxinus excelsior*, Fom: *Fomes fomentarius*, Gen: *Genista pilosa*, Hed: *Hedera helix*, Her: *Hericium clathroides*, Lon: *Lonicera periclymenum*, Ple: *Pleurotus ostreatus*, Pru: Prunus avium, Rub: *Rubus ulmifolius*, Sal: *Salvia sp*, Sen: *Senecio inaequendis*, Sor: *Sorbus aria*.

Supplementary Figure 2 : Example of horizontal transfer of a Copia between Ash and Alder (MaCo01) illustrated by high sequence similarity (HS) and phylogenetic incongruence (PI). a) Graph representing the alignment of the two homologs of the transferred element between the two species involved in the transfer. b) Histogram of the distribution of sequence identity of the orthologous genes (BUSCO) conserved between the two species. The N indicate the number of BUSCO genes analyzed. The red arrow corresponds to the sequence identity of the transferred Copia between the two species. c) Phylogenetic tree of the transferred Copia homologs identified in 400 plant genomes. d) Phylogenetic tree of the species for which at least one Copia homologous sequence has been identified. The colors of the branches correspond to the different plant families indicated in the box.

Supplementary Figure 3-14 : Patchy distribution of the horizontally transferred LTR-retrotransposons in the phylogenetic tree of 400 plant species. Maco1 to Maco11 (Figure 2-12) and MaGy01 (Figure 13).

Supplementary Figure 15 : Graphical visualisation of Blastn alignment of transferred LTRs-retrotransposons (Maco2) identified and assembled using INTERCHANGE pipeline against Nanopore reads of two *Hedera helix* genomes corresponding to two ivy individuals A and B. Visual representation was achieved using http://kablammo.wasmuthlab.org/ software.

Supplementary Figure 16 : Graphical visualisation of Blastn alignment of two transferred LTRs-retrotransposons (Maco3 and Maco11) identified and assembled using

INTERCHANGE pipeline against Nanopore reads of *Fagus sylvatica* genome. Visual representation was achieved using http://kablammo.wasmuthlab.org/ software.

References

Aubin, E., El Baidouri, M., and Panaud, O. (2021). Horizontal Gene Transfers in Plants. *Life* 11, 857. doi:10.3390/life11080857.

Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Loudet, O., et al. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. *Genome Biology* 22, 138. doi:10.1186/s13059-021-02348-5.

El Baidouri, M. El, Carpentier, M. C., Cooke, R., Gao, D., Lasserre, E., Llauro, C., et al. (2014). Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Research* 24, 831–838. doi:10.1101/gr.164400.113.

Boto, L. (2014). Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proceedings of the Royal Society B: Biological Sciences* 281, 20132450. doi:10.1098/rspb.2013.2450.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12, 59–60. doi:10.1038/nmeth.3176.

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi:10.1093/bioinformatics/btp348.

Chen, R., Huangfu, L., Lu, Y., Fang, H., Xu, Y., Li, P., et al. (2021). Adaptive innovation of green plants by horizontal gene transfer. *Biotechnology Advances* 46, 107671. doi:10.1016/j.biotechadv.2020.107671.

Christin, P. A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., et al. (2012). Adaptive evolution of C 4 photosynthesis through recurrent lateral gene transfer. *Current Biology* 22, 445–449. doi:10.1016/j.cub.2012.01.054.

Davies, J. (1994). Inactivation of antibiotics and the dissemination of resistance genes. *Science* 264, 375–382. doi:10.1126/science.8153624.

Davis, C. C. (2004). Host-to-Parasite Gene Transfer in Flowering Plants: Phylogenetic Evidence from Malpighiales. *Science* 305, 676–678. doi:10.1126/science.1100671.

Debladis, E., Llauro, C., Carpentier, M.-C., Mirouze, M., and Panaud, O. (2017). Detection of active transposable elements in Arabidopsis thaliana using Oxford Nanopore Sequencing technology. *BMC Genomics* 18, 537. doi:10.1186/s12864-017-3753-z.

Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., et al. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences*, 201810031. doi:10.1073/pnas.1810031116.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9. doi:10.1186/1471-2105-9-18.

Fuentes, I., Stegemann, S., Golczyk, H., Karcher, D., and Bock, R. (2014). Horizontal genome transfer as an asexual path to the formation of new species. *Nature* 511, 232–235. doi:10.1038/nature13291.

Gilbert, C., Schaack, S., Pace II, J. K., Brindley, P. J., and Feschotte, C. (2010). A role for host–parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464, 1347–1350. doi:10.1038/nature08939.

Graham, L. A., Li, J., Davidson, W. S., and Davies, P. L. (2012). Smelt was the likely beneficiary of an antifreeze gene laterally transferred between fishes. *BMC Evolutionary Biology* 12, 190. doi:10.1186/1471-2148-12-190.

Grandbastien, M.-A. (2015). LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1849, 403–416. doi:10.1016/j.bbagrm.2014.07.017.

Hibdige, S. G. S., Raimondeau, P., Christin, P.-A., and Dunning, L. T. (2021). Widespread lateral gene transfer among grasses. *New Phytologist* 230, 2474–2486. doi:10.1111/nph.17328.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472, 115–119. doi:10.1038/nature09861.

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772–780. doi:10.1093/molbev/mst010.

Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* 9, 605–618. doi:10.1038/nrg2386.

Kim, G., LeBlanc, M. L., Wafula, E. K., DePamphilis, C. W., and Westwood, J. H. (2014). Genomic-scale exchange of mRNA between a parasitic plant and its hosts. *Science* 345, 808–811. doi:10.1126/science.1253122.

Kurtz, S., Narechania, A., Stein, J. C., and Ware, D. (2008). A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. *BMC genomics* 9, 517. doi:10.1186/1471-2164-9-517.

Lanciano, S., Carpentier, M. C., Llauro, C., Jobet, E., Robakowska-Hyzorek, D., Lasserre, E., et al. (2017). Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genetics* 13, 1–20. doi:10.1371/journal.pgen.1006630.

Leclercq, S., Thézé, J., Chebbi, M. A., Giraud, I., Moumen, B., Ernenwein, L., et al. (2016). Birth of a W sex chromosome by horizontal transfer of Wolbachia bacterial symbiont genome. *Proceedings of the National Academy of Sciences of the United States of America* 113, 15036–15041. doi:10.1073/pnas.1608979113.

Lee, S. C., Ernst, E., Berube, B., Borges, F., Parent, J.-S., Ledon, P., et al. (2020). Arabidopsis LTR retrotransposons and their regulation by epigenetically activated small RNA. 576–588. doi:10.1101/2020.01.24.919167.

Lee, S. C., and Martienssen, R. A. (2021). Regulation of retrotransposition in Arabidopsis. *Biochemical Society Transactions* 49, 2241–2251. doi:10.1042/BST20210337.

Li, F. W., Villarreal, J. C., Kelly, S., Rothfels, C. J., Melkonian, M., Frangedakis, E., et al. (2014). Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proceedings of the National Academy of Sciences of the United States of America* 111, 6672–6677. doi:10.1073/pnas.1319929111.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191.

Magri, D. (2008). Patterns of post-glacial spread and the extent of glacial refugia of European beech (Fagus sylvatica). *Journal of Biogeography* 35, 450–463. doi:10.1111/j.1365-2699.2007.01803.x.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genetics* 45, 1029–

1039. doi:10.1038/ng.2703.

Miele, V., Penel, S., and Duret, L. (2011). Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* 12, 116. doi:10.1186/1471-2105-12-116.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., et al. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* 461, 427–430. doi:10.1038/nature08328.

Moore, R. M., Harrison, A. O., McAllister, S. M., Polson, S. W., and Wommack, K. E. (2020). Iroki: automatic customization and visualization of phylogenetic trees. *PeerJ* 8, e8584. doi:10.7717/peerj.8584.

Neumann, P., Novák, P., Hoštáková, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA* 10, 1. doi:10.1186/s13100-018-0144-1.

Ouyang, S. (2004). The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research* 32, 360D – 363. doi:10.1093/nar/gkh099.

Park, M., Sarkhosh, A., Tsolova, V., and El-Sharkawy, I. (2021). Horizontal Transfer of LTR Retrotransposons Contributes to the Genome Diversity of Vitis. *International Journal of Molecular Sciences* 22, 10446. doi:10.3390/ijms221910446.

Peccoud, J., Loiseau, V., Cordaux, R., and Gilbert, C. (2017). Massive horizontal transfer of transposable elements in insects. *Proceedings of the National Academy of Sciences* 114, 4721–4726. doi:10.1073/pnas.1621178114.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* 26, 1641–1650. doi:10.1093/molbev/msp077.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* 70. doi:10.1002/cpbi.102.

Pushkova, E. ., Beniaminov, A. ., Krasnov, G. ., Novakovskiy, R. ., Povkhova, N. ., Melnikova, N. ., et al. (2019). Extraction of high-molecular-weight DNA from poplar plants for Nanopore sequencing. in *Plant genetic resources for breeding and producing functional nutraceutical food* (Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences Novosibirsk State University), 158–160. doi:10.18699/ICG-PlantGen2019-51.

Richards, T. A., Soanes, D. M., Foster, P. G., Leonard, G., Thornton, C. R., and Talbot, N. J. (2009). Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21, 1897–1911. doi:10.1105/tpc.109.065805.

Sabot, F., and Schulman, A. H. (2006). Parasitism and the retrotransposon life cycle in plants : a hitchhiker ' s guide to the genome. 381–388. doi:10.1038/sj.hdy.6800903.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi:10.1093/bioinformatics/btr026.

Sibbald, S. J., Eme, L., Archibald, J. M., and Roger, A. J. (2020). Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. *Trends in Parasitology* 36, 927–941. doi:10.1016/j.pt.2020.07.014.

Soucy, S. M., Huang, J., and Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics* 16, 472–482. doi:10.1038/nrg3962.

Stegemann, S., and Bock, R. (2009). Exchange of Genetic Material. *Science* 649. doi:10.1126/science.1170397.

Van Etten, J., and Bhattacharya, D. (2020). Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends in Genetics*. doi:10.1016/j.tig.2020.08.006.

Wallau, G. L., Ortiz, M. F., and Loreto, E. L. S. (2012). Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. *Genome Biology and Evolution* 4, 689–699. doi:10.1093/gbe/evs055.

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., et al. (2020). Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. *Science*, eaba5435. doi:10.1126/science.aba5435.

Wicker, T., and Keller, B. (2007). Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Research* 17, 1072–1081. doi:10.1101/gr.6214107.

Wiedenbeck, J., and Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* 35, 957–976. doi:10.1111/j.1574-6976.2011.00292.x.

Wos, G., Choudhury, R. R., Kolář, F., and Parisod, C. (2021). Transcriptional activity of transposable elements along an elevational gradient in Arabidopsis arenosa. *Mobile DNA* 12, 7. doi:10.1186/s13100-021-00236-0.

Xi, Z., Bradley, R. K., Wurdack, K. J., Wong, K. M., Sugumaran, M., Bomblies, K., et al. (2012). Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics* 13, 227. doi:10.1186/1471-2164-13-227.

Xi, Z., Wang, Y., Bradley, R. K., Sugumaran, M., Marx, C. J., Rest, J. S., et al. (2013). Massive Mitochondrial Gene Transfer in a Parasitic Flowering Plant Clade. *PLoS Genetics* 9, 1–10. doi:10.1371/journal.pgen.1003265.

Xia, J., Guo, Z., Yang, Z., Han, H., Wang, S., Xu, H., et al. (2021). Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell*, 1–13. doi:10.1016/j.cell.2021.02.014.

Yang, Z., Wafula, E. K., Kim, G., Shahid, S., McNeal, J. R., Ralph, P. E., et al. (2019). Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants*. doi:10.1038/s41477-019-0458-0.

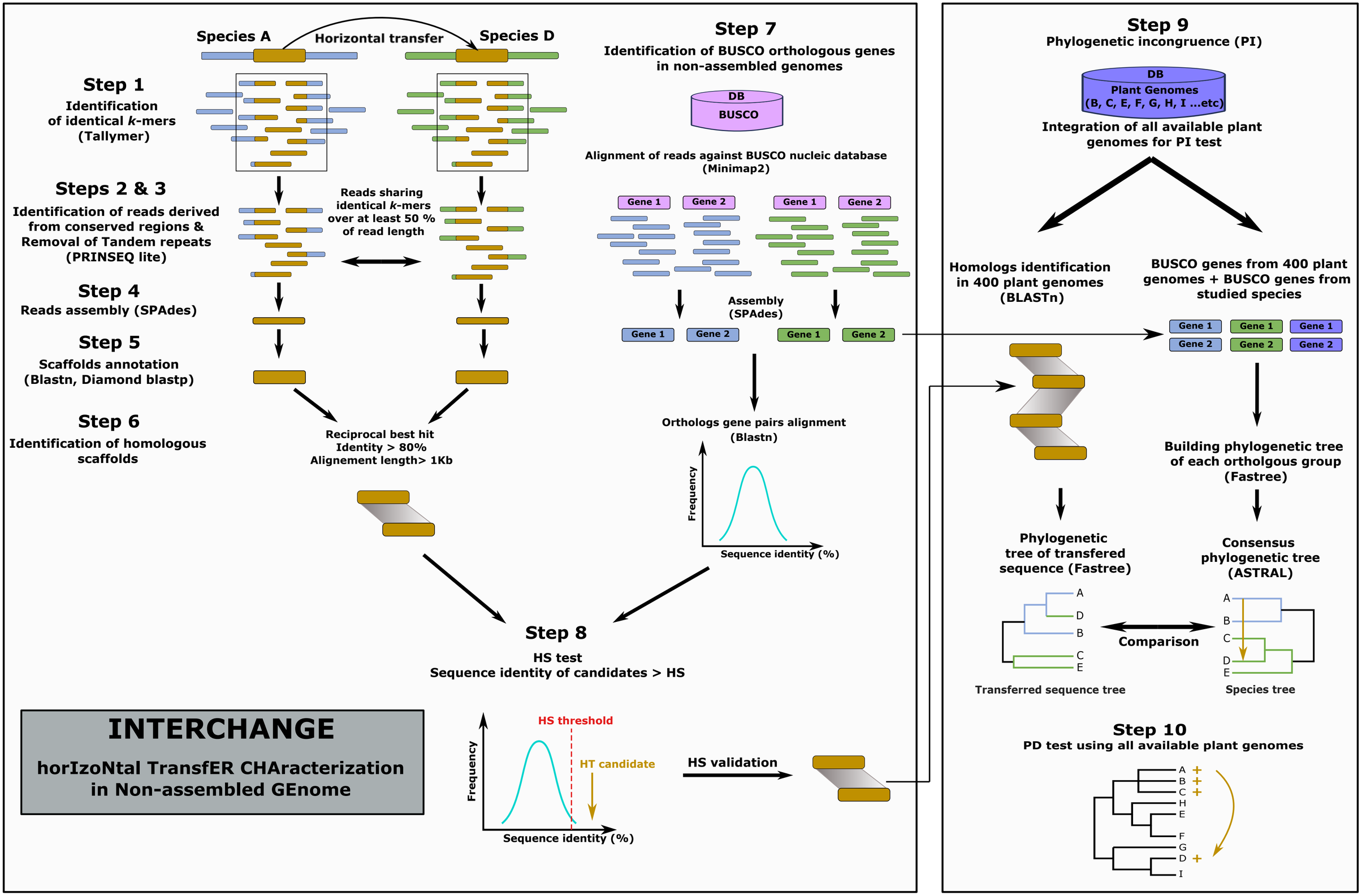
Yang, Z., Zhang, Y., Wafula, E. K., Honaas, L. A., Ralph, P. E., Jones, S., et al. (2016). Horizontal gene transfer is more frequent with increased heterotrophy and contributes to parasite adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 113, E7010–E7019. doi:10.1073/pnas.1608765113.

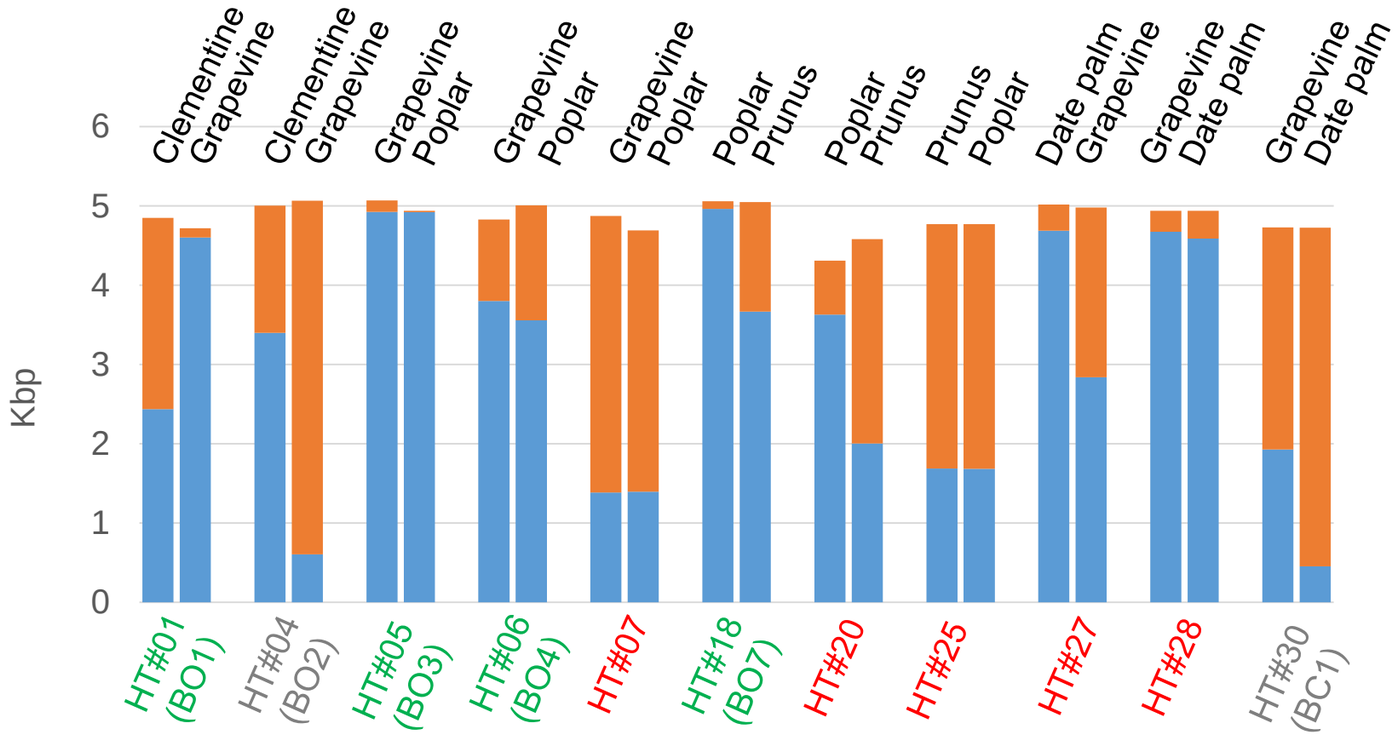
Yoshida, S., Maruyama, S., Nozaki, H., and Shirasu, K. (2010). Horizontal gene transfer by the parasitic plant striga hermonthica. *Science* 328, 1128. doi:10.1126/science.1187145.

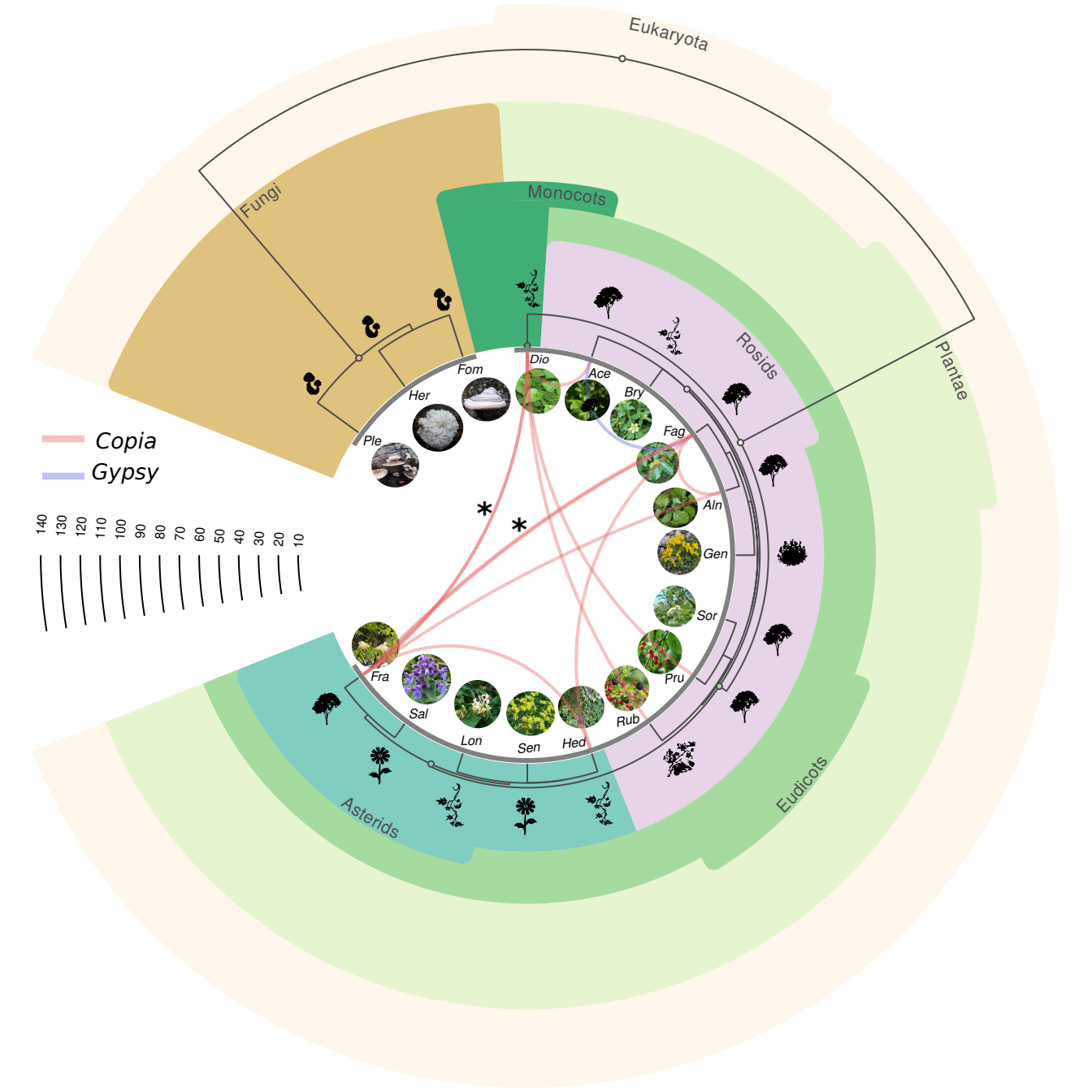
Yue, J., Hu, X., Sun, H., Yang, Y., and Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. *Nature Communications* 3, 1152. doi:10.1038/ncomms2148.

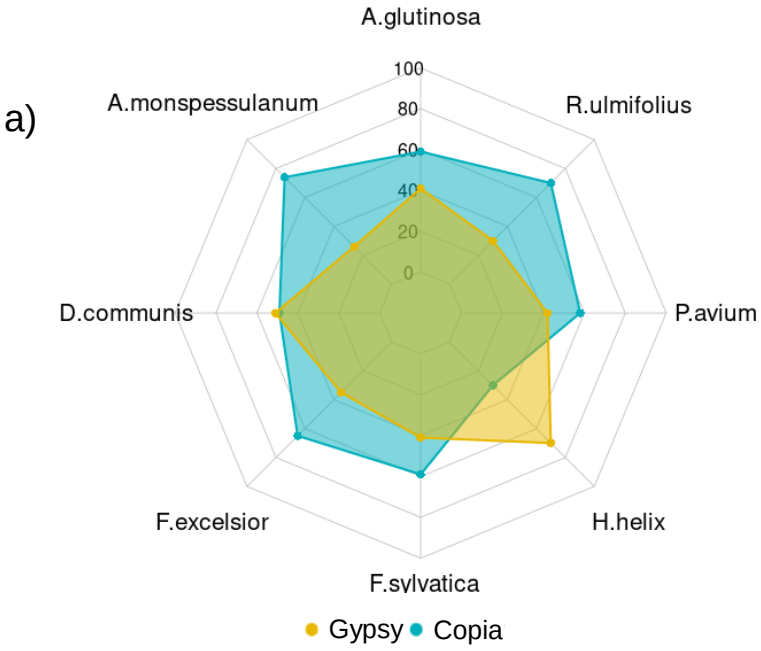
Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19, 153. doi:10.1186/s12859-018-2129-y.

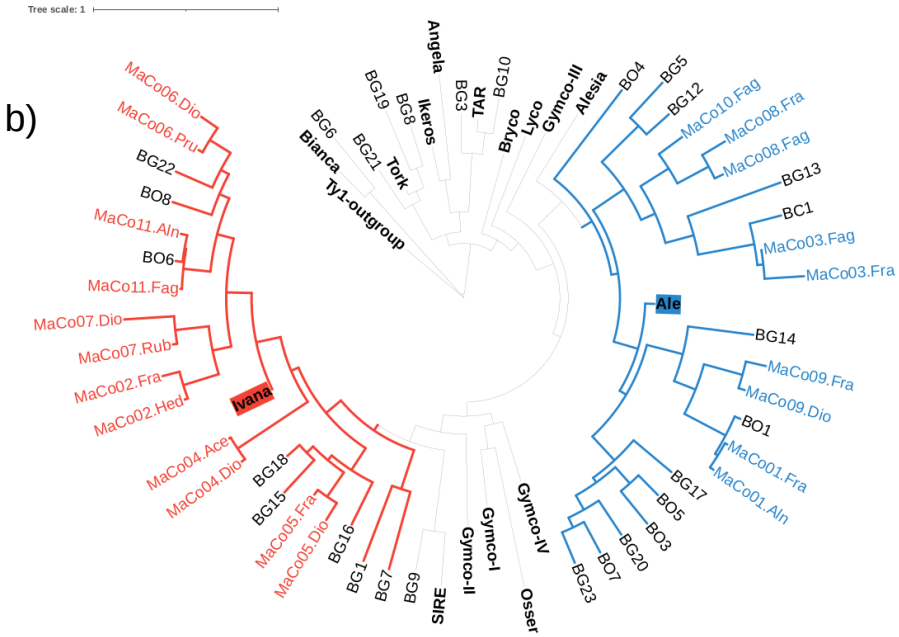
Zhang, H.-H., Peccoud, J., Xu, M.-R.-X., Zhang, X.-G., and Gilbert, C. (2020). Horizontal transfer and evolution of transposable elements in vertebrates. *Nature Communications* 11, 1362. doi:10.1038/s41467-020-15149-4.

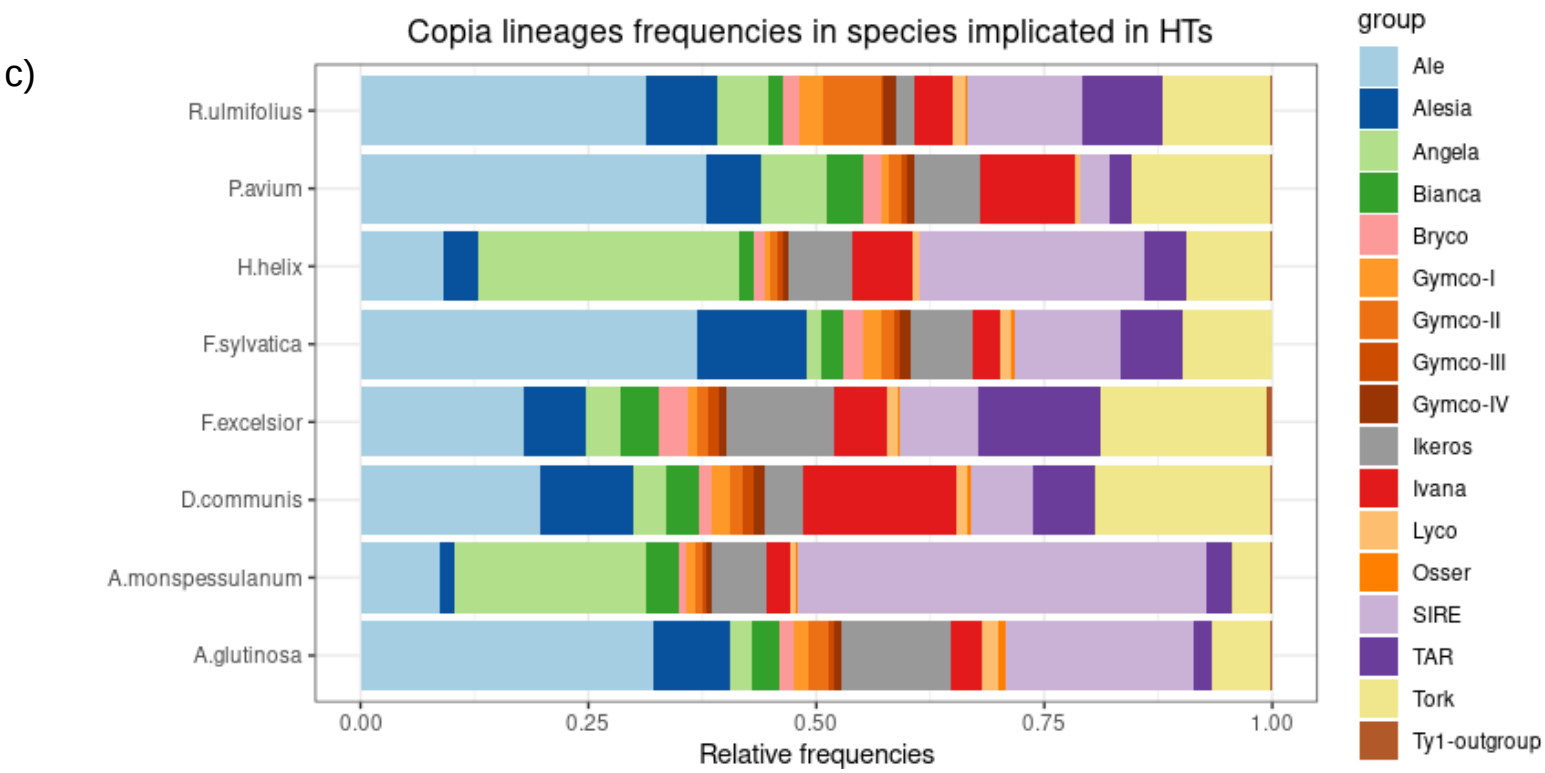

Species A
Horizontal transfer
Species D
Step 1
Identification of identical k-mers (Tallymer)
Steps 2 & 3
Identification of reads derived from conserved regions & Removal of Tandem repeats (PRINSEQ lite)
Reads sharing identical k-mers over at least 50 % of read length
Step 4
Reads assembly (SPAdes)
Step 5
Scaffolds annotation (Blastn, Diamond blastp)
Step 6
Identification of homologous scaffolds
Reciprocal best hit
Identity > 80%
Alignement length> 1Kb
Step 7
Identification of BUSCO orthologous genes in non-assembled genomes
DB
BUSCO
Alignment of reads against BUSCO nucleic database (Minimap2)
Gene 1
Gene 2
Assembly (SPAdes)
Orthologs gene pairs alignment (Blastn)
Frequency
Sequence identity (%)
Step 8
HS test
Sequence identity of candidates > HS
INTERCHANGE
horIzoNtal TransfER CHAracterization in Non-assembled GEnome
HS threshold
HT candidate
HS validation
Step 9
Phylogenetic incongruence (PI)
DB
Plant Genomes
(B, C, E, F, G, H, I ...etc)
Integration of all available plant genomes for PI test
Homologs identification in 400 plant genomes (BLASTn)
BUSCO genes from 400 plant genomes + BUSCO genes from studied species
Building phylogenetic tree of each ortholgous group (Fastree)
Phylogenetic tree of transfered sequence (Fastree)
Consensus phylogenetic tree (ASTRAL)
Comparison
Transferred sequence tree
Species tree
Step 10
PD test using all available plant genomes


Kbp
0
1
2
3
4
5
6
HT#01 (BO1)
HT#04 (BO2)
HT#05 (BO3)
HT#06 (BO4)
HT#07
HT#18 (BO7)
HT#20
HT#25
HT#27
HT#28
HT#30 (BC1)
Clementine
Grapevine
Clementine
Grapevine
Grapevine
Poplar
Grapevine
Poplar
Grapevine
Poplar
Poplar
Prunus
Poplar
Prunus
Prunus
Poplar
Date palm
Grapevine
Grapevine
Date palm
Grapevine
Date palm


Eukaryota
Fungi
Monocots
Rosids
Plantae
Eudicots
Asterids
Copia
Gypsy
140
130
120
110
100
90
80
70
60
50
40
30
20
10
Ple
Her
Fom
Dio
Ace
Bry
Fag
Aln
Gen
Sor
Pru
Rub
Hed
Sen
Lon
Sal
Fra
*
*


a)
A.glutinosa
R.ulmifolius
P.avium
H.helix
F.sylvatica
F.excelsior
D.communis
A.monspessulanum
100
80
60
40
20
0
Gypsy
Copia


b)
Tree scale: 1
Ivana
Ale
Ty1-outgroup
Bianca
Tork
Ikeros
Angela
TAR
Bryco
Lyco
Gymco-III
Alesia
Gymco-IV
Gymco-I
Gymco-II
SIRE
Osser
MaCo06.Dio
MaCo06.Pru
BG22
BO8
MaCo11.Aln
BO6
MaCo11.Fag
MaCo07.Dio
MaCo07.Rub
MaCo02.Fra
MaCo02.Hed
MaCo04.Ace
MaCo04.Dio
BG18
BG15
MaCo05.Fra
MaCo05.Dio
BG16
BG1
BG7
BG9
BG6
BG21
BG19
BG8
BG3
BG10
BO4
BG5
BG12
MaCo10.Fag
MaCo08.Fra
MaCo08.Fag
BG13
BC1
MaCo03.Fag
MaCo03.Fra
BG14
MaCo09.Fra
MaCo09.Dio
BO1
MaCo01.Fra
MaCo01.Aln
BG17
BO5
BO3
BG20
BO7
BG23


c)
Copia lineages frequencies in species implicated in HTs
R.ulmifolius
P.avium
H.helix
F.sylvatica
F.excelsior
D.communis
A.monspessulanum
A.glutinosa
0.00
0.25
0.50
0.75
1.00
Relative frequencies
group
Ale
Alesia
Angela
Bianca
Bryco
Gymco-I
Gymco-II
Gymco-III
Gymco-IV
Ikeros
Ivana
Lyco
Osser
SIRE
TAR
Tork
Ty1-outgroup