This Appendix includes Propositions 1 and 2, Theorems 1, 2 and 3 with their proofs as well as an example showing inappropriateness of the RF metric in the HGT recovery context.

Properties of the Bipartition Dissimilarity

The Propositions 1 and 2 establish some interesting properties of the bipartition dissimilarity. Thus, Proposition 1 states the sufficiency condition that ensures that a bipartition dissimilarity (BD) satisfies the triangle inequality (and is a metric), and Proposition 2 gives the maximum values of this measure depending on the number of tree leaves.

The bipartition $a$ of a tree $T$ is associated to the bipartition $b$ of the tree $T'$ (this association is denoted by $a \rightarrow b$), if the Hamming distance between the bipartition vectors corresponding to $a$ and $b$ is the smallest among all possible distances computed between $a$ and all the bipartition vectors corresponding to the branches of the tree $T'$. A sufficient metricality condition is as follows:

**Proposition 1.** Let $T_1$, $T_2$ and $T_3$ be phylogenetic trees with the same number of internal branches and the same sets of leaves. Then, if:

1. For any two bipartitions $a$ and $b$ from different trees: $a \rightarrow b$ implies that $b \rightarrow a$, and
2. For any three bipartitions $a \in T_1$, $b \in T_2$ and $c \in T_3$: $a \rightarrow b$ and $b \rightarrow c$ implies that $a \rightarrow c$,

then, the triangle inequality, $bd(T_1, T_2) \leq bd(T_1, T_3) + bd(T_2, T_3)$, holds.
Proof. On one hand, considering the first statement of Proposition:
\[ bd(T_1, T_2) = (\sum_{a \in BT_1} d(a, b) + \sum_{b \in BT_2} d(b, a)) \div 2 = \sum_{a \in BT_1} d(a, b), \text{ where } (a \in BT_1 \text{ and } b \in BT_2) \rightarrow a \]
\[ (b \in BT_2) \rightarrow a \] means that the sum is taken for all the \(a\)'s belonging to the bipartition table \(BT_1\) corresponding to the tree \(T_1\) and all the \(b\)'s associated with these \(a\)'s. In a similar way:
\[ bd(T_1, T_3) = (\sum_{a \in BT_1} d(a, c) + \sum_{b \in BT_3} d(c, a)) \div 2 = \sum_{a \in BT_1} d(a, c), \text{ and } \]
\[ bd(T_2, T_3) = (\sum_{b \in BT_2} d(c, b) + \sum_{b \in BT_3} d(c, b)) \div 2 = \sum_{a \in BT_1} d(c, b). \]
Consider the following three sums:
\[ \sum_{a \in BT_1} d(a, b), \sum_{a \in BT_1} d(a, c) \text{ and } \sum_{a \in BT_1} d(b, c). \]
Because the Hamming distance \(d\) satisfies the triangle inequality, for any term \(d(a, b)\) from the first sum we have the term \(d(a, c)\) from the second sum and the term \(d(b, c)\) from the third sum such that:
\[ d(a, b) \leq d(a, c) + d(b, c). \]
Because each of the bipartition vectors included in the bipartition tables \(BT_1\), \(BT_2\) and \(BT_3\) appears only once in each of the three sums we conclude that:
\[ bd(T_1, T_2) \leq bd(T_1, T_3) + bd(T_2, T_3). \]

Proposition 2. The value of the bipartition dissimilarity between two phylogenetic trees on the same sets of \(n\) leaves ranges from 0 to \(n(n-3)/2\) if \(n\) is even, and from 0 to \((n-1)(n-3)/2\) if \(n\) is odd.

Proof. For any two binary vectors \(a\) and \(b\) of size \(n\), the maximum value of the quantity \(\text{Min}(d(a, b); d(a, \bar{b}))\), where \(d(a, b)\) is the Hamming distance between \(a\) and \(b\), and \(\bar{a}\) and \(\bar{b}\) are their complements, is \(n/2\) when \(n\) is even and \((n-1)/2\) when \(n\) is odd. On the other hand, the maximum number of internal branches in a phylogenetic tree (i.e., number of rows of the corresponding bipartition table) with \(n\) leaves is \(n-3\). Consequently, according to Formula 2, the maximum value of the bipartition dissimilarity between two trees with \(n\) leaves is \(n(n-3)/2\) if \(n\) is even, and \((n-1)(n-3)/2\) if \(n\) is odd.
Theorem 1. If the newly-formed subtree $\text{Sub}_{yw}$ resulting from the HGT (i.e. the subtree rooted by the branch $(x,a)$ in Fig. 2) is present in the gene tree $T'$, and the bipartition vector associated with the branch $(x,x_1)$ in the transformed species tree $T_1$ (Fig. OA1) is present in the bipartition table of $T'$, then the HGT from $(x,y)$ to $(z,w)$, transforming $T$ into $T_1$, is a part of a minimum-cost HGT scenario transforming $T$ into $T'$ and satisfying the subtree constraint.

Proof. The four possible cases leading to the formation of the subtree $\text{Sub}_{yw}$ are the following: 1) HGT from $(x,y)$ to $(z,w)$; 2) HGT from $(z,w)$ to $(x,y)$; 3) HGT from $(x',x)$ to $(z,z')$; 4) HGT from $(z,z')$ to $(x',x)$. When the path $(x,z)$ in $T$ consists of two or more branches, the HGTs corresponding to the cases (3) and (4) will not produce the subtree $\text{Sub}_{yw}$, but bring the vertices $x$ and $z$ closer to each other by reducing the number of branches of the path $(x,z)$. The HGT cases (3) and (4) will induce the bipartition $b$, which will be present in the bipartition table the gene tree $T'$ because of the subtree constraint, such that the leaves of the subtree located to the left of $x'$ and those of the subtree located to the right of $z'$ (Fig. OA1) belong to the same part of it (e.g., they are denoted by 1’s in the bipartition table of $T'$), whereas the leaves of the subtree located below the vertices $y$ and $w$ belong to a different part of it (e.g., they are denoted by 0’s in the bipartition table of $T'$). According to the Theorem condition, the bipartition corresponding to the branch $(x,x_1)$ in the tree $T_1$ obtained from the initial species tree $T$ after the HGT from $(x,y)$ to $(z,w)$ was carried out, and denoted here $b_1$, is also present in the bipartition table of $T'$. This means that the leaves of the subtree located to the left of $x'$ and those of the subtree located below the vertices $y$ and $w$ belong to the same part of it, whereas the leaves of the subtree located to the right of $z'$ belong a different part of it. Obviously, the bipartitions $b$ and $b_1$ are incompatible (i.e., they cannot be present together in the same bipartition table.
associated with a phylogenetic tree) meaning that the HGTs from \((x',x)\) to \((z,z')\) and from \((z,z')\) to \((x',x)\) are impossible. Moreover, the HGT from \((z,w)\) to \((x,y)\) is possible only when the path \((x,z)\) in \(T\) consists of a single branch (in this case the opposite HGTs from \((x,y)\) to \((z,w)\) and from \((z,w)\) to \((x,y)\) will lead to the same topological transformation of \(T\)) because this HGT would induce a bipartition, denoted here \(b_2\), which is incompatible with \(b_1\) if the path \((x,z)\) in \(T\) consist of two or more branches. Indeed, in \(b_2\) the leaves of the subtree located to the right of \(z'\) and those of the subtree located below the vertices \(y\) and \(w\) belong to the same part of it, whereas the leaves of the subtree located to the left of \(x'\) belong a different part of it. Consequently, the HGT from \((x,y)\) to \((z,w)\) is necessary to transform \(T\) into \(T'\). The only exception from this would be the case of the opposite HGT from \((z,w)\) to \((x,y)\) which is possible only if the path \((x,z)\) consists of (or was reduced to) a single branch. In this case the opposite HGTs will lead to the same topological transformation and any of them is a part of a minimum-cost HGT scenario transforming \(T\) into \(T'\) and satisfying the subtree constraint. □

**Figure OA1.** HGT from the branch \((x,y)\) to the branch \((z,w)\) is a part of a minimum-cost HGT scenario transforming the species tree \(T\) into the gene tree \(T'\) if the bipartition corresponding to the branch \((x,x_1)\) in the transformed species tree \(T_1\) is present in the bipartition table of \(T'\) and the subtree \(Sub_{yw}\) (i.e., obtained by the SPR move induced by this HGT, see Fig. 2) is present in \(T'\);
**Theorem 2.** If the newly-formed subtree $\text{Sub}_{yw}$ resulting from the HGT (i.e., the subtree rooted by the branch $(x,a)$ in Fig. 2) is present in the gene tree $T'$, and all the bipartition vectors associated with the branches of the path $(x',z')$ in the transformed species tree $T_1$ (Fig. OA2) are present in the bipartition table of $T'$, and the path $(x',z')$ in $T_1$ consists of at least 3 branches, then the HGT from $(x,y)$ to $(z,w)$, transforming $T$ into $T_1$, is a part of any minimum-cost HGT scenario transforming $T$ into $T'$ and satisfying the subtree constraint.

**Proof.** The bipartition vectors corresponding to the branches $(x',x)$ and $(z,z')$ of the transformed species tree $T_1$ obtained from $T$ after the HGT from $(x,y)$ to $(z,w)$ are also present in the bipartition table of the species tree $T$ and gene tree $T'$. Thus, the four possible cases leading to the formation of the subtree $\text{Sub}_{yw}$ are the following: 1) HGT from $(x,y)$ to $(z,w)$; 2) HGT from $(z,w)$ to $(x,y)$; 3) HGT from $(x',x)$ to $(z,z')$; 4) HGT from $(z,z')$ to $(x',x)$. When the path $(x,z)$ in $T$ consists of two or more branches, the HGTs corresponding to the cases (3) and (4) will not produce the subtree $\text{Sub}_{yw}$, but bring the vertices $x$ and $z$ closer to each other by reducing the number of branches of the path $(x,z)$.

According to the Theorem condition, all the bipartitions of the non-empty path $(x,z)$ in $T_1$ obtained from the initial species tree $T$ after the HGT from $(x,y)$ to $(z,w)$ are also present in the bipartition table of the gene tree $T'$. Consequently, the leaves of the subtree located to the left of $x'$ and those of the subtrees located below the vertices $y$ and $w$ (Fig. OA2) belong to a different part (e.g., they are denoted by 1’s in the bipartition table of $T'$) of these bipartitions than the leaves of the subtree located to the right of $z'$ (e.g., they are denoted by 0’s in the bipartition table of $T'$). This means that there is no bipartition in $T'$ such that all the leaves located in the subtrees to the left of $x'$ and to the right of $z'$ would belong to one part of it and those from the subtrees located below the vertices $y$ and $w$, to the other. Thus, the HGT from $(x',x)$ to $(z,z')$, case (3), as well as the opposite HGT from
(z,z') to (x',x), case (4), will violate the subtree constraint. Obviously, any HGT from the branches (x',x) and (z,z') to the branches of the path (x,z) will also violate the subtree constraint.

**Figure OA2.** HGT from the branch (x,y) to the branch (z,w) is a part of any minimum-cost HGT scenario transforming the species tree $T$ into the gene tree $T'$ if all the bipartitions corresponding to the branches of the path (x',z') in the transformed species tree $T_1$ are present in the bipartition table of $T'$ and the subtree $Sub_{yw}$ (i.e., obtained by the SPR move induced by this HGT, see Fig. 2) is present in the tree $T'$.

Therefore, either the HGT from (x,y) to (z,w) or the opposite HGT from (z,w) to (x,y) is a part of any minimum-cost HGT scenario transforming $T$ into $T'$ and satisfying the subtree constraint. After the HGT from (x,y) to (z,w), all the bipartition vectors corresponding to the branches of a non-empty path (x',z'), in Figure OA2, will be present in the bipartition table of $T'$, and none of them in the case of the opposite HGT from (z,w) to (x,y). As the bipartitions associated with the branches $(x_i,x_{i+1})$ and $(x_{i+1},x_{i+2})$, where $i = 0,\ldots, k-1$, and $x_0 = x'$ and $x_{k+1} = z'$ (Fig. OA2), are present in the bipartition table of $T'$, the bipartition associated with the branch $(x_{i+1}, y_{i+1})$ is also present in the bipartition table of $T'$. This means that the subtrees rooted by the branches $(x_1,y_1)$ to $(x_k,y_k)$ can be arranged independently (according to the topology of the gene tree $T'$) if it is not done already, from each other and from the rest of the tree $T_1$ (i.e., this means that the SPR
operations will be carried out only within these subtrees and inter-subtree SPRs will not be necessary). In the same way, in a minimum-cost scenario the arrangements of the subtrees located to the left of $x'$ and those located to the right of $z'$ (Fig. OA2) should be done independently of the rest of the tree and will take the same minimum number of SPR operations in the case of the HGT from $(x,y)$ to $(z,w)$ and the opposite HGT from $(z,w)$ to $(x,y)$. Consequently, in the case of the opposite HGT from $(z,w)$ to $(x,y)$, the SPR transformation of the tree $T_1$ into the gene tree $T'$ will take at least one SPR operation more, needed to arrange the branches of the path $(x,z)$, than in the case of the HGT from $(x,y)$ to $(z,w)$.

**Theorem 3.** If the bipartition vectors corresponding to the branches $(x,x')$ and $(z,z')$ of the species tree $T$ (Fig. OA3) are present in the bipartition table of the gene tree $T'$ and the newly-formed subtree, denoted here Sub$_{yz}$, induced by the HGT (e.g., the subtree rooted by the branch $(x,a)$ in Fig. 2) is present in $T'$, then either the HGT from $(x,y)$ to $(z,w)$ or the opposite HGT from $(z,w)$ to $(x,y)$, transforming the species tree $T$ into $T_1$, is a part of a minimum-cost HGT scenario transforming $T$ into $T'$ and satisfying subtree constraint.

**Proof.** The species tree $T$ (Fig. OA3) can be subdivided into three subtrees by cutting the branches $(x,x')$ and $(z,z')$. Subtree 1 is rooted by the vertex $x'$ and located to the left of $x'$; Subtree 2 is rooted by the vertex $z'$ and located to the right of $z'$; and, Subtree 3 is formed by the subtrees grafted to the path $(x,z)$ and by the branches $(x,x')$ and $(z,z')$. The fact that the bipartitions associated with $(x,x')$ and $(z,z')$ of the species three $T$ are present in the bipartition table of the gene tree $T'$ means that any minimum-cost scenario transforming $T$ into $T'$ does not include HGTs between the branches of different Subtrees, but only those within each of them because any HGT between the branches of two different
Subtrees will result in the violation of the subtree constraint (Fig. 2). Any HGT satisfying this constraint preserves all existing identical bipartitions in $T$ and $T'$.

Consider now Subtree 3. The bipartition vectors corresponding to the branches $(x,x')$ and $(z,z')$ of the species tree $T$ are also present in the bipartition table of the gene tree $T'$. Assume that the path $(x,z)$ in $T$ consists of a single branch. In this case, the four possible cases leading to the formation of the subtree $Sub_{yw}$ are the following: 1) HGT from $(x,y)$ to $(z,w)$; 2) HGT from $(z,w)$ to $(x,y)$; 3) HGT from $(x',x)$ to $(z,z')$; 4) HGT from $(z,z')$ to $(x',x)$. Each of these HGTs leads to the same topology of the transformed species tree $T_1$ and satisfies the subtree constraint, and, consequently, is a part of a minimum-cost scenario transforming $T$ into $T'$. Thus, when the path $(x,z)$ in $T$ consists of a single branch, the real HGT direction is undetectable.

**Figure OA3.** Either the HGT from $(x,y)$ to $(z,w)$ or the opposite transfer from $(z,w)$ to $(x,y)$ is a part of a minimum-cost HGT scenario transforming $T$ into $T'$ if the bipartitions induced by the branches $(x,x')$ and $(z,z')$ in $T$ are present in the bipartition table of $T'$ and the newly-formed subtree $Sub_{yw}$ resulting from one of these HGTs is present in the tree $T'$.

Assume now that the path $(x,z)$ in $T$ consists of more than one branch. To form the subtree $Sub_{yw}$ and satisfy the subtree constraint, we can either directly carry out the HGTs (cases 1 and 2) from $(x,y)$ to $(z,w)$ or from $(z,w)$ to $(x,y)$, or regraft by SPR moves all the subtrees of the path $(x,z)$, except those including the branches $(x,y)$ and $(z,w)$, to the
branches \((x',x)\) or \((z,z')\), and then proceed by the SPR moves (cases 3 and 4) from \((x',x)\)
to \((z,z')\), or from \((z,z')\) to \((x',x)\).

Assume that a minimum-cost scenario \(S_{\text{min}}\) of the SPR reconciliation of the trees \(T\) and
\(T'\) does not include the HGTs from \((x,y)\) to \((z,w)\) and from \((z,w)\) to \((x,y)\), and proceeds as
follows: first, it reduces the path \((x,z)\) to a single branch, and at the last step, merge the
vertices \(x\) and \(z\) to form the subtree \(Sub_{yw}\) by the SPR move from \((x',x)\) to \((z,z')\) or from
\((z,z')\) to \((x',x)\). It is worth noting that at the last step of the reduction process, the branches
\((x',x)\) and \((z,z')\) can be substituted by the other ones before the last SPR move if a HGT
between them has taken place beforehand.

We will now show that there is another SPR scenario of the same length including
either the HGT from \((x,y)\) to \((z,w)\) or that from \((z,w)\) to \((x,y)\). Without loss of generality
assume that in the scenario \(S_{\text{min}}\) there is a HGT from the branch \((x',x)\) to a subtree grafted
to the path \((x,z)\) and induced by the branch denoted here by \((x_i,y_i)\), see Figure OA3, except
those induced by \((x,y)\) and \((z,w)\), and that this HGT reduces the path \((x,z)\) to a single
branch. In \(S_{\text{min}}\), the latter HGT should be followed by another HGT, from \((x',x)\) to \((z,z')\)
or from \((z,z')\) to \((x',x)\), initiating the formation of the subtree \(Sub_{yw}\). However, there exists
another SPR scenario \(S\), of the same length that \(S_{\text{min}}\), which starts by the HGT from \((z,w)\)
to \((x,y)\), thus eliminating the vertex \(x\), and brings all the subtrees grafted on the path \((x,y)\),
including that induced by the branch \((x_i,y_i)\), one branch closer to the vertex \(x'\). The latter
HGT will make the transfer from \((x',x)\) to \((x_i,y_i)\), of the scenario \(S_{\text{min}}\), unnecessary. All the
other HGTs of \(S_{\text{min}}\), that are necessary to arrange the branches grafted to the path \((x',z')\)
according to the topology of the gene tree \(T'\), will be similar in the scenarios \(S\) and \(S_{\text{min}}\),
thus confirming the optimality of the HGT scenario \(S\). □
Notice: Obviously, in the proofs of Theorems 1, 2 and 3 we assume that the branches \((x,y)\) and \((z,w)\) do not belong to the same lineage. Otherwise, the HGT between them is impossible due to the evolutionary constraints. Also, without loss of generality we assume that \(T\) and \(T'\) are binary trees.
The RF Metric and SPR Distance

Figure OA4 illustrates a typical situation when the RF metric is unsuitable for finding an optimal scenario of SPR transformations. It shows a HGT in a binary “caterpillar-shaped” tree with \( n \) leaves. Here, the species phylogeny \( T \) is the tree before the transfer and the gene phylogeny \( T' \) is the tree after it. Thus, the SPR distance between \( T \) and \( T' \) is 1, whereas the RF distance between them equals to its maximum possible value \( 2n-6 \). This example suggests that the RF metric is not a very appropriate measure to approximate the SPR distance. On the other hand, the value of bipartition dissimilarity between \( T \) and \( T' \) is \( n-3 \), whereas its maximum value for the case of two binary trees with \( n \) leaves is \( n(n-3)/2 \) when \( n \) is even, and \( (n-1)(n-3)/2 \) when \( n \) is odd (see Proposition 2).

**Figure OA4:** The SPR move, representing a HGT, from the branch \((x_1, y)\) to the branch \((x_n, z)\) transforms the species tree \( T \) into the gene tree \( T' \). The RF distance between \( T \) and \( T' \) equals to its maximum value \( 2n-6 \), while the SPR distance between \( T \) and \( T' \) is 1. In this example, the tree root node is incident to a node of the path \((y, z)\), and the tree leaves are denoted by \( x_1 \ldots x_n \).
**Figure OA5.** Percentage of instances the algorithms recover: (a) Correct horizontal gene transfers, and (b) Complete correct HGT scenario, versus the number of HGTs, under the condition of known species and gene trees (i.e., tree-like data). The four compared algorithmic strategies were based on RF, QD, LS and BD. *Each reported value* represents the average result obtained for random trees with 10, 20 … 100 leaves (1000 replicates were generated for each tree size).
Figure OA6. Percentage of instances when LatTrans (white columns) and algorithm based on the bipartition dissimilarity (grey columns) recover: Correct HGTs (cases a and b), Correct total number of HGTs (cases c and d) and Complete correct HGT scenario (cases e and f) depending on the number of tree leaves (cases a, c and e) and number of HGTs (cases b, d and f). Each reported value represents the average result obtained for the set of random trees with 1 to 10 HGTs (cases a, c and e) and 10, 20 … 100 leaves (cases b, d and f); 1000 replicates were generated for each number of HGTs and each tree size.
**Figure OA7.** Running time in seconds for LatTrans (white squares) and algorithmic strategy based on the bipartition dissimilarity (grey squares) depending on the: (a) Number of transfers, and (b) Number of tree leaves. *Each reported value* represents the average result obtained for the set of random trees with: (a) 1 to 10 HGTs, and with (b) 10, 20 … 100 leaves (1000 replicates were generated for each number of HGTs and each tree size).
Online Appendix 3

This Appendix includes:

1) An illustration of computing HGT bootstrap support by the *RIATA-HGT* program.

2) The input data for the *rpl12e* and *PheRS Synthetase* examples and the exact output data provided by the *RIATA-HGT* program. Both the text output and solution screenshots are reported in this Appendix.

![Species tree + HGT](image)

**Figure OA8.** Computing the bootstrap support of a HGT branch with *RIATA-HGT*. The score of the HGT branch $X \rightarrow Y$ added to the species tree is defined as the maximum bootstrap score of all internal branches of the path linking the nodes $Z$ and $X$ in the gene tree. The bootstrap support of the event $X \rightarrow Y$ given by *RIATA-HGT* in this case is 100%. In our method, the bootstrap support of this HGT event would be at most 10%.

**RIATA-HGT output for the rpl12e and PheRS Synthetase examples**

**Input data 1 (Example of the gene rpl12e):**

```plaintext
(((A.pernix, S.solfataricus), P.aerophilum), (((P.abyssi, P.horikoshii), P.furiosus), (M.jannaschii, M.thermoaut.), ((T.acidophilum, F.acidarananus), ((Halobacterium.sp., H.marismortui), M.barkeri), A.fulgidus))));
```

```plaintext
(((P.aerophilum::0.0, S.solfataricus::0.0)::74.0, A.pernix::0.0)::79.0, T.acidophilum::0.0)::79.0, F.acidarananus::0.0)::100.0, ((P.horikoshii::0.0, P.furiosus::0.0)::61.0, P.abyssi::0.0)::81.0, (((Halobacterium.sp::0.0, H.marismortui::0.0)::100.0, M.thermoaut::0.0)::56.0, M.barkeri::0.0)::56.0, A.fulgidus::0.0)::51.0, M.jannaschii::0.0)::65.0)::100.0);
```

*Notice:* Bootstrap scores in the gene tree are indicated after “::”. Bootstrap scores of the gene tree branches adjacent to the leaves were set to 0.0 in the Newick string, otherwise *RIATA-HGT* was unable to compute the correct HGT bootstrap support.
Output data 1:

species tree:


gene tree:

(((F.acidinaranus,(((P.aerophilum,S.solfataricus),A.pernix),T.acidophilum)),(((P.horikoshii,P.furiosus),P.abyssi),(((Halobacterium.sp.,H.marismortui)I0,M.thermoaut.)M.barkeri,A.fulgidus)I12);

There are 3 component(s), which account(s) for 9 solution(s), each of size 5

Component I12:

Subsolution1:
I0 -> M.thermoaut. (56.0)
I11 -> F.acidinaranus (100.0)
I11 -> T.acidophilum (100.0)

Component I11:

Subsolution1:
I11 -> A.pernix (74.0) [time violation?]
Subsolution2:
S.solfataricus -> P.aerophilum (74.0)
Subsolution3:
P.aerophilum -> S.solfataricus (74.0)

Component I8:

Subsolution1:
P.horikoshii -> P.furiosus (61.0)
Subsolution2:
P.furiosus -> P.horikoshii (61.0)
Subsolution3:
I8 -> P.abyssi (61.0) [time violation?]

Consensus network for this set of gene trees:

For the example of the rpl12e data RIATA-HGT found 9 solutions, each of size 5. Five of these solutions contradict the same lineage constraint (they include HGTs marked by [time violation?] in the program output) and four of them satisfy all plausible evolutionary constraints (e.g., the solution represented in Fig. 6 is among the four eligible solutions). HGT bootstrap scores are indicated between the parentheses in the program output.

**Input data 2 (Example of PheRS Synthetase):**

```
```

*Notice:* Bootstrap scores in the gene tree are indicated after “::”. Bootstrap scores of the gene tree branches adjacent to the leaves were set to 0.0 in the Newick string, otherwise RIATA-HGT was unable to compute the correct HGT bootstrap support.
Output data 2:

species tree:

gene tree:

There are 3 component(s), which account(s) for 12 solution(s), each of size 14

-----------------------------------------------------------------------------------------------------
Component I21:
Subsolution1:
   I17 -> P.hori
   P.hori -> I14 (100.0)
   I16 -> M.ther (25.0)
   I16 -> A.fulg (25.0)
Subsolution2:
   P.hori -> I14 (100.0)
   P.hori -> I17 (85.0)
   I16 -> A.fulg (25.0)
   I16 -> M.ther (25.0)
Subsolution3:
   A.fulg -> I16 (100.0)
   I17 -> M.jann (25.0)
   I17 -> P.hori
   P.hori -> I14 (100.0)
Subsolution4:
   I19 -> M.jann (88.0) [time violation?]
   I16 -> A.fulg (25.0)
   P.hori -> I14 (100.0)
   I16 -> M.ther (25.0)
-----------------------------------------------------------------------------------------------------
Component I2:
Subsolution1:
   R.prow -> H.pilo (67.0)
   R.caps -> T.mari (31.0)
   I4 -> I3 (85.0)
   I11 -> R.prow (0.0)
   I4 -> I13 (100.0)
   R.caps -> M.tube
   P.aeru -> N.gono (55.0)
   M.tube -> C.acet (59.0)
   I4 -> I15 (19.0)
-----------------------------------------------------------------------------------------------------
Component I3:
Subsolution1:
   P.ging -> C.trac (85.0)
Subsolution2:
   C.trac -> P.ging (85.0)
Subsolution3:
   I3 -> C.tepi (85.0) [time violation?]
***************************************************************************
Consensus network for this set of gene trees
P.ging -> C.trac
C.trac -> P.ging
I3 -> C.tepi
R.prow -> H.pilo
R.caps -> T.mari
I4 -> I3
I11 -> R.prow
I4 -> I13
R.caps -> M.tube
P.aeru -> N.gono
M.tube -> C.acet
I4 -> I15
I17 -> P.hori
P.hori -> I14
I16 -> M.ther
I16 -> A.fulg
P.hori -> I17
A.fulg -> I16
I17 -> M.jann
I19 -> M.jann
Figure OA10. For the example of the *PheRS Synthetase* data, *RIATA-HGT* found 12 solutions, each of size 14. The HGTs contradicting the same lineage constraint are marked by [time violation?] in the program output. Five initial tree transformations indicated by the dashed ellipses were made by *RIATA-HGT* prior to carrying out HGT detection (see the transformed input Newick string of the species tree in the program input). Each of these transformations corresponds to a trivial HGT (i.e., HGTs between the sister taxa from the same multifurcation). Thus, the presented solution actually consists of 19 HGTs, including 14 regular and 5 trivial HGTs. The minimum-cost solution presented in Figure 8, and comprising of 7 regular and 10 trivial HGTs, was not found by *RIATA-HGT*. HGT bootstrap scores are indicated between the parentheses in the program output.