INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 7, Issue 4 - V7I4-1862)
Available online at: https://www.ijariit.com

Pareto-optimal phylogenetic tree reconciliation considering duplications, transfers, losses, and incomplete lineage sorting

Rahnuma Tasmin
rahnuma053@yahoo.com
Military Institute of Science and Technology (MIST), Dhaka, Bangladesh

Dr. Md. Abul Kashem Mia
kashem@cse.buet.ac.bd
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

ABSTRACT

Phylogenetic tree is a representation of the evolutionary histories of different species from which the gene tree is constructed from some specific sequence copies of gene which are sampled from a group of species. For some biological events like gene Duplication (D), Horizontal Gene Transfer (T), Loss (L) and Incomplete Lineage Sorting (ILS), with the passage of time, the history of gene evolution and species evolution might show discord in most cases. Though many researchers do not account ILS event as discordance between gene tree and species tree because of in theory, ILS is not a true "gene event" such as duplication or a transfer, since nothing "happens" to the gene during incomplete lineage sorting. But still this phenomenon can lead a gene tree differing from the species tree just like speciation acts on populations. Reconciliation is a process of resolving disagreement between gene and species tree with least possible evolutionary event costs. Addressing this problem (not accounting ILS event), we propose an efficient pareto-optimal algorithm to reconcile a binary gene tree with binary species tree under Duplication, Transfer, Loss and Incomplete Lineage Sorting parsimony criterion, which is the extension of the previous pareto-optimal algorithm developed by Libeskind-Hadas et al., 2014[5]. Through this research it has been shown that how to properly cost DTL events considering ILS event and then give a fixed-parameter tractable (FPT) algorithm which calculates the most parsimonious Duplication (D), Horizontal Gene Transfer (T), Loss (L) and Incomplete Lineage Sorting (ILS) reconciliations. Comparing with the previous pareto-optimal algorithm, the space complexity will remain same after considering the ILS events. This is the first pareto-optimal reconciliation algorithm to consider all of the four evolutionary process driving tree incongruence.

Keywords: *Phylogenetic, Incomplete Lineage Sorting, Reconciliation, Coalescence, Pareto-optimality.*

1 INTRODUCTION

A Phylogeny is a measure of shared history and separate history for any two species. The larger two species have a common history, the more similar they are expected to be, on average. This is a tree showing relationship between lineages which might be computed for genome-wide DNA or from only a single gene. If the phylogenetic tree is computed from genome-wide DNA, then it is actually computing a species tree, although some sister lineages might not perfectly fit the definition of species; on the other hand, if the phylogenetic tree is computed from data coming from a single gene, then it is considered as gene tree.

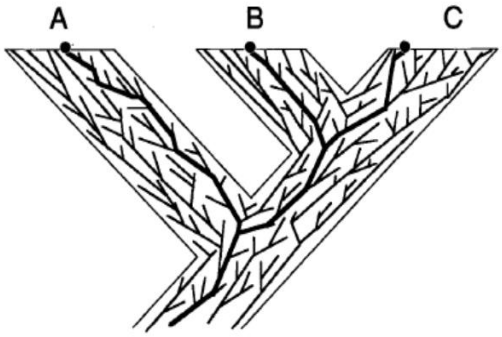

Fig. 1: A gene tree contained within the species tree leading to three extant species: A, B and C. Bold branches of gene tree show relationships among the sampled copies of gene ( ). Sampled copies from sister species B and C are sister copies.

Hence, the evolutionary history of a group of gene tree is strongly influenced by the evolutionary history of a group of species tree. Gene Duplication, Gene Loss, Horizontal Gene Transfer (HGT) or Incomplete Lineage Sorting (ILS) can result in a gene tree that differs from the species tree (Maddison, 1997) [6].

Basically, two methods have been proposed to reconcile the gene tree with the corresponding species tree, using gene-specific events – one is called Probabilistic Methods, which notice the most probable reconciliation under the applied mathematics model of evaluation and another is Maximum Parsimony, which assumes that the evolutionary change is rare and minimizes the amount of character – state changes (e.g., number of DNA substitutions). In this paper, the second method is used to coerce the newly developed algorithm more efficient and versatile.

Whole genome sequencing data are revealing an ever-growing number of cases where all four processes are active (e.g., Andersson, 2009; Serres et al., 2009; Zhaxybayeva and Doolittle, 2011), [1] [10] [14] [4] leading to calls for algorithms that model multiple evolutionary processes (Degnan and Rosenberg, 2009; Edwards, 2009).

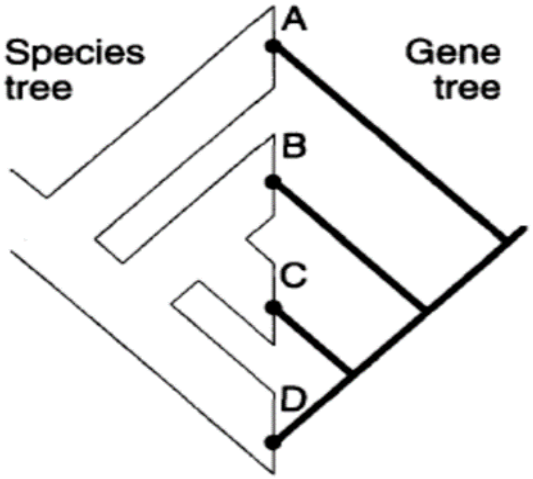

Fig.2: Discord between gene and species trees. At left is the species tree of four species, A, B, C and D, and at the right is the tree of a gene sampled one copy per species. Species B and C are sister species, but their gene copies are not sister copies.

It has been realized for some years than when ancestral polymorphisms persist through several speciation events, the subsequent loss or failure to sample some of the gene forms in the various species can give a gene tree with a topology different from that of the species tree (i.e., lineage sorting; Avise et al., 1983; Tajima, 1983; Takahata and Nei, 1985; Neigel and Avise, 1986; Nei, 1987) [2] [9] [12]. In deep coalescent event where two or more lineages fail to coalesce; the outcome of this failure can be formed a new speciation or can be duplicated gene variants (alleles). This phenomenon can be easily understood by Fig:2.

Here, we actually develop the previous pareto-optimal algorithm which considered only three evolutionary events except Incomplete Lineage Sorting. Algorithm that does not count ILS event will overestimate the number of duplication and/or transfers [Strolzer et al., 2012] [11] For example, a recent analysis based on a model that did not consider ILS, reported an inexplicable but dramatic increase in duplications in recently sequenced mammalian genomes (Milinkovitch et al., 2010) [8].

Maximum parsimony reconciliation depends on the event cost. In the DTLI model, speciation and Incomplete Lineage Sorting are considered "null events" and are therefore typically assigned a cost of 0 while duplication, transfers and losses are assigned positive cost.

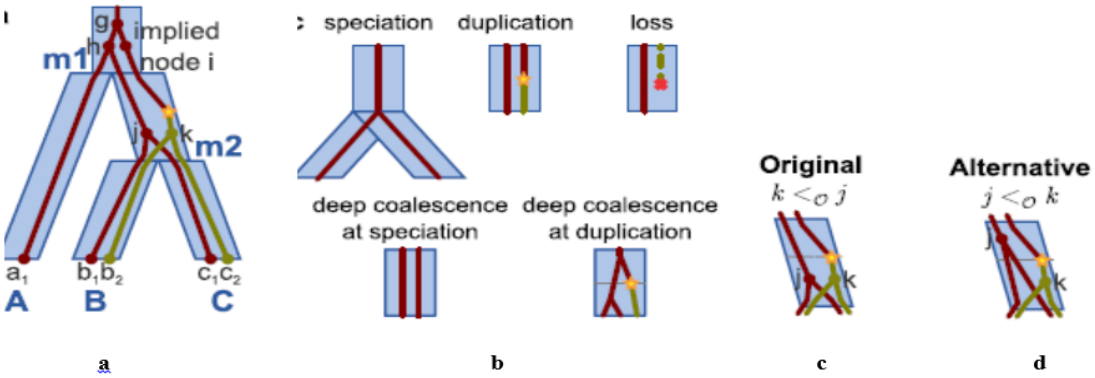

Fig.3: The Labeled coalescent tree. **a** Evolution is represented using the LCT. In this example, a duplication (yellow star) creates a new locus, "locus 2" (yellow), from the original locus "locus 1" (red), and lineages j and k fail to coalesce within species m2. **b** Evolutionary events are depicted in the LCT. Except for speciation evolution within a single species tree branch is shown. **c** An alternative scenario is presented for evolution in species m2. The new partial order induces an extra lineage at the time of the duplication. [Figure and caption adapted with permission from Du et al.,2019, and Wu et al.,2014]

As we work on the pareto-optimal algorithm considering ILS events, the main concept of pareto-optimal event count vectors will be same as the previous algorithm. Maximum Parsimony reconciliation depends on the event cost. Partitioning the space of possible event cost assignments into equivalence classes, or 'regions', such that any two event cost assignments within the same region lead to the same optimal reconciliation (Libeskind-Hadas et al., 2014) [5].

1.1 Previous Work

TreeMap (Charleston, 1998) [3] was the first to consider the problem to solve with pareto-optimal solution but the algorithm has worst case exponential time and can be applied only on small tree. Later, Tofigh (2009) [13] considering to solve the problem only accounting duplication and transfer but not accounting for losses. Later, Libeskind-Hadas et al., (2014) [5] considering to solve the problem accounting duplication, transfer and loss but not accounting incomplete lineage sorting. Recently, Mawhorter et al., (2019) [7], develop the inferring pareto-optimal algorithm considering duplication, loss and incomplete lineage sorting but not considering horizontal gene transfer.

To our knowledge, no paroto-optimal algorithms that consider all four evolutionary events to correctly estimate the event costs.

1.2 Our Contribution

(a) We extend the previously developed dp-algorithm of Libeskind-Hadas et al. (2014) [5] by considering the one of the important evolutionary event ILS. By considering this event the computation of event cost will more appropriate than the previous developed algorithm.
(b) Most importantly, the space complexity will be the same as the previous developed algorithm with considering ILS, as partitioning the space of possible event cost happened due to ILS will be fall in duplication or transfer or loss regions.

2 DEFINITIONS AND PRELIMINARIES

Preliminaries

Here, we formalize the concept of reconciliations and maximum parsimony reconciliation under DTLI model. We actually follow the basic definition and notation from Libeskind-Hadas et al., 2014 [5]. Here, the labeled coalescent tree is added (LCT, **Fig:3 a**) which formalizes the notation of a reconciliation in the DTLI model.

Given a tree T, denoting its node, edge and leaf sets by V(T), E(T) and Le(T), respectively. If T is rooted, the root node of T is denoted by rt(T), the parent of a node $v \in V(T)$ by $pa_T(v)$, its set of children by $Ch_T(v)$ and the (maximal) subtree of T rooted at v by T(v). The set of internal nodes of T, denoted I(T), is defined to be V(T)\Le(T). Here, $\leq_T$ is defined to be the partial order on V(T) where $x \leq_T y$ if y is a node on the path between rt(T) and x. The partial order $\geq_T$ is defined analogously, i.e. $x \geq_T y$ if x is a node on the path between rt(T) and y. It is said that y is an ancestor of x, or that x is a descendant of y, if $x \leq_T y$ (note that, under this definition, every node is a descendant as well as ancestor of itself). It is said that x and y are incomparable if neither $x \leq_T y$ nor $y \leq_T x$. Given a non-empty subset $L \subseteq Le(T)$, denoted by $lca_T(L)$ the last common ancestor (LCA) of all the leaves in L in tree T, that is, $lca_T(L)$ is the unique smallest upper bound of L under $\leq_T$. Given $x, y \in V(T)$, $x \rightarrow_T y$ denotes the unique path from x to y in T. $d_T(x,y)$ is denoted the number of edges on the path $x \rightarrow_T y$; note that if x = y then $d_T(x, y) = 0$. Throughout this work, the term tree refers to a rooted binary tree.

It will be convenient to consider the restriction of an LCT to a single species branch or to a subtree rooted at a species, in each case considering only the parts of the gene tree that evolve within the considered species (Mawhorter et al., 2019) [7]. Here, the locus will be denoted by either l_1 or l_2 as we take binary gene tree and species tree.

Given a species node s, let nodes(s) is set of gene nodes which will be mapped to s; bottoms(s) are the subset of nodes(s) that can be denoted by leaves and tops(s) = bottoms(pa(s)) if $s \neq rt(S)$ and tops(s) = {rt(G)}, here, pa(s) denotes the parent of s and rt(S) and rt(G) denote the root of species and gene trees, respectively. Here, we actually consider the tops(s) as the tree will be traversed as pre-order. As we know, the ILS are considered as "null events", so if the ILS event will be detected then the locus will be marked and will be counted that event either in duplication or transfer or in loss region.

In **Fig:3** LCT have been depicted briefly with the various evolutionary events. A locus present at the bottom of the species branch continuing at the same locus denoting the speciation event. The creation of a new locus along a gene branch, which is occurred when a gene node and its parents are mapped to different loci, this event is denoting the duplication event. If a locus is present at either the top of a species branch or created via a duplication being no longer present at the bottom of the species branch, this event is called loss event. If two or more lineages are being failed to coalesce; then this failure can be resulted in multiple lineages at speciation or duplication, this event actually denoted by deep coalescence.

2.1 Reconciliation and DTLI scenarios

DEFINITION 2.1 (DTLI-scenario). A DTLI-scenario for T and S is a seven-tuple $\langle L, M, \sum, \Delta, \Delta', \Theta, \Xi, \tau \rangle$, where L: Le(T)→ Le(S) represents the leaf-mapping from T to S, M : V(T)→V(S) maps each node of T to a node of S, the sets $\sum, \Delta, \Theta$ partition I(T) into speciation (or co-speciation), duplication and transfer nodes, respectively, Δ' is the duplication created for deep coalescence, Ξ is a subset of edges of T that represent transfer edges and $\tau: \Theta \rightarrow V(S)$ specifies the recipient for each transfer event, subject to the following constraints:

(1) if $t \in Le(T)$, then M(t) = L(t). This ensures that the mapping M is consistent with the leaf-mapping L.
(2) if $t \in I(T)$ and t' and t'' denote the children of t, then, (a) M(t) not less than M(t') and M(t) not less than M(t''). This imposes on
   (a) M the temporal constraints implied by S
   (b) At least one of M(t') and M(t") is a descendant of M(t). This implies that any internal node in G may represent at

most one transfer event.

(3) Given any edge (t,t') ϵ E(T), (t,t') ϵ Ξ if and only if M(t) and M(t') are incomparable. This determines the edges of T that are transfer edges.

(4) If t ϵ I(T) and t' and t" denote the children of t, then,

(a) t ϵ ∑ only if M(t) = lca(M(t'), M(t")) and M(t') and M(t") are incomparable, this state the conditions under which an internal node of T may represent a speciation.

(b) t ϵ Δ only if M(t)≥$_s$ lca(M(t'), M(t")), this state the conditions under which an internal node of T may represent a duplication.

(c) t ϵ Θ if and only if either (t,t') ϵ Ξ or (t,t") ϵ Ξ. this state the conditions under which an internal node of T may represent a transfer.

(d) if t ϵ Θ and (t,t') ϵ Ξ, then M(t) and τ(t) must be incomparable, and M(t') must be a descendant of τ(t), this specifies which species may be designated as the recipient species for any given transfer event.

(5) A coalescence event is, in fact, a deep coalescence, in which two or more lineages fail to coalesce; such failure can result in multiple lineages at speciation or duplications.

(a) if t ϵ tops(s), t' ϵ nodes(s), M(t) and M(t') are incomparable and if |d$_s$(M(t), M(t'))| > 1, then this induces a coalescence at speciation event with |d$_s$(M(t), M(t'))|-1 extra lineages. This is applicable for t'' node also.

(b) Given any edge (t_1,t_1') ϵ E(T), such that t_1 ϵ { r(T) U tops(s) U M(t_1)≥$_s$ lca(M(t_1')}, t_1' ϵ nodes(s) and at position l_1 at species s. For each duplication node d ϵ M(t_1)≥s lca(M(t_1')}, let N(s, l_1 , d) denote the set of gene lineages in species s at position l_1 contemporaneous with duplication d, where a lineage is contemporaneous with a different position say l_2 if it starts before and ends after the duplication node. Position l_1 can be created before l_2 vice versa. If | N(s, l_1, d) | > 1, then N(s, l_1 , d) induces a coalescence at duplication event with | N(s, l_1 , d) |-1 extra gene lineages. This is applicable for t_1'' node also.

DEFINITION 2.2 (Losses). Given a DTLI-scenario α=<L, M, ∑, Δ, Δ',Θ, Ξ, τ> for T and S, let t ϵ V(T) and {t', t"} = Ch(t)(Libeskind-Hadas et al., 2014) [5]. The number of losses Loss$_α$(t) at node t is defined to be

- |d$_s$(M(t), M(t'))-1| + |d$_s$(M(t), M(t"))-1|, if t ϵ ∑.
- d$_s$(M(t), M(t')) + d$_s$(M(t), M(t")), if t ϵ Δ.
- |d$_s$(M(t), M(t')) + d$_s$(M(t), M(t"))| > 1, if t ϵ Δ'
- d$_s$(M(t), M(t")) + d$_s$(τ(t), M(t')), if (t, t') ϵ Ξ

The total number of losses in the reconciliation corresponding to the DTLI-scenario α is defined to be Loss$_α$ = $\sum_{t \epsilon I(T)}$ Loss$_α$(t).

Assuming that, the speciation and Incomplete Lineage Sorting have zero cost, and let C$_Δ$, C$_Θ$ and C$_L$ denote the assigned positive costs for duplication, transfer and loss events, respectively.

DEFINITION 2.3 (Reconciliation cost of a DTLI-scenario). Given a DTLI-scenario α=<L, M, ∑, Δ, Δ', Θ, Ξ, τ> for T and S, the reconciliation cost associated with α is given by C$_Δ$. |Δ| + C$_{Δ'}$. |Δ'| + C$_Θ$. |Θ| + C$_L$. Loss$_α$ (Libeskind-Hadas et al., 2014) [5].

The traditional goal of DTLI-reconciliation is to find a most parsimonious reconciliation, i.e. a DTLI-scenario for G and S with minimum reconciliation cost. In this work as the previous work (Libeskind-Hadas et al., 2014) [5] we assume that the exact cost assignments, C$_Δ$, C$_{Δ'}$, C$_Θ$ and C$_L$ are unknown; therefore, we focus on the inferred event counts rather than the reconciliation cost itself.

The most important two terms are used in pareto-optimal algorithm- one is **Pareto-optimal vector (PV)** and another one is **Equivalent Region Partition (ERP).** In our newly developed algorithm we will modify this PV and ERP concept a little bit which will be explained in the next section.

3 ALGORITHM

Tofigh (2009) [13] was the first to adapt the dynamic programming algorithm for computing Pareto-optimal event count vectors, but the loss event was discarded to simply the computation. Later, Libeskind-Hadas et al., 2014 [5] developed the previous dp-algorithm accounting duplication, transfer and loss but discarded ILS event. Later, Mawhorter et al., 2019 [7], developed DLCparETO algorithm but not accounting transfer event. In developing an efficient algorithm for the PV problem, we will account all four evolutionary events and we will show that the space complexity will remain same as the previous algorithm that was developed by Libeskind-Hadas et al., 2014 [5].

The algorithm is as follows: (the description of this algorithm will be briefly discussed after the pseudo-code)

ALGORITHM(Updated Algorithm Considering Incomplete Lineage Sorting):

Algorithm : Pareto-Reconcile (T,S,L)

1: **for** each t ϵ V(T) and s ϵ V(S) **do**
2: Initialize P(t,s), P$_∑$(t,s), P$_Δ$(t,s), P$_{Δ'}$(t,s), P$_Θ$(t,s), in(t,s), inAlt(t,s) and out(t,s) to Ø
3: **for** each t ϵ Le(T) **do**
4: P(t,L(t)) = {<0,0,0>}, and, **for** each s ≥$_s$ L(t), in(t,s) = {<0,0,d$_s$(s,L(t))>} and inAlt(t,s) = {<0,0,0>}
5: **for** each t ϵ I(T) in post-order **do**
6: **for** each s ϵ V(S) in post-order **do**
7: Let {t',t''} = Ch$_T$(t)
8: **if** s ϵ Le(S) **then**

9: $P_{\sum}(t,s) = \{< \infty, \infty, \infty>\}$
10: $P_{\Delta}(t,s) = (P(t',s) \otimes P(t'',s)) + (\Delta,1)$
11: $P_{\Theta}(t,s) = ((in(t',s) \otimes out(t'',s)) \oplus (in(t'',s) \otimes out(t',s))) + (\Theta,1)$
12: $P(t,s) = P\sum(t,s) \oplus P\Delta(t,s) \oplus P\Theta(t,s)$
13: $in(t,s) = P(t,s)$
14: $inAlt(t,s) = P(t,s)$
15: **else**
16: Let $\{s',s''\} = Ch_s(s)$
17: $P_{\sum}(t,s) = (in(t',s') \otimes in(t'',s'')) \oplus (in(t'',s') \otimes in(t',s''))$
18: **for** each gene node $t \in V(T)$ in pre-order do
19: **if** $t_1 \in \{r(T)\} \cup tops(s)$ **then**
20: Set $pos(t_1) = l_1(t_1)$
21: $P_{\Delta}(t_1,s) = (in(t_1',s) \otimes in(t_1'',s)) + (\Delta,1)$
22: **else if** another duplication occur at t_2 **then**
23: Set $pos(t_2) = l_2(t_2)$
24: Label t_2 marked as deep coalesence
25: $P_{\Delta'}(t_2,s) = (in(t_2',s) \otimes in(t_2'',s)) + (\Delta',1)$
26: **else**
27: Set pos(t) = pos (p(t))
28: **if** $s \neq rt(S)$ **then** $P_{\Theta}(t,s) = ((in(t',s) \otimes out(t'',s)) \oplus (in(t'',s) \otimes out(t',s))) + (\Theta,1)$
29: $P(t,s) = P_{\sum}(t,s) \oplus P_{\Delta}(t,s) \oplus P_{\Theta}(t,s) \oplus P_{\Delta}'(t,s)$
30: $in(t,s) = P(t,s) \oplus (in(t,s') + (L,1)) \oplus (in(t,s'') + (L,1))$
31: $inAlt(t,s) = P(t,s) \oplus inAlt(t,s') \oplus inAlt(t,s'')$
32: **for** each $s \in I(S)$ in pre-order do
33: Let $\{s',s''\} = Ch_s(s)$
34: $out(t,s') = out(t,s) \oplus inAlt(t,s'')$ and
35: $out(t,s'') = out(t,s) \oplus inAlt(t,s')$
36: **Return** $\oplus_{s \in V(S)} P(rt(T),s)$

Brief explanation of the algorithm:

The notation used in this algorithm are shortly briefed in **Definition 2.1(mentioned above).**

Solving the PV problem:

Given any $t \in I(T)$ and $s \in V(S)$, let $P_{\sum}(t,s)$ denote the set of pareto-optimal event count vectors for reconciling T(t) with S such that t maps to s and $t \in \sum$. The terms $P_{\Delta}(t,s)$ and $P_{\Theta}(t,s)$ are defined similarly for $t \in \Delta$ and $t \in \Theta$, respectively. P(t,s) is the set of Pareto-optimal event count vectors for reconciling T(t) with S such that t maps to s.

$P_{\Delta'}(t,s)$ counts for duplication that occurred for deep coalescence.

Suppose, there are two sets of Pareto-optimal event count vectors A and B, the $A \oplus B$ means the set obtained by taking the union of the event count vectors in A and B and then selecting the subset of event count vectors that are pareto-optimal.

Similarly, $A \otimes B$ means the set obtained by first computing the Cartesian product of A and B then converting each resulting ordered pair into a single event count vector by adding the two vectors of the ordered pair and finally taking only the subset of pareto-optimal event count vectors.

Given any two sets, A and B, of Pareto-optimal event count vectors, the set $A \oplus B$ can be computed in $O(m^4)$ time. [Libeskind-Hadas et al., 2014] [5].

Given any two sets, A and B, of Pareto-optimal event count vectors, the set $A \otimes B$ can be computed in $O(m^{4l}\log m)$ time. [Libeskind-Hadas et al., 2014] [5].

Here, the dynamic programming table for P(t,s) is initialized and computed as shown below:

$P(t,s) = <0,0,0>$ if $t \in Le(T)$ and $s = M(t)$

$P(t,s) = <\infty, \infty, \infty>$ if $t \in Le(T)$ and $s \neq M(t)$

$P(t,s) = P_{\sum}(t,s) \oplus P_{\Delta}(t,s) \oplus P_{\Theta}(t,s) \oplus P_{\Delta}'(t,s)$ otherwise.

Again, here,

$in(t,s) = \oplus_{s \in V(S(s))}(P(t,x) + (L, d_s(s,x)))$,

$out(t,s) = \oplus_{s \in V(S) \text{ incomparable to } s} P(t,x)$ and

$inAlt(t,s) = \oplus_{s \in V(S(s))} P(t,x)$

This algorithm computes $P_{\sum}(t,s)$, $P_{\Delta}(t,s)$ and $P_{\Theta}(t,s)$ by performing a nested post-order traversal of T ans S. $P_{\Delta'}(t,s)$ is computed by performing pre-order traversal of T and S. The values in(. , .) , out(. , .) and inAlt(. , .) help to reuse previously computed information to efficiently compute the values $P_{\sum}(t,s)$, $P_{\Delta}(t,s)$, $P_{\Delta'}(t,s)$ and $P_{\Theta}(t,s)$ at each step.

Here in this algorithm:

Line 1 to 2 : initialize $P_{\sum}(t,s)$, $P_{\Delta}(t,s)$, $P_{\Delta'}(t,s)$, $P_{\Theta}(t,s)$, in(t,s), inAlt(t,s) and out(t,s) to null.

Line 3 to 4: Mapping gene leaf node to species and compute P(t,s), in(t,s) and inAlt(t,s)

Line 5 to 17: Considering the gene trees internal node and nodes of species tree and calculating the values or speciation,

duplication and transfer

Line 18 to 27: By traversing the gene tree in pre order if the duplication is occurred due to deep coalescence then that position will be marked and that event will be counted as $P_{\Delta'}(t,s)$. t_1 and t_2 are considered the child of t and l_1 and l_2 are marked for normal duplication or the duplication is occurred for deep coalescence.

Line 29 to 30: Computing every events including Loss calculation.

Complexity:

Based on the pseudo-code for Algorithm Pareto-Reconcile, and on the previous three lemmas which was developed in Libeskind-Hadas et al., 2014 [5] algorithm, the next two theorems can be followed easily. In that paper three lemma had been proved.

i. Lemma-1: The cardinality of any set of Pareto-optimal event count vectors can be no longer than $\Theta(m^2)$.
ii. Lemma-2: Given any two sets, A and B, of Pareto-optimal event count vectors, the set $A \oplus B$ can be computed in $O(m^4)$.
iii. Lemma-3: Given any two sets, A and B, of Pareto-optimal event count vectors, the set $A \otimes B$ can be computed in $O(m^4 \log m)$.

So, the derived theorem will be:

a. Theorem -1: Algorithm Pareto-Reconcile correctly solves the PV problem.

b.Theorem-2: The total time complexity of Algorithm Pareto-Reconcile is $O(m^6 n \log m)$ and the space complexity will remain same as the Libeskind-Hadas et al., 2014 [5], is $O(m^3 n)$. The space complexity will remain same because if the incomplete lineage sorting event is found then this will be result either duplication or speciation, so the region will not be increased.

Equivalent Region Partition

Here, the region partition concept is same as the previous developed algorithm by Libeskind-Hadas et al., 2014[5] because Speciation and Incomplete Lineage Sorting events have cost 0 (zero), so the region will not be changed.

c. Theorem 3: Given a set of Pareto-optimal event count vectors, the corresponding regions can be found in time $O(m^4 \log m)$.

4 CONCLUSIONS

Through this research, it has been shown that how to cost DTL events considering ILS event. It gives a fixed parameter tractable algorithm which calculates the most parsimonious duplication, horizontal gene transfer, loss and incomplete lineage sorting reconciliation. Comparing with the previous pareto-optimal algorithm, the space complexity will remain same after considering the incomplete lineage sorting event. This research is based on binary gene tree and binary species tree, but this algorithm can be considered non binary gene tree and non-binary species tree. Accounting non-binary gene tree and non-binary species tree the outcome will be more versatile and efficient.

REFERENCES

[1] Andersson, J.O.: Horizontal gene transfer between microbial eukaryotes. Horizontal Gene Transfer pp. 473-487 (2009)
[2] Avise, J.C., Shapira, J., Daniel, S.W., Aquadro, C.F., Lansman, R.A.: Mitochondrial dna differentiation during the speciation process in peromyscus. Molecular Biology and Evolution 1(1), 38-56 (1983)
[3] Charleston, M.: Jungles: a new solution to the host/parasite phylogeny reconciliation problem. Mathematical biosciences 149(2), 191-223 (1998)
[4] Edwards, S.V.: Is a new and general theory of molecular systematics emerging? Evolution: International Journal of Organic Evolution 63(1), 1-19 (2009)
[5] Libeskind-Hadas, R., Wu, Y.C., Bansal, M.S., Kellis, M.: Pareto-optimal phylogenetic tree reconciliation. Bioinformatics 30(12), i87-i95 (2014)
[6] Maddison, W.P.: Gene Trees in Species Trees. Systematic Biology 46(3), 523-536 (09 1997). https://doi.org/10.1093/sysbio/46.3.523, https://doi.org/10.1093/sysbio/46.3.523
[7] Mawhorter, R., Liu, N., Libeskind-Hadas, R., Wu, Y.C.: Inferring pareto-optimal reconciliations across multiple event costs under the duplication-loss-coalescence model. BMC bioinformatics 20(20), 1-13 (2019)
[8] Milinkovitch, M.C., Helaers, R., Depiereux, E., Tzika, A.C., Gabaldon, T.: " considerations for the inclusion of 2x mammalian genomes in phylogenetic analyses (a response to vilella et al.'s correspondence). Genome Biology 12, 401 (2010)
[9] Neigel, J.E.: Phylogenetic relationships of mitochondrial dna under various models of speciation. Evolutionary processes and theory (1986)
[10] Serres, M.H., Kerr, A.R., McCormack, T.J., Riley, M.: Evolution by leaps: gene duplication in bacteria. Biology direct 4(1), 1-17 (2009)
[11] Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., Durand, D.: Inferring duplications, losses, transfers and incomplete lineage sorting with non-binary species trees. Bioinformatics 28(18), i409-i415 (2012)
[12] Tajima, F.: Evolutionary relationship of dna sequences in finite populations. Genetics 105(2), 437{460 (1983)
[13] Tofigh, A.: Using trees to capture reticulate evolution: lateral gene transfers and cancer progression. Ph.D. thesis, KTH (2009)
[14] Zhaxybayeva, O., Doolittle, W.F.: Lateral gene transfer. Current Biology 21(7),R242-R246 (2011)

