Unique perfect phylogeny is NP-hard

Michel Habib[1] and Juraj Stacho[2]

[1] LIAFA – CNRS and Université Paris Diderot – Paris VII,
Case 7014, 75205 Paris Cedex 13, France (habib@liafa.jussieu.fr)
[2] Caesarea Rothschild Institute, University of Haifa
Mt. Carmel, 31905 Haifa, Israel (stacho@cs.toronto.edu)

Abstract. We answer, in the affirmative, the following question proposed by Mike Steel as a \$100 challenge: *"Is the following problem NP-hard? Given a ternary*[†] *phylogenetic X-tree $\mathcal{T}$ and a collection $\mathcal{Q}$ of quartet subtrees on X, is $\mathcal{T}$ the only tree that displays $\mathcal{Q}$?"* [28, 29] As a particular consequence of this, we show that the unique chordal sandwich problem is also NP-hard.

1 Introduction

One of the major efforts in molecular biology has been the computation of phylogenetic trees, or *phylogenies*, which describe the evolution of a set of species from a common ancestor. A phylogenetic tree for a set of species is a tree in which the leaves represent the species from the set and the internal nodes represent the (hypothetical) ancestral species. One standard model for describing the species is in terms of *characters*, where a character is an equivalence relation on the species set, partitioning it into different *character states*. In this model, we also assign character states to the (hypothetical) ancestral species. The desired property is that for each state of each character, the set of nodes in the tree having that character state forms a connected subgraph. When a phylogeny has this property, we say it is *perfect*. The Perfect Phylogeny problem [18] then asks *for a given set of characters defining a species set, does there exist a perfect phylogeny?* Note that we allow that states of some characters are unknown for some species; we call such characters *partial*, otherwise we speak of *full* characters. This approach to constructing phylogenies has been studied since the 1960s [7, 23–25, 32] and was given a precise mathematical formulation in the 1970s [10–13]. In particular, Buneman [6] showed that the Perfect Phylogeny problem reduces to a specific graph-theoretic problem, the problem of finding a chordal completion of a graph that respects a prescribed colouring. In fact, the two problems are polynomially equivalent [21]. Thus, using this formulation, it has been proved that the Perfect Phylogeny problem is NP-hard in [3] and independently in [30]. These two results rely on the fact that the input may contain partial characters. In fact, the characters in these constructions only have two states. If we insist on full characters, the situation is different as for any fixed number r of character

[†] Some formulations of this question use the term "binary", as in "rooted binary tree".

states, the problem can be solved in time polynomial [1] in the size of the input (and exponential in r). In particular, for $r = 2$ (or $r = 3$), the solution exists if and only if it exists for every pair (or triple) of characters [13, 22]. Also, when the number of characters is k (even if there are partial characters), the complexity [26] is polynomial in the number of species (and exponential in k).

Another common formulation of this problem is the problem of a *consensus tree* [9, 17, 30], where a collection of subtrees with labelled leaves is given (for instance, the leaves correspond to species of a partial character). Here, we ask for a (phylogenetic) tree such that each of the input subtrees can be obtained by contracting edges of the tree (we say that the tree *displays* the subtree). The problem does not change [28] if we only allow particular input subtrees, the so-called *quartet trees*, which have exactly six vertices and four leaves. This follows from the fact that every ternary phylogenetic tree can be uniquely described by a collection of quartet trees [28]. However, a collection of quartet trees does not necessarily uniquely describe a ternary phylogenetic tree.

This leads to a natural question: *what is the complexity of deciding whether or not a collection of quartet trees uniquely describes a (ternary) phylogenetic tree?* This question was first posed in 1992 in [30], later conjectured to be NP-hard [28] and listed on M. Steel's personal webpage [29] where he offers \$100 for the first proof of NP-hardness.

In this paper, we are the first to answer this question by showing that the problem is indeed NP-hard. That is, we prove the following theorem.

Theorem 1. *It is NP-hard to determine, given a ternary phylogenetic X-tree $\mathcal{T}$ and a collection $\mathcal{Q}$ of quartet subtrees on X, whether or not $\mathcal{T}$ is the only phylogenetic tree that displays $\mathcal{Q}$.*

(We note that an alternative proof of this theorem recently appeared on arxiv [4]. The proof uses different techniques and extends the hardness result of [30].)

In light of this, we note that there are special cases of the problem that are known to be solvable in polynomial time. For instance, this is so if the collection $\mathcal{Q}$ contains a subcollection $\mathcal{Q}'$ with the same set $\mathcal{L}$ of labels of leaves and with $|\mathcal{Q}'| = |\mathcal{L}|-3$. However, finding such a subcollection is known to be NP-complete. For these and similar results, we refer the reader to [2].

We prove Theorem 1 by describing a polynomial-time reduction from the uniqueness problem for ONE-IN-THREE-3SAT, which is NP-hard by [20].†

Theorem 2. [20] *It is NP-hard to decide, given an instance I to* ONE-IN-THREE-3SAT*, and a truth assignment σ that satisfies I, whether or not σ is the unique satisfying truth assignment for I.*

Our construction in the reduction is essentially a modification of the construction of [3] which proves NP-hardness of the Perfect Phylogeny problem.

† We extract this from [20] by encoding the problem as the relation $\{001, 010, 100\}$. We check that this relation is not: 0-valid, 1-valid, Horn, anti-Horn, affine, 2SAT, or complementive. Then the uniqueness of the satisfiability problem corresponding to this relation is $CoNP$-hard by [20] and thus NP-hard (assuming Turing reductions).

Recall that the construction of [3] produces instances $\mathcal{Q}$ that have a perfect phylogeny if and only if a particular boolean formula Φ is satisfiable. We immediately observed that these instances $\mathcal{Q}$ have, in addition, the property that Φ has a unique satisfying assignment if and only if there is a unique minimal restricted chordal completion of the partial partition intersection graph of $\mathcal{Q}$ (for definitions see Section 2). This is precisely one of the two necessary conditions for uniqueness of perfect phylogeny as proved by Semple and Steel in [27] (see Theorem 4). Thus by modifying the construction of [3] to also satisfy the other condition of uniqueness of [27], we obtained the construction that we present in this paper. Note that, however, unlike [3] which uses 3SAT, we had to use a different NP-hard problem in order for the construction to work correctly. Also, to prove that the construction is correct, we employ a variant of the characterization of [27] that uses the more general chordal sandwich problem [15] instead of the restricted chordal completion problem (see Theorem 7). In fact, by way of Theorems 5 and 6, we establish a direct connection between the problem of perfect phylogeny and the chordal sandwich problem, which apparently has not been yet observed. (Note that the connection to the (restricted) chordal completion problem of coloured graphs as mentioned above [6, 21] is a special case of this.)

Finally, as a corollary, we obtain the following result using [8].

Corollary 1. *The Unique chordal sandwich problem is NP-hard. Counting the number of minimal chordal sandwiches is #P-complete.*

The paper is structured as follows. In Section 2, we introduce definitions and some preliminary work. In Sections 3 and 4, we present our hardness reduction, first informally and then formally. Then we sketch the proof of one of the main theorems (Theorem 8) in Section 5, and conclude with some open questions.

2 Preliminaries

We mostly follow the terminology of [27, 28] and graph-theoretical notions of [31].

Let X be a non-empty set. An *X-tree* is a pair (T, ϕ) where T is tree and $\phi : X \to V(T)$ is a mapping such that $\phi^{-1}(v) \neq \emptyset$ for all vertices $v \in V(T)$ of degree at most two. An X-tree (T, ϕ) is *ternary* if all internal vertices of T have degree three. Two X-trees (T_1, ϕ_1), (T_2, ϕ_2) are *isomorphic* if there exists an isomorphism $\psi : V(T_1) \to V(T_2)$ between T_1 and T_2 that satisfies $\phi_2 = \psi \circ \phi_1$.

An X-tree (T, ϕ) is a *phylogenetic X-tree* (or a *free X-free* in [27]) if ϕ is a bijection between X and the set of leaves of T. A *partial partition* of X is a partition of a non-empty subset of X into at least two sets. If A_1, A_2, ..., A_t are these sets, we call them *cells* of this partition, and denote the partition $A_1|A_2|\ldots|A_t$. If $t = 2$, we call the partition a *partial split*. A partial split $A_1|A_2$ is trivial if $|A_1| = 1$ or $|A_2| = 1$. A *quartet tree* is a ternary phylogenetic tree with a label set of size four, that is, a ternary tree $\mathcal{T}$ with 6 vertices, 4 leaves labelled a, b, c, d, and with only one non-trivial partial split $\{a, b\}|\{c, d\}$ that it displays. Note that such a tree is unambiguously defined by this partial split. Thus, in the subsequent text, we identify the quartet tree $\mathcal{T}$ with the partial split $\{a, b\}|\{c, d\}$, that is, we say that $\{a, b\}|\{c, d\}$ is both a quartet tree and a partial split.

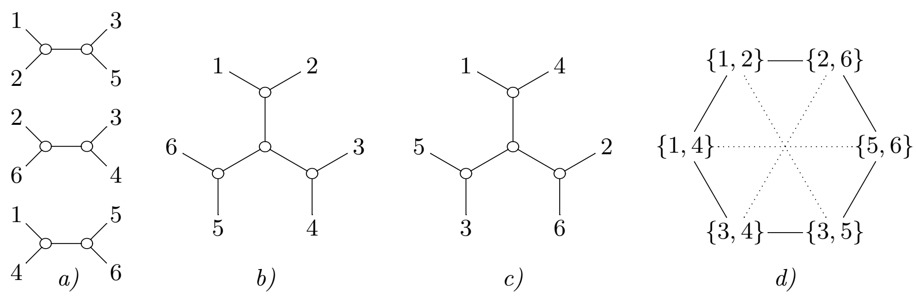

Fig. 1. *a)* quartet trees $\mathcal{Q}$, *b)* and *c)* two X-trees displaying $\mathcal{Q}$ and distinguished by $\mathcal{Q}$, *d)* the graph $\text{int}^*(\mathcal{Q})$; the dotted lines represent the edges of $\text{forb}(\mathcal{Q})$.

Let $\mathcal{T} = (T, \phi)$ be an X-tree, and let $\pi = A_1|A_2|\dots|A_t$ be a partial partition of X. Let $F \subseteq E(T)$ be a set of edges of T. We say that F *displays* π in $\mathcal{T}$ if for all distinct $i, j \in \{1 \dots t\}$, the sets $\phi(A_i)$ and $\phi(A_j)$ are subsets of the vertex sets of different connected components of $T - F$. We say that $\mathcal{T}$ *displays* π if there is a set of edges that displays π in $\mathcal{T}$. Further, an edge e of T is *distinguished* by π if every set of edges that displays π in $\mathcal{T}$ contains e.

Let $\mathcal{Q}$ be a collection of partial partitions of X. An X-tree $\mathcal{T}$ *displays* $\mathcal{Q}$ if it displays every partial partition in $\mathcal{Q}$. An X-tree $\mathcal{T} = (T, \phi)$ is *distinguished* by $\mathcal{Q}$ if every internal edge of T is distinguished by some partial partition in $\mathcal{Q}$; we also say that $\mathcal{Q}$ *distinguishes* $\mathcal{T}$. The set $\mathcal{Q}$ *defines* $\mathcal{T}$ if $\mathcal{T}$ displays $\mathcal{Q}$, and all other X-trees that display $\mathcal{Q}$ are isomorphic to $\mathcal{T}$. Note that if $\mathcal{Q}$ defines $\mathcal{T}$, then $\mathcal{T}$ is necessarily a ternary phylogenetic X-tree, since otherwise "resolving" any vertex either of degree four or more, or with multiple labels results in a non-isomorphic X-tree that also displays $\mathcal{Q}$ (also, see Proposition 2.6 in [27]). See Figure 1 for an illustration of these concepts.

The *partial partition intersection graph* of $\mathcal{Q}$, denoted by $\text{int}(\mathcal{Q})$, is a graph whose vertex set is $\{(A, \pi) \mid \text{where } A \text{ is a cell of } \pi \in \mathcal{Q}\}$ and two vertices (A, π), (A', π') are adjacent just if the intersection of A and A' is non-empty.

A graph is *chordal* if it contains no induced cycle of length four or more. A *chordal completion* of a graph $G = (V, E)$ is a chordal graph $G' = (V, E')$ with $E \subseteq E'$. A *restricted chordal completion* of $\text{int}(\mathcal{Q})$ is a chordal completion G' of $\text{int}(\mathcal{Q})$ with the property that if A_1,A_2 are cells of $\pi \in \mathcal{Q}$, then (A_1, π) is not adjacent to (A_2, π) in G'. A restricted chordal completion G' of $\text{int}(\mathcal{Q})$ is *minimal* if no proper subgraph of G' is a restricted chordal completion of $\text{int}(\mathcal{Q})$.

The problem of perfect phylogeny is equivalent to the problem of determining the existence of an X-tree that display the given collection $\mathcal{Q}$ of partial partitions. In [6], it was given the following graph-theoretical characterization.

Theorem 3. [6, 28, 30] *Let $\mathcal{Q}$ be a set of partial partitions of a set X. Then there exists an X-tree that displays $\mathcal{Q}$ if and only if there exists a restricted chordal completion of* $\text{int}(\mathcal{Q})$.

Of course, the X-tree in the above theorem might not be unique. For the problem of uniqueness, Semple and Steel [27, 28] describe necessary and sufficient conditions for when a collection of partial partitions defines an X-tree.

Theorem 4. [27] *Let $\mathcal{Q}$ be a collection of partial partitions of a set X. Let $\mathcal{T}$ be a ternary phylogenetic X-tree. Then $\mathcal{Q}$ defines $\mathcal{T}$ if and only if:*

(i) $\mathcal{T}$ displays $\mathcal{Q}$ and is distinguished by $\mathcal{Q}$, and
(ii) there is a unique minimal restricted chordal completion of $\mathrm{int}(\mathcal{Q})$.

In order to simplify our proof of Theorem 1, we now describe a variant of the above theorem that, instead, deals with the notion of chordal sandwich [15].

Let $G = (V, E)$ and $H = (V, F)$ be two graphs on the same set of vertices with $E \cap F = \emptyset$. A *chordal sandwich* of (G,H) is a chordal graph $G' = (V, E')$ with $E \subseteq E'$ and $E' \cap F = \emptyset$.[†] A chordal sandwich G' of (G,H) is *minimal* if no proper subgraph of G' is a chordal sandwich of (G,H).

The *cell intersection graph* of $\mathcal{Q}$, denoted by $\mathrm{int}^*(\mathcal{Q})$, is the graph whose vertex set is $\{A \mid \text{where } A \text{ is a cell of } \pi \in \mathcal{Q}\}$ and two vertices A, A' are adjacent just if the intersection of A and A' is non-empty. Let $\mathrm{forb}(\mathcal{Q})$ denote the graph whose vertex set is that of $\mathrm{int}^*(\mathcal{Q})$ in which there is an edge between A and A' just if A,A' are cells of some $\pi \in \mathcal{Q}$. See Figure 1d for an example.

The correspondence between the partial partition intersection graph and the cell intersection graph is captured by the following theorem.

Theorem 5. *Let $\mathcal{Q}$ be a collection of partial partitions of a set X. There is a one-to-one mapping between the minimal restricted chordal completions of* $\mathrm{int}(\mathcal{Q})$ *and the minimal chordal sandwiches of* $(\mathrm{int}^*(\mathcal{Q}), \mathrm{forb}(\mathcal{Q}))$.

The proof of this theorem follows easily from the following lemma.

Lemma 1. *Let G be a graph, and let G^+ be a graph obtained from G by substituting complete graphs*[‡] *for the vertices of G. Then there is a one-to-one mapping between minimal chordal completions of G and G^+.*

This combined with Theorem 3 yields that there is a phylogenetic X-tree that displays $\mathcal{Q}$ if and only if there exists a chordal sandwich of $(\mathrm{int}^*(\mathcal{Q}), \mathrm{forb}(\mathcal{Q}))$. Conversely, we can express every instance to the chordal sandwich problem as a corresponding instance to the problem of perfect phylogeny as follows.

Theorem 6. *Let (G, H) be an instance to the chordal sandwich problem. Then there is a collection $\mathcal{Q}$ of partial splits such that there is a one-to-one mapping between the minimal chordal sandwiches of (G, H) and the minimal restricted chordal completions of* $\mathrm{int}(\mathcal{Q})$. *In particular, there exists a chordal sandwich for (G, H) if and only if there exists a phylogenetic tree that displays $\mathcal{Q}$.*

Proof. (Sketch) Without loss of generality, we may assume that each connected component of G has at least three vertices. As usual, $G = (V, E)$ and $H = (V, F)$ where $E \cap F = \emptyset$. The collection $\mathcal{Q}$ satisfying the claim is defined as follows: for every edge $xy \in F$, we construct the partial split $D_x|D_y$, where D_x are the edges of E incident to x, and D_y are the edges of E incident to y. $\square$

As a corollary, we obtain the following desired characterization.

[†] we say that E are the *forced* edges and F are the *forbidden* edges.
[‡] replacing v by a complete graph K and adding edges $\{ux \mid x \in V(K) \wedge uv \in E(G)\}$

Theorem 7. *Let $\mathcal{Q}$ be a collection of partial partitions of a set X. Let $\mathcal{T}$ be a ternary phylogenetic X-tree. Then $\mathcal{Q}$ defines $\mathcal{T}$ if and only if:*

(i) $\mathcal{T}$ displays $\mathcal{Q}$ and is distinguished by $\mathcal{Q}$, and
(ii) there is a unique minimal chordal sandwich of $\Big(\mathrm{int}^*(\mathcal{Q}), \mathrm{forb}(\mathcal{Q})\Big)$.

The main technical advantage of this theorem over Theorem 4 is that it is less restrictive; it allows us to construct instances with arbitrary sets of forbidden edges rather than just with forbidden edges between vertices of the same colour. This makes our proof of Theorem 1 much simpler and more manageable.

3 Construction

Consider an instance I to ONE-IN-THREE-3SAT. The instance I consists of n *variables* $v_1, \ldots, v_n$ and m *clauses* $\mathcal{C}_1, \ldots, \mathcal{C}_m$ each of which is a disjunction of exactly three *literals* (i.e., variables v_i or their negations $\overline{v_i}$).

To simplify the presentation, we shall denote literals by capital letters X, Y, etc., and indicate their negations by $\overline{X}$, $\overline{Y}$, etc. (For instance, if $X = v_i$ then $\overline{X} = \overline{v_i}$, and if $X = \overline{v_i}$ then $\overline{X} = v_i$.)

By standard arguments, we may assume that no variable appears twice in the same clause. First, we discuss how to find a collection $\mathcal{Q}_I$ of quartet trees arising from the instance I that satisfies the following theorem.

Theorem 8. *There is a one-to-one mapping between the satisfying assignments of the instance I and the minimal chordal sandwiches of* $(\mathrm{int}^*(\mathcal{Q}_I), \mathrm{forb}(\mathcal{Q}_I))$.

Before describing the collection $\mathcal{Q}_I$, let us briefly review the construction from [3] that proves NP-hardness of the Perfect Phylogeny problem. For convenience, we describe it in terms of the chordal sandwich problem whose input is a graph with (forced) edges and forbidden edges. In the construction from [3], one similarly considers a collection $\mathcal{C}_1, \ldots, \mathcal{C}_m$ of 3-literal clauses, and treats it as an instance of 3-SATISFIABILITY. From this instance, one constructs a graph where each variable v_i corresponds to two *shoulders* S_{v_i} and $S_{\overline{v_i}}$, and where each literal W in a clause $\mathcal{C}_j$ corresponds to a pair of *knees* K^j_W and $K^j_{\overline{W}}$. In addition, there are two special vertices *the head* H and *the foot* F. All shoulders are adjacent to the head while all knees are adjacent to the foot. Further, if v_i occurs in the clause $\mathcal{C}_j$ (positively or negatively), then the vertices H, S_{v_i}, $K^j_{\overline{v_i}}$, F, $K^j_{v_i}$, $S_{\overline{v_i}}$ form an induced 6-cycle (see Figure 2a). Also, if $\mathcal{C}_j = X \vee Y \vee Z$, then the vertices K^j_X, K^j_Y, K^j_Z induce a triangle with pendant edges $K^j_X K^j_{\overline{Y}}$, $K^j_Y K^j_{\overline{Z}}$, and $K^j_Z K^j_{\overline{X}}$ (we call this the *clause gadget*, see Figure 2b).

Finally, the edge between H and F is forbidden in the desired chordal sandwich, and so is the edge between S_{v_i} and $S_{\overline{v_i}}$, and between $K^j_{v_i}$ and $K^j_{\overline{v_i}}$ for all meaningful indices i and j (the dotted edges in Figure 2).

The main idea of this construction is that each of the 6-cycles allows only two possible chordal sandwiches: either the path $H, K^j_{v_i}, S_{v_i}, F$ is added, or the path $H, K^j_{\overline{v_i}}, S_{\overline{v_i}}, F$ is added (the authors of [3] call this path the "Mark of Zorro"). These two choices correspond to assigning v_i the value *true* or *false*, respectively,

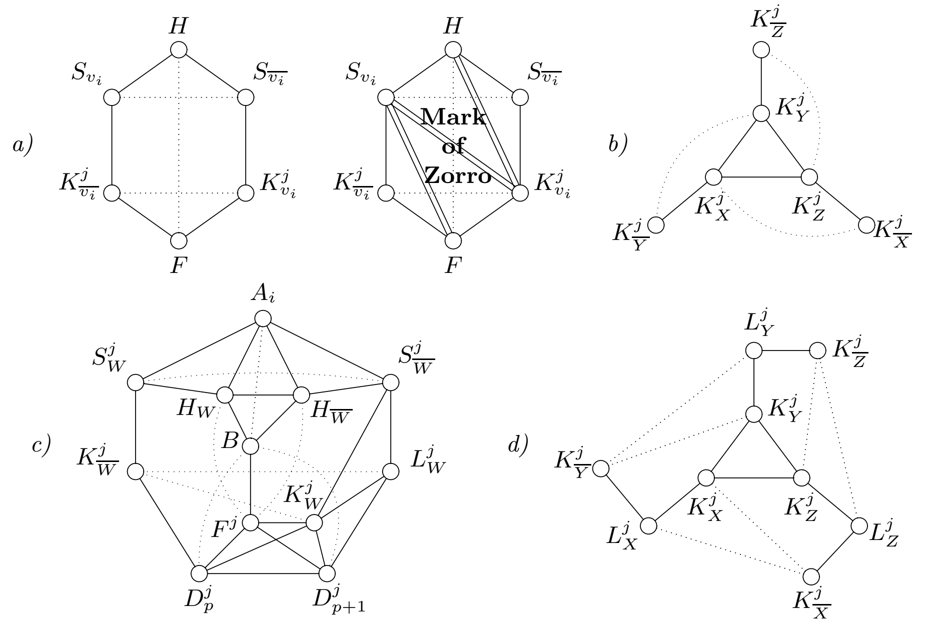

Fig. 2. *a)* and *b)* configurations from [3], *c)* and *d)* configurations from our construction (note that in *c)* the literal W is either v_i or $\overline{v_i}$, and is the p-th literal of the clause $\mathcal{C}_j$)

and the construction ensures that this choice is consistent over all clauses. This only produces satisfying assignments to 3-SATISFIABILITY, since we notice that no chordal sandwich adds a triangle on $K^j_{\overline{X}}, K^j_{\overline{Y}}, K^j_{\overline{Z}}$.

We can try to use this construction to prove Theorem 1. We immediately observe that the truth assignments satisfying the clauses $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are in one-to-one correspondence with the minimal chordal sandwiches of the above graph G. This is a little technical to prove. To do this, one first observes that different assignments add a different mark of Zorro to at least one 6-cycle. For the converse, one needs to find out which edges are forced in the sandwich after the marks of Zorro are added according to a satisfying assignment. It turns out that these edges yield a chordal sandwich, and thus a minimal chordal sandwich.

From G, using Theorems 5 and 6, one can construct a collection $\mathcal{Q}$ of partial splits (phylogenetic trees) such that the satisfying assignments of the clauses $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are in one-to-one correspondence with the minimal chordal sandwiches of $(\mathrm{int}^*(\mathcal{Q}), \mathrm{forb}(\mathcal{Q}))$. In particular, this collection $\mathcal{Q}$ satisfies the condition (ii) of Theorem 7 if and only if the clauses $\mathcal{C}_1, \ldots, \mathcal{C}_m$ have a unique satisfying assignment. Since this is NP-hard to determine [20], it would seem like we almost have a proof of Theorem 1. Unfortunately, we are missing a crucial piece which is the phylogenetic tree $\mathcal{T}$ satisfying the condition (i) of Theorem 7 for the collection $\mathcal{Q}$. A straightforward construction of such a tree based on [27] yields a phylogenetic tree that is distinguished by $\mathcal{Q}$, but whose internal nodes may have degree higher than three. If we try to fix this (by "resolving" the high-degree nodes in order to get a ternary tree), the resulting tree may no longer be distinguished by $\mathcal{Q}$. Moreover, the collection $\mathcal{Q}$ may not consist of quartet trees only. For all these reasons, we need to modify the construction of G.

First, we discuss how to modify G so that it corresponds to a collection of quartet trees. To do this, we must ensure that the neighbourhood of each vertex consists of two cliques (with possibly edges between them). We construct a new graph G_I by modifying G as follows. Instead of one head H, we now have, for each variable v_i, two *heads* H_{v_i}, $H_{\overline{v_i}}$, and an *auxiliary head* A_i. For a literal W in the clause $\mathcal{C}_j$, we now have two *shoulders* S^j_W and $S^j_{\overline{W}}$, and, as before, two *knees* K^j_W and $K^j_{\overline{W}}$, but also an additional *auxiliary knee* L^j_W. Further, for each clause $\mathcal{C}_j$, we have a *foot* F^j and three *auxiliary feet* D^j_1, D^j_2, and D^j_3. Finally, we have one additional vertex B known as *the backbone*. The resulting modifications to the 6-cycles and the clause gadgets can be seen in Figures 2c and 2d. (The forbidden edges are again indicated by dotted lines.) Note that, unlike in the case of G, this is not a complete description of G_I as we need to add some additional (forced) edges and forbidden edges not shown in these diagrams in order to make the reduction work. This is rather technical and we omit this for brevity.

From the construction, we conclude that, just like in G, the "6-cycles" of G_I (Figure 2c) admit only two possible kinds of sandwiches, and this is consistent over different clauses. However, unlike in G, the chordal sandwiches of G_I no longer correspond to satisfying assignments of 3-SATISFIABILITY but rather to satisfying assignments of ONE-IN-THREE-3-SAT. Fortunately, this problem is also NP-hard as is its uniqueness variant as previously discussed (see Theorem 2).

Now, from G_I, we construct a collection $\mathcal{Q}_I$ of quartet trees. To do this, we cannot just use Theorem 6 as before, since this may create partial partitions that do not correspond to quartet trees. Moreover, even if we use [28] to replace these partitions by an equivalent collection of quartet trees, this process may not preserve the number of solutions. We need a more careful construction.

We recall that the each vertex v of G_I belongs to two cliques that completely cover its neighbourhood; we assign greek letters to these two cliques (to distinguish them from vertices), and associate them with v.

In particular, we use the following symbols: α_W, β^j_W, $\gamma^j_1, \gamma^j_2, \gamma^j_3$, λ^j, δ, μ where W is a literal and $j \in \{1 \dots m\}$. They define specific cliques of G_I as follows. The letter α_W defines the clique of G_I consisting of all heads and shoulders of W. The letter β^j_W corresponds to the clique formed by the shoulder S^j_W and the knees $K^j_{\overline{W}}$, $L^j_{\overline{W}}$ (if exists). Further, λ^j yields a clique on F^j, D^j_1, D^j_2, D^j_3, K^j_X, K^j_Y, K^j_Z where $\mathcal{C}_j = X \vee Y \vee Z$, while the clique for γ^j_p where $p \in \{1,2,3\}$ is formed by D^j_p, $K^j_{\overline{W}}$, L^j_U where W and U are the p-th and $(p-1)$-th (modulo 3) literals of $\mathcal{C}_j$. Finally, δ corresponds to the clique containing B and all heads H_W whereas μ correspond to the clique with B and all feet F^j.

From this, we construct the collection $\mathcal{Q}_I$ by considering every forbidden edge uv of G_I and by constructing a partial partition with two cells in which one cell is the set of cliques assigned to u and the other is the set of cliques assigned to v. Since we assign to each vertex of G_I exactly two cliques, this yields partitions corresponding to quartet trees. For instance, in Figure 2d, we have a forbidden edge $K^j_X K^j_{\overline{X}}$ where K^j_X is assigned cliques $\beta^j_{\overline{X}}, \lambda^j$ and $K^j_{\overline{X}}$ is assigned β^j_X, γ^j_1. This yields a quartet tree $\{\beta^j_{\overline{X}}, \lambda^j\}|\{\beta^j_X, \gamma^j_1\}$. The complete definition of $\mathcal{Q}_I$ can be found in Section 4. Finally, since by construction every vertex of G_I is incident to at least one forbidden edge, we conclude that $G_I = \text{int}^*(\mathcal{Q}_I)$.

This completes the overview of the proof of Theorem 8. Its actual proof is quite technical and involved, but it is along the same lines as the uniqueness property we discussed for G, i.e., one describes the edges forced by an assignment and proves that this yields a chordal sandwich. (We sketch this in Section 5.)

To complete the result, we need to explain how to construct a phylogenetic tree corresponding to a satisfying assignment for $\mathcal{C}_1, \ldots, \mathcal{C}_m$ (as an instance of ONE-IN-THREE-3SAT) and show that it displays and is distinguished by the trees in $\mathcal{Q}_I$. Instead of giving a formal definition here, we discuss a small example. (The complete description is rather technical and is presented in Section 4.)

The example instance I^+ consists of four variables v_1, v_2, v_3, v_4 and three clauses $\mathcal{C}_1 = v_1 \vee v_2 \vee v_3$, $\mathcal{C}_2 = \overline{v_1} \vee v_2 \vee v_4$, and $\mathcal{C}_3 = v_3 \vee \overline{v_2} \vee \overline{v_4}$. The unique satisfying assignment assigns true to v_1, v_4 and false to v_2, v_3. The corresponding phylogenetic tree $\mathcal{T} = (T, \phi)$ is shown in Figure 3.

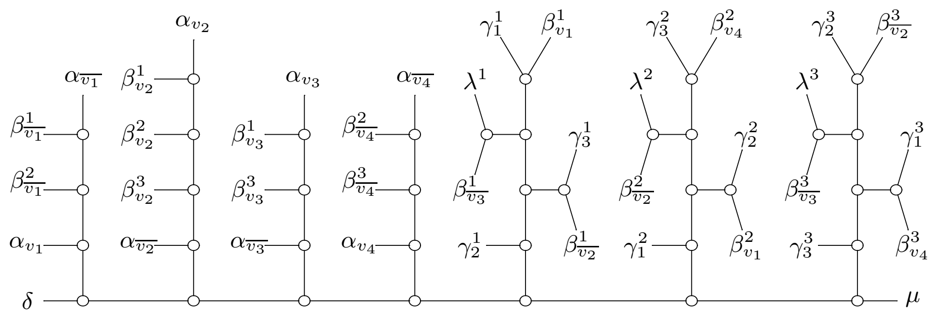

Fig. 3. The phylogenetic tree for the example instance I^+.

For instance, one of the quartet trees in $\mathcal{Q}_{I^+}$ is $\pi = \{\alpha_{v_1}, \beta^1_{v_1}\}|\{\alpha_{\overline{v_1}}, \beta^1_{\overline{v_1}}\}$ representing the forbidden edge of G_{I^+} between $S^1_{v_1}$ and $S^1_{\overline{v_1}}$. It is easy to verify $\mathcal{T}$ displays π. Another example from $\mathcal{Q}_{I^+}$ is $\pi' = \{\beta^1_{\overline{v_1}}, \lambda^1\}|\{\beta^1_{v_1}, \gamma^1_1\}$ representing the forbidden edge $K^1_{v_1}K^1_{\overline{v_1}}$. Again, it is displayed by $\mathcal{T}$, but this time one internal edge of T is contained in every set of edges of T that displays π' in $\mathcal{T}$; hence, this edge is distinguished by π'. This way we can verify all other quartet trees in $\mathcal{Q}_{I^+}$ and conclude that they are displayed by $\mathcal{T}$ and they distinguish $\mathcal{T}$.

Now, with the help of Theorem 7, this allows us to prove that given an instance I to ONE-IN-THREE-3SAT and a satisfying assignment φ for I, one can in polynomial time construct a phylogenetic tree $\mathcal{T}$ and a collection of quartet trees $\mathcal{Q}$ such that $\mathcal{T}$ is the unique tree defined by $\mathcal{Q}$ if and only if φ is the unique satisfying assignment for I. Combined with Theorem 2, this proves Theorem 1.

That concludes this section. In the next sections, we formally describe the above constructions and sketch some proofs. For full details of proofs, we invite the reader to see our arxiv version of this paper [19].

4 Formal description

Let I be an instance to ONE-IN-THREE-3SAT consisting of variables $v_1, \ldots, v_n$ and clauses $\mathcal{C}_1, \ldots, \mathcal{C}_m$. A *truth assignment* σ assigns to each variable v_i a truth value *true* or *false*; we indicate this by writing $v_i = 1$ or $v_i = 0$, respectively, and extend this notation to literals. A truth assignment σ is a *satisfying assignment for* I if in each clause $\mathcal{C}_j$ exactly one the three literals evaluates to true.

For each $i \in \{1 \ldots n\}$, we let Δ_i denote all indices j such that v_i or $\overline{v_i}$ appears in the clause $\mathcal{C}_j$. Let $\mathcal{X}_I$ be the set consisting of the elements:

a) α_{v_i}, $\alpha_{\overline{v_i}}$ for each $i \in \{1 \ldots n\}$,
b) $\beta^j_{v_i}$, $\beta^j_{\overline{v_i}}$ for each $i \in \{1 \ldots n\}$ and each $j \in \Delta_i$,
c) γ^j_1, γ^j_2, γ^j_3, λ^j for each $j \in \{1 \ldots m\}$, and
d) δ and μ.

Consider the following collection of 2-element subsets of $\mathcal{X}_I$:

a) $B = \Big\{\mu, \delta\Big\}$,

b) for each $i \in \{1, \ldots, n\}$:
$H_{v_i}{=}\{\alpha_{v_i},\, \delta\Big\}$, $H_{\overline{v_i}}{=}\{\alpha_{\overline{v_i}},\, \delta\Big\}$, $A_i = \Big\{\alpha_{v_i}, \alpha_{\overline{v_i}}\Big\}$,
$S^j_{v_i} = \Big\{\alpha_{v_i}, \beta^j_{v_i}\Big\}$, $S^j_{\overline{v_i}} = \Big\{\alpha_{\overline{v_i}}, \beta^j_{\overline{v_i}}\Big\}$ for all $j \in \Delta_i$,

c) for each $j \in \{1 \ldots m\}$ where $C_j = X \vee Y \vee Z$:

$$K^j_{\overline{X}} = \Big\{\beta^j_X, \gamma^j_1\Big\},\ K^j_{\overline{Y}} = \Big\{\beta^j_Y, \gamma^j_2\Big\},\ K^j_{\overline{Z}} = \Big\{\beta^j_Z, \gamma^j_3\Big\},$$
$$K^j_X = \Big\{\beta^j_{\overline{X}}, \lambda^j\Big\},\ K^j_Y = \Big\{\beta^j_{\overline{Y}}, \lambda^j\Big\},\ K^j_Z = \Big\{\beta^j_{\overline{Z}}, \lambda^j\Big\},$$
$$L^j_X = \Big\{\beta^j_{\overline{X}}, \gamma^j_2\Big\},\ L^j_Y = \Big\{\beta^j_{\overline{Y}}, \gamma^j_3\Big\},\ L^j_Z = \Big\{\beta^j_{\overline{Z}}, \gamma^j_1\Big\},$$
$$D^j_1 = \Big\{\gamma^j_1, \lambda^j\Big\},\quad D^j_2 = \Big\{\gamma^j_2, \lambda^j\Big\},\quad D^j_3 = \Big\{\gamma^j_3, \lambda^j\Big\},\quad F^j = \Big\{\lambda^j, \mu\Big\}.$$

The collection $\mathcal{Q}_I$ of quartet trees is defined as follows:

$$\mathcal{Q}_I = \bigcup_{i\in\{1\ldots n\}} \Big\{A_i|B\Big\} \cup \bigcup_{j\in\{1\ldots m\}} \Big\{D^j_1|B, D^j_2|B, D^j_3|B\Big\}$$
$$\cup \bigcup_{\substack{i\in\{1\ldots n\}\\ j,j'\in\Delta_i}} \Big\{S^j_{v_i}|S^{j'}_{\overline{v_i}}\Big\} \cup \bigcup_{\substack{i\in\{1\ldots n\}\\ j,j'\in\Delta_i \text{ and } j<j'}} \Big\{S^j_{v_i}|K^{j'}_{\overline{v_i}}, S^j_{\overline{v_i}}|K^{j'}_{v_i}\Big\} \cup \bigcup_{\substack{i\in\{1\ldots n\}\\ j\in\Delta_i \text{ and } j<j'\leq m}} \Big\{K^j_{\overline{v_i}}|F^{j'}, K^j_{v_i}|F^{j'}\Big\}$$
$$\cup \bigcup_{\substack{1\leq i'<i\leq n\\ j\in\Delta_i}} \Big\{H_{v_{i'}}|S^j_{v_i}, H_{\overline{v_{i'}}}|S^j_{v_i}, H_{v_{i'}}|S^j_{\overline{v_i}}, H_{\overline{v_{i'}}}|S^j_{\overline{v_i}}\Big\} \cup \bigcup_{\substack{i\in\{1\ldots n\}\\ j\in\{1\ldots m\}}} \Big\{H_{\overline{v_i}}|F^j, H_{v_i}|F^j\Big\}$$
$$\cup \bigcup_{\substack{j\in\{1\ldots m\}\\ \text{where } \mathcal{C}_j=X\vee Y\vee Z}} \left\{\begin{matrix} K^j_{\overline{X}}|K^j_X,\ K^j_{\overline{Y}}|K^j_Y,\ K^j_{\overline{Z}}|K^j_Z,\ K^j_{\overline{X}}|L^j_X,\ K^j_{\overline{Y}}|L^j_Y,\ K^j_{\overline{Z}}|L^j_Z \\ S^j_Y|K^j_X,\ \ S^j_Z|K^j_Y,\ \ S^j_X|K^j_Z,\ S^j_Z|L^j_X,\ \ S^j_X|L^j_Y,\ \ S^j_Y|L^j_Z \end{matrix}\right\}$$

Let T_I be the tree defined as follows:

$$V(T_I) = \Big\{y_0, y_1, y'_1, \ldots, y_n, y'_n\Big\} \cup \Big\{a_1, a'_1, \ldots, a_n, a'_n\Big\} \cup \Big\{u_0, u_1, \ldots, u_m\Big\}$$
$$\cup \Big\{x^j_1, x^j_2, x^j_3, x^j_4, x^j_5, x^j_6, b^j_1, b^j_2, b^j_3, g^j_1, g^j_2, g^j_3, \ell^j\Big\}_{j=1}^{m} \cup \Big\{c^j_i, z^j_i \mid j \in \Delta_i\Big\}_{i=1}^{n}$$

$$E(T_I) = \Big\{y_1y_1', y_2y_2', \ldots, y_ny_n'\Big\} \cup \Big\{a_1y_1', a_2y_2', \ldots a_ny_n'\Big\} \cup \Big\{c_i^jz_i^j \mid j \in \Delta_i\Big\}_{i=1}^{n}$$
$$\cup \Big\{y_0y_1, y_1y_2, y_2y_3, \ldots, y_{n-1}y_n\Big\} \cup \Big\{y_nu_1, u_1u_2, u_2u_3, \ldots, u_{m-1}u_m, u_mu_0\Big\}$$
$$\cup \Big\{u_jx_1^j, x_1^jx_2^j, x_2^jx_3^j, x_2^jx_4^j, x_4^jx_5^j, x_4^jx_6^j, b_1^jx_6^j, b_2^jx_3^j, b_3^jx_5^j, g_1^jx_6^j, g_2^jx_1^j, g_3^jx_3^j, \ell^jx_5^j\Big\}_{j=1}^{m}$$
$$\cup \Big\{a_i'z_i^{j_1}, z_i^{j_1}z_i^{j_2}, \ldots, z_i^{j_{t-1}}z_i^{j_t}, z_i^{j_t}y_i' \;\Big|\; j_1 < j_2 < \ldots < j_t \text{ are elements of } \Delta_i\Big\}_{i=1}^{n}$$

Let σ be a satisfying assignment for the instance I, and let ϕ_σ be the mapping of $\mathcal{X}_I$ to $V(T_I)$ defined as follows:

a) for each $i \in \{1 \ldots n\}$:
if $v_i = 1$, then $\phi_\sigma(\alpha_{v_i}) = a_i$, $\phi_\sigma(\alpha_{\overline{v_i}}) = a_i'$, $\phi_\sigma(\beta^j_{\overline{v_i}}) = c_i^j$ for all $j \in \Delta_i$,
if $v_i = 0$, then $\phi_\sigma(\alpha_{\overline{v_i}}) = a_i$, $\phi_\sigma(\alpha_{v_i}) = a_i'$, $\phi_\sigma(\beta^j_{v_i}) = c_i^j$ for all $j \in \Delta_i$,

b) for each $j \in \{1 \ldots m\}$ where $\mathcal{C}_j = X \vee Y \vee Z$:
if $X = 1$, then $\phi_\sigma(\beta^j_X) = b_1^j$, $\phi_\sigma(\beta^j_{\overline{Y}}) = b_2^j$, $\phi_\sigma(\beta^j_{\overline{Z}}) = b_3^j$,
$\phi_\sigma(\gamma_1^j) = g_1^j$, $\phi_\sigma(\gamma_2^j) = g_2^j$, $\phi_\sigma(\gamma_3^j) = g_3^j$, $\phi_\sigma(\lambda^j) = \ell_j$,
if $Y = 1$, then $\phi_\sigma(\beta^j_Y) = b_1^j$, $\phi_\sigma(\beta^j_{\overline{Z}}) = b_2^j$, $\phi_\sigma(\beta^j_{\overline{X}}) = b_3^j$,
$\phi_\sigma(\gamma_2^j) = g_1^j$, $\phi_\sigma(\gamma_3^j) = g_2^j$, $\phi_\sigma(\gamma_1^j) = g_3^j$, $\phi_\sigma(\lambda^j) = \ell_j$,
if $Z = 1$, then $\phi_\sigma(\beta^j_Z) = b_1^j$, $\phi_\sigma(\beta^j_{\overline{X}}) = b_2^j$, $\phi_\sigma(\beta^j_{\overline{Y}}) = b_3^j$,
$\phi_\sigma(\gamma_3^j) = g_1^j$, $\phi_\sigma(\gamma_1^j) = g_2^j$, $\phi_\sigma(\gamma_2^j) = g_3^j$, $\phi_\sigma(\lambda^j) = \ell_j$,

c) $\phi_\sigma(\delta) = y_0$ and $\phi_\sigma(\mu) = u_0$.

Theorem 9. *If σ is a satisfying assignment for I, then $\mathcal{T}_\sigma = (T_I, \phi_\sigma)$ is a ternary phylogenetic $\mathcal{X}_I$-tree that displays $\mathcal{Q}_I$ and is distinguished by $\mathcal{Q}_I$.*

5 Proof of Theorem 8

To explain the proof, we need the following naming convention adopted from [3]. If W is a literal in the clause $\mathcal{C}_j$, we say that S^j_W is a *shoulder of the clause* $\mathcal{C}_j$ as well as a *shoulder of the literal* W. It is a *a true shoulder* if $W = 1$; otherwise, a *false shoulder*. Similarly, the vertex K^j_W and L^j_W (if exists) are *knees of the clause* $\mathcal{C}_j$ as well as *knees of the literal* W. A knee of W is a *true knee* if $W = 1$; otherwise, a *false knee*. The vertices A_i, D^j_p, H_W, F^j for all meaningful choices of indices are respectively called *A-vertices*, *D-vertices*, *H-vertices*, and *F-vertices*.

Let G_σ be the graph constructed from $\mathrm{int}^*(\mathcal{Q}_I)$ by performing the following.

(i) make B adjacent to all true knees and true shoulders.

Let G'_σ be constructed from G_σ by performing the following steps.

(ii) make {true knees, true shoulders} into a complete graph,
(iii) for all $i \in \{1 \ldots n\}$, make A_i adjacent to all true knees of the literals v_i,$\overline{v_i}$,
(iv) for all $1 \le i' \le i \le n$, make H_{v_i}, $H_{\overline{v_i}}$ adjacent to all true knees and true shoulders of the literals $v_{i'}$, $\overline{v_{i'}}$,
(v) for all $1 \le j \le j' \le m$, make F^j adjacent to all true knees and true shoulders of the clause $\mathcal{C}_{j'}$,

(vi) for all $1 \le i \le n$ and all $j, j' \in \Delta_i$ such that $j \le j'$:

a) if $v_i = 1$, make $S^{j'}_{\overline{v_i}}$ adjacent to $K^j_{v_i}$, $L^j_{v_i}$ (if exists),

b) if $v_i = 0$, make $S^{j'}_{v_i}$ adjacent to $K^j_{\overline{v_i}}$, $L^j_{\overline{v_i}}$ (if exists).

Finally, let G^*_σ be constructed from G'_σ by adding the following edges.

(vii) for all $j \in \{1 \dots m\}$ where $\mathcal{C}_j = X \vee Y \vee Z$:

a) if $X = 1$, then add edges $F^j L^j_Z$, $K^j_X L^j_Z$, $K^j_Y K^j_{\overline{Z}}$, $D^j_2 K^j_{\overline{Z}}$, $D^j_2 S^j_{\overline{Y}}$, $D^j_3 S^j_{\overline{Y}}$ and make $\{D^j_1, D^j_2, D^j_3, S^j_X, S^j_{\overline{Z}}, L^j_Z, K^j_Y\}$ into a complete graph,

b) if $Y = 1$, then add edges $F^j L^j_X$, $K^j_Y L^j_X$, $K^j_Z K^j_{\overline{X}}$, $D^j_3 K^j_{\overline{X}}$, $D^j_3 S^j_{\overline{Z}}$, $D^j_1 S^j_{\overline{Z}}$ and make $\{D^j_1, D^j_2, D^j_3, S^j_Y, S^j_{\overline{X}}, L^j_X, K^j_Z\}$ into a complete graph,

c) if $Z = 1$, then add edges $F^j L^j_Y$, $K^j_Z L^j_Y$, $K^j_X K^j_{\overline{Y}}$, $D^j_1 K^j_{\overline{Y}}$, $D^j_1 S^j_{\overline{X}}$, $D^j_2 S^j_{\overline{X}}$ and make $\{D^j_1, D^j_2, D^j_3, S^j_Z, S^j_{\overline{Y}}, L^j_Y, K^j_X\}$ into a complete graph.

Lemma 2. *G'_σ is a subgraph of every chordal sandwich of $(G_\sigma, \mathrm{forb}(\mathcal{Q}_I))$.*

Lemma 3. *If σ is a satisfying assignment for I, then G^*_σ is a subgraph of every chordal sandwich of $(G_\sigma, \mathrm{forb}(\mathcal{Q}_I))$.*

Lemma 4. *For every chordal sandwich G' of $(\mathrm{int}^*(\mathcal{Q}_I), \mathrm{forb}(\mathcal{Q}_I))$, there is σ such that G_σ is a subgraph of G', and such that σ is a satisfying assignment for I.*

Lemma 5. *If σ is a satisfying assignment for I, then G^*_σ is chordal.*

Proof. (Sketch) Assume that σ is a satisfying assignment for I, i.e., in each clause $\mathcal{C}_j$ exactly one literal evaluates to 1 by the assignment.

Consider the partition $V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5$ of $V(G^*_\sigma)$ where $V_1 = \{$false knees, D-vertices$\}$, $V_2 = \{$false shoulders$\}$, $V_3 = \{A$-vertices$\}$, $V_4 = \{H$-vertices, F-vertices$\}$, and $V_5 = \{$true knees, true shoulders, the vertex $B\}$.

Let π be an enumeration of $V(G^*_\sigma)$ constructed by listing the elements of V_1, V_2, V_3, V_4, V_5 in this order such that:

(•) the elements of V_1 are listed by considering each clause $\mathcal{C}_j = X \vee Y \vee Z$ and listing vertices (based on the truth assignment) as follows:

a) if $X = 1$, then list $K^j_{\overline{X}}$, K^j_Z, L^j_Y, L^j_Z, D^j_1, K^j_Y, D^j_3, D^j_2 in this order,

b) if $Y = 1$, then list $K^j_{\overline{Y}}$, K^j_X, L^j_Z, L^j_X, D^j_2, K^j_Z, D^j_1, D^j_3 in this order,

c) if $Z = 1$, then list $K^j_{\overline{Z}}$, K^j_Y, L^j_X, L^j_Y, D^j_3, K^j_X, D^j_2, D^j_1 in this order,

(•) the elements of V_2 (the false shoulders) are listed by listing the false shoulders of the clauses $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m$ in this order,

(•) the elements of V_4 are listed as follows: first H_{v_1}, $H_{\overline{v_1}}$, H_{v_2}, $H_{\overline{v_2}}$, $\dots H_{v_n}$, $H_{\overline{v_n}}$ in this order, then F^m, F^{m-1}, $\dots$, F^1 in this order,

(•) the elements of V_3 and V_5 are listed in any order.

A simple but tedious analysis shows that π is a perfect elimination ordering of the vertices of G^*_σ. This proves that G^*_σ is indeed a chordal graph (see [14]). □

Proof of Theorem 8. Let G' be a minimal chordal sandwich of $(\mathrm{int}^*(\mathcal{Q}_I), \mathrm{forb}(\mathcal{Q}_I))$. By Lemma 4, there exists σ, a satisfying assignment for I, such that

G_σ is a subgraph fo G'. Thus, G' is also a chordal sandwich of $(G_\sigma, \text{forb}(\mathcal{Q}_I))$, and hence, G^*_σ is a subgraph of G' by Lemma 3. But by Lemma 5, G^*_σ is chordal, and so G' is isomorphic to G^*_σ by the minimality of G'.

Conversely, if σ is a satisfying assignment for I, then the graph G^*_σ is chordal by Lemma 5. Moreover, $\text{int}^*(\mathcal{Q}_I)$ is a subgraph of G^*_σ, by definition, and G^*_σ contains no edges of $\text{forb}(\mathcal{Q}_I)$, also by definition. Thus, G^*_σ is a chordal sandwich of $(\text{int}^*(\mathcal{Q}_I), \text{forb}(\mathcal{Q}_I))$, and it is minimal by Lemma 3.

This proves that by mapping each satisfying assignment σ to the graph G^*_σ, we obtain the required bijection. That concludes the proof. $\square$

6 Conclusion

In this paper, we have shown that determining whether a given phylogenetic tree represents the unique evolution of given species is an NP-hard problem. This implies that the problem is actually $CoNP$-complete, as it can be defined by the formula "for every pair of trees, if they are solutions, they are isomorphic". Moreover, the problem clearly remains NP-hard even if the tree is not provided and we only want to test whether there is a unique solution. (For this, note that isomorphism of trees and testing if a tree is a solution takes polynomial time.)

In addition, we proved that the unique chordal sandwich problem is NP-hard. Following this direction, it would be interesting to consider the complexity of uniqueness of other sandwich problems, for instance, interval sandwich (DNA physical mapping) or cograph sandwich (genome comparison); the decision problem for the former is NP-hard [16] while it is polynomial for the latter [5, 15].

References

1. Agarwala, R., Fernández-Baca, D.: A polynomial-time algorithm for the perfect phylogeny problem when the number of character states is fixed. SIAM Journal of Computing 23, 1216–1224 (1994)
2. Böcker, S., Bryant, D., Dress, A.W.M., Steel, M.A.: Algorithmic aspects of tree amalgamation. Journal of Algorithms 37, 522–537 (2000)
3. Bodlaender, H.L., Fellows, M.R., Warnow, T.J.: Two strikes against perfect phylogeny. In: Proceedings of ICALP 1992, Lecture Notes in Computer Science 623. pp. 273–283. Springer Berlin/Heidelberg (1992)
4. Bonet, M.L., Linz, S., John, K.S.: The complexity of finding multiple solutions to betweenness and quartet compatibility. CoRR abs/1101.2170 (2011), http://arxiv.org/abs/1101.2170
5. Bui-Xuan, B.M., Habib, M., Paul, C.: Competitive graph searches. Journal Theoretical Computer Science 393, 72–80 (2008)
6. Buneman, P.: A characterization of rigid circuit graphs. Discrete Mathematics 9, 205–212 (1974)
7. Camin, J., Sokal, R.: A method for deducing branching sequences in phylogeny. Evolution 19, 311–326 (1965)
8. Creignou, N., Hermann, M.: Complexity of generalized satisfiability counting problems. Information and Computation 125, 1–12 (1996)

9. Dekker, M.C.H.: Reconstruction methods for derivation trees. Master's thesis, Vrije Universiteit, Amsterdam (1986)
10. Estabrook, G.F.: Cladistic methodology: a discussion of the theoretical basis for the induction of evolutionary history. Annual Review of Ecology and Systematics 3, 427–456 (1972)
11. Estabrook, G.F., C. S. Johnson, J., McMorris, F.R.: An idealized concept of the true cladistic character. Mathematical Biosciences 23, 263–272 (1975)
12. Estabrook, G.F., C. S. Johnson, J., McMorris, F.R.: An algebraic analysis of cladistic characters. Discrete Mathematics 16, 141–147 (1976)
13. Estabrook, G.F., C. S. Johnson, J., McMorris, F.R.: A mathematical foundation for the analysis of cladistic character compatibility. Mathematical Biosciences 29, 181–187 (1976)
14. Golumbic, M.C.: Algorithmic Graph Theory and Perfect Graphs (2nd edition). North Holland (2004)
15. Golumbic, M.C., Kaplan, H., Shamir, R.: Graph sandwich problems. Journal of Algorithms 19, 449–473 (1995)
16. Golumbic, M.C., Shamir, R.: Complexity and algorithms for reasoning about time: a graph-theoretic approach. Journal of the ACM 40, 1108–1133 (1993)
17. Gordon, A.D.: Consensus supertrees: The synthesis of rooted trees containing overlapping sets of labeled leaves. Journal of Classification 3, 335–348 (1986)
18. Gusfield, D.: Efficient algorithms for inferring evolutionary trees. Networks 21, 19–28 (1991)
19. Habib, M., Stacho, J.: Unique perfect phylogeny is NP-hard. CoRR abs/1011.5737 (2010), http://arxiv.org/abs/1011.5737
20. Juban, L.: Dichotomy theorem for the generalized unique satisfiability problem. In: Proceedings of FCT 1999, Lecture Notes in Computer Science 1684. pp. 327–337. Springer Berlin/Heidelberg (1999)
21. Kannan, S.K., Warnow, T.J.: Triangulating 3-colored graphs. SIAM Journal on Discrete Mathematics 5, 249–258 (1992)
22. Lam, F., Gusfield, D., Sridhar, S.: Generalizing the splits equivalence theorem and four gamete condition: Perfect phylogeny on three state characters. In: Algorithms in Bioinformatics (WABI 2009), Lecture Notes in Computer Science 5724. pp. 206–219. Springer Berlin/Heidelberg (2009)
23. LeQuesne, W.J.: Further studies on the uniquely derived character concept. Systematic Zoology 21, 281–288 (1972)
24. LeQuesne, W.J.: The uniquely evolved character concept and its cladistic application. Systematic Zoology 23, 513–517 (1974)
25. LeQuesne, W.J.: The uniquely evolved character concept. Systematic Zoology 26, 218–223 (1977)
26. McMorris, F.R., Warnow, T., Wimer, T.: Triangulating vertex colored graphs. SIAM Journal on Discrete Mathematics 7, 296–306 (1994)
27. Semple, C., Steel, M.: A characterization for a set of partial partitions to define an X-tree. Discrete Mathematics 247, 169–186 (2002)
28. Semple, C., Steel, M.: Phylogenetics. Oxford lecture series in mathematics and its applications, Oxford University Press (2003)
29. Steel, M.: personal webpage, http://www.math.canterbury.ac.nz/~m.steel/
30. Steel, M.: The complexity of reconstructing trees from qualitative characters and subtrees. Journal of Classification 9, 91–116 (1992)
31. West, D.: Introduction to Graph Theory. Prentice Hall (1996)
32. Wilson, E.O.: A consistency test for phylogenies based upon contemporaneous species. Systematic Zoology 14, 214–220 (1965)