



Corrigendum to “Best match graphs”

David Schaller, et al. [full author details at the end of the article]

Received: 4 September 2020 / Revised: 9 February 2021 / Accepted: 24 March 2021 /
Published online: 5 April 2021
© The Author(s) 2021

Abstract

Two errors in the article *Best Match Graphs* (Geiß et al. in JMB 78: 2015–2057, 2019) are corrected. One concerns the tacit assumption that digraphs are sink-free, which has to be added as an additional precondition in Lemma 9, Lemma 11, Theorem 4. Correspondingly, Algorithm 2 requires that its input is sink-free. The second correction concerns an additional necessary condition in Theorem 9 required to characterize best match graphs. The amended results simplify the construction of least resolved trees for n -cBMGs, i.e., Algorithm 1. All other results remain unchanged and are correct as stated.

Keywords Corrigendum · Phylogenetic Combinatorics · Colored digraph · Rooted triples

Mathematics Subject Classification 05C05 · 68R10 · 92-10 · 92B10

Best match graphs (BMGs) must be sink-free

Throughout Geiß et al. (2019) we have tacitly assumed that all vertex-colored digraphs (G, σ) satisfy the following property, which, by construction, is true for all colored best match graphs (cBMGs):

For each vertex x with color $\sigma(x)$, there is an arc xy to at least one vertex of y every other color $\sigma(y) \neq \sigma(x)$.

All properly 2-colored digraphs appearing in the text are therefore assumed to be sink-free, i.e., the out-neighborhoods of their \simeq -classes are assumed to be non-empty:

(N4) $N(\alpha) \neq \emptyset$ for all $\alpha \in \mathcal{N}$.

This assumption was not clearly stated in the text.

Property (N4) is required in the proof of Lemma 9 [last line on page 2032]: Here $R(\beta) \cap R(\alpha^*) = \emptyset$ only implies $\beta \subseteq R(\alpha)$ for the \simeq classes β with $R(\beta) \subseteq R(\alpha)$

✉ Peter F. Stadler
studla@bioinf.uni-leipzig.de

Extended author information available on the last page of the article

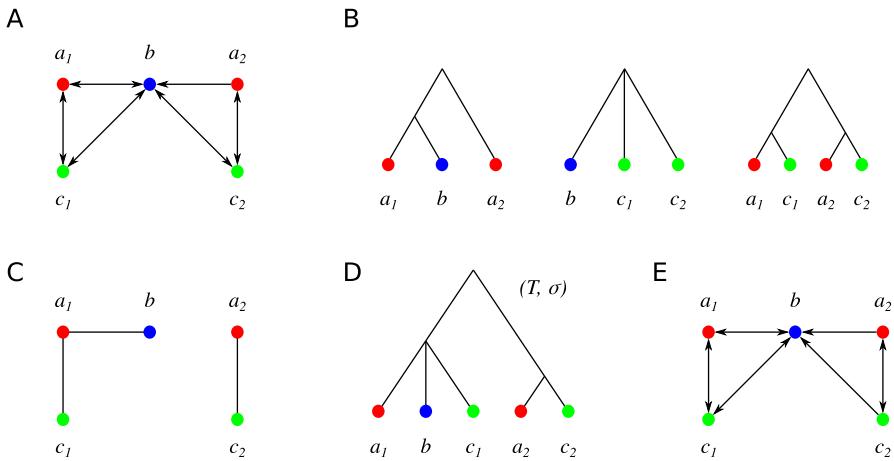


Fig. 1 Counterexample for the original version of Theorem 9. **a** A colored digraph with vertex set L that is not a 3-cBMG. **b** The least resolved subtrees for the three 2-colored induced subgraphs. The union of their triples is $R := \{a_1b|a_2, a_1c_1|a_2, a_1c_1|c_2, a_2c_2|a_1, a_2c_2|c_1\}$. **c** The Aho-graph $[R, L]$. In particular, R forms a consistent set. **d** The tree $T := \text{Aho}(R)$. **e** The 3-cBMG $G(T, \sigma)$. The arc bc_2 that was present in (G, σ) is missing in $G(T, \sigma)$

if $R(\beta) \neq \emptyset$, which in turn is equivalent to $N(\beta) \neq \emptyset$. Furthermore, $N(\alpha) = \emptyset$ implies $Q(\alpha) = \alpha$ and thus $R'(\alpha) = \alpha$. Property (N4) is therefore also necessary to ensure that $|R'(\alpha)| > 1$ and thus that the tree $T(\mathcal{H}')$ is phylogenetic [page 2035, just before Theorem 4]. In summary, Lemma 9, Lemma 11, and Theorem 4 require (N4) as additional precondition.

Corrected characterization of n -cBMGs

The second paragraph of the proof of Theorem 9 in (Geiß et al. 2019, page 2045) incorrectly states that “for any $\alpha \in \mathcal{N}$ and any color $s \neq \sigma(\alpha)$ the out-neighborhood $N_s(\alpha)$ is the same w.r.t. (T_{st}, σ_{st}) and w.r.t. $\text{Aho}(R)$.”, leading to the incorrect conclusion that $G(\text{Aho}(R), \sigma) = (G, \sigma)$ whenever $\text{Aho}(R)$ exists. We recall that the triple set R is defined as the union

$$R := \bigcup_{s,t \in S} r(T_{st}^*)$$

of all triples in the least resolved trees (T_{st}^*, σ_{st}) that explain the induced subgraphs (G_{st}, σ_{st}) of (G, σ) , and the Aho tree $\text{Aho}(R)$ is defined on the leaf set $L = V(G)$ (which may not have been clear from the wording in the text). We shall show in Proposition 3 below that $G(\text{Aho}(R), \sigma)$ is always a subgraph of (G, σ) whenever R is consistent. The example in Fig. 1 shows, however, that $G(\text{Aho}(R), \sigma) \neq (G, \sigma)$ is possible because $\text{Aho}(R)$ can contain triples that are not present in any of the 2-colored trees (T_{st}, σ_{st}) .

As a consequence, the characterization of n -cBMGs requires the equality $(G, \sigma) = G(\text{Aho}(R), \sigma)$ as an additional condition. The corrected result, with the correction underlined, reads as follows:

Theorem 9 *A connected colored digraph (G, σ) is an n -cBMG if and only if (1) all induced subgraphs (G_{st}, σ_{st}) on two colors are 2-cBMGs, (2) the union R of all triples obtained from their least resolved trees (T_{st}, σ_{st}) forms a consistent set, and (3) $G(\text{Aho}(R), \sigma) = (G, \sigma)$. In particular, $(\text{Aho}(R), \sigma)$ is the unique least resolved tree that explains (G, σ) .*

Condition (1) in Theorem 9, i.e., the requirement that all 2-colored induced subgraphs (G_{st}, σ_{st}) of (G, σ) are 2-cBMGs, is necessary to ensure that the least resolved trees (T_{st}, σ_{st}) exist and thus that the triple sets $r(T_{st})$ – and therefore also the set R of all triples displayed by the 2-colored induced subgraphs – are well-defined. Consistency of R is necessary for the existence of $\text{Aho}(R)$. Clearly, Condition (3) is sufficient to ensure that (G, σ) is an n -cBMG. Hence, it remains to show that Condition (3) is also necessary. This is achieved in Proposition 2 below.

Proof of theorem 9

Instead of proving the corrected version of Theorem 9 directly, we first state and prove a slightly stronger and more convenient result, Theorem 1 below, and then proceed to derive Theorem 9. To this end, we first generalize Definition 8 in (Geiß et al. (2019), page 2036) to digraphs with an arbitrary number of colors:

Definition 1 (Schaller et al. (2021), Definition 2.7) Let (G, σ) be a colored digraph. We say that a triple $ab|b'$ is *informative* for (G, σ) if a, b and b' are pairwise distinct vertices in G such that (i) $\sigma(a) \neq \sigma(b) = \sigma(b')$ and (ii) $ab \in E(G)$ and $ab' \notin E(G)$. The set of informative triples is denoted by $\mathcal{R}(G, \sigma)$.

We briefly argue that, for 2-colored digraphs, the definition of informative triples given here is equivalent to the one given in Geiß et al. (2019): By definition, an informative triple of some colored digraph has vertices with exactly two colors, and thus is also an informative triple in one of its 2-colored induced subgraphs. It is easy to check that, for 2-colored digraphs, Definition 1 is equivalent to Definition 8 in Geiß et al. (2019), since the four induced subgraphs shown in Fig. 8 in (Geiß et al. (2019), page 2036) correspond to the presence or absence of the two optional arcs ba and ca in the informative triple $ab|c$ (as defined here).

We will also make use of a generalization of Lemma 12 in (Geiß et al. (2019), page 2036):

Lemma 1 (Schaller et al. (2021), Lemma 2.8) *Let (G, σ) be an n -cBMG and $ab|b'$ an informative triple for (G, σ) . Then, every tree (T, σ) that explains (G, σ) displays the triple $ab|b'$, i.e. $\text{lca}_T(a, b) <_T \text{lca}_T(a, b') = \text{lca}_T(b, b')$.*

Given a digraph (G, σ) for which R exists, Lemma 1 in particular implies that

$$\mathcal{R}(G, \sigma) \subseteq R. \quad (1)$$

With these preliminaries, we are ready to formulate our new main result as

Theorem 1 *A colored digraph (G, σ) is an n -cBMG if and only if $G(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma) = (G, \sigma)$. Moreover, $\text{Aho}(\mathcal{R}(G, \sigma))$ is the unique least resolved tree explaining an n -cBMG (G, σ) .*

In order to prove Theorem 1, we will first provide several technical results that make use of the notion of (non-)redundant tree edges and, in particular, of least resolved trees. Recall that an inner edge e in a leaf-colored tree (T, σ) is *redundant* if the tree (T_e, σ) obtained from T by contraction of e explains the same n -cBMG, i.e., if $G(T, \sigma) = G(T_e, \sigma)$. A tree (T, σ) is called *least resolved* if it does not contain any redundant edges. We will need the following, simplified, characterization of redundant edges:

Lemma 2 (Schaller et al. (2021), Lemma 2.10) *Let (G, σ) be an n -cBMG explained by a tree (T, σ) . The edge $e = uv$ with $v \prec_T u$ in (T, σ) is redundant w.r.t. (G, σ) if and only if (i) e is an inner edge of T and (ii) there is no arc $ab \in E(G)$ such that $\text{lca}_T(a, b) = v$ and $\sigma(b) \in \sigma(L(T(u)) \setminus L(T(v)))$.*

We note that the proofs of Lemma 1 (Schaller et al. (2021), Lemma 2.8) and Lemma 2 (Schaller et al. (2021), Lemma 2.10) only require the definition of best match graphs, and are thus independent of the results proved in Geiß et al. (2019).

Following Bryant and Steel (1995), an inner edge e of a rooted tree T is *distinguished* by a triple $ab|c \in r(T)$ if the path from a to c in T intersects the path from b to the root ρ_T precisely on the edge e . In other words, $e = uv$ with $v \prec_T u$ is distinguished by $ab|c$ if $\text{lca}_T(a, b) = v$ and $\text{lca}_T(a, b, c) = u$. Lemma 2 immediately implies the following generalization of Lemma 13 in Geiß et al. (2019):

Corollary 1 *Let (G, σ) be an n -cBMG explained by a tree (T, σ) . An inner edge e of (T, σ) is non-redundant w.r.t. (G, σ) if and only if it is distinguished by an informative triple $ab|b'$ for (G, σ) . In particular, if (T, σ) is least resolved, then each of its inner edges is distinguished by an informative triple.*

In addition, we will need the following two technical results relating subtrees and induced subgraphs of n -cBMGs.

Lemma 3 *Let (T, σ) be a tree explaining an n -cBMG (G, σ) . Then $G(T(u), \sigma|_{L(T(u))}) = (G[L(T(u))], \sigma|_{L(T(u))})$ holds for every $u \in V(T)$.*

Proof Let $(G_1, \sigma') := (G(T(u), \sigma|_{L(T(u))})$ and $(G_2, \sigma') := (G[L(T(u))], \sigma|_{L(T(u))})$. By definition, we have $V(G_1) = V(G_2) = L(T(u))$. First assume that $xy \in E(G_1)$ for some $x, y \in L(T(u))$. Hence, it holds that $\text{lca}_{T(u)}(x, y) \preceq_{T(u)} \text{lca}_{T(u)}(x, y')$ for all y' with $\sigma(y) = \sigma(y')$ in $T(u)$ and thus, since $T(u)$ is a subtree of T , we have $\text{lca}_T(x, y) \preceq_T \text{lca}_T(x, y')$ for all y' with $\sigma(y) = \sigma(y')$ in T . Therefore, $xy \in E(G)$. Since $x, y \in L(T(u))$ and G_2 is the subgraph of G induced by $L(T(u))$, we have $xy \in E(G_2)$ and thus $E(G_1) \subseteq E(G_2)$. Now assume $xy \in E(G_2)$ for some $x, y \in L(T(u))$. Hence, $xy \in E(G)$. Consequently, there is no leaf y' in T with $\sigma(y') = \sigma(y) \neq \sigma(x)$ such that $\text{lca}_T(x, y') \prec_T \text{lca}_T(x, y) \preceq_T u$. This clearly also holds for the subtree $T(u)$. Therefore, we have $xy \in E(G_1)$ and thus $E(G_2) \subseteq E(G_1)$.

□

Lemma 4 *If (T, σ) is least resolved for an n -cBMG (G, σ) , then the subtree $T(u)$ is least resolved for the n -cBMG $G(T(u), \sigma|_{L(T(u))})$ for each $u \in V(T)$.*

Proof The statement is trivially satisfied if $T(u)$ does not contain any inner edges, which is exactly the case if either $u \in L(T)$ or $u \in V^0(T)$ with $\text{child}_T(u) \subseteq L(T)$. Thus, let $u \in V^0(T)$ and $\text{child}_T(u) \cap V^0(T) \neq \emptyset$. Since (T, σ) is least resolved, it does not contain redundant edges. Let vw be an inner edge of $T(u)$ with $w \prec_T v \preceq_T u$, and note that vw must also be an inner edge in T . By Lemma 2 and since vw is not redundant in T , there is an arc $ab \in E(G)$ such that $\text{lca}_T(a, b) = w$ and $\sigma(b) \in \sigma(L(T(v)) \setminus L(T(w)))$. Since $u \succeq_T v$, Lemma 3 implies that ab is also an arc in $G(T(u), \sigma|_{L(T(u))})$ and $\text{lca}_{T(u)}(a, b) = v$. Hence, in particular, we have $\sigma(b) \in \sigma(L(T(v)) \setminus L(T(w)))$. We can now apply Lemma 2 to conclude that vw is not redundant in $T(u)$. Since vw was chosen arbitrarily, we conclude that $T(u)$ does not contain any redundant edge and thus, it must be least resolved for $G(T(u), \sigma|_{L(T(u))})$ for all $u \in V(T)$. \square

We finally relate the subtrees $T(u)$ to the construction of the Aho-graph as specified in (Geiß et al. (2019), Sec. 3.4). Given a set of triples R on L , we will write $R|_{L'}$ for the set of triples $ab|c \in R$ with $a, b, c \in L' \subseteq L$.

Lemma 5 *Let (T, σ) be least resolved for an n -cBMG (G, σ) with informative triple set $\mathcal{R} := \mathcal{R}(G, \sigma)$. Then, $L(T(v))$ is a connected component in the Aho-graph $[\mathcal{R}|_{L(T(u))}, L(T(u))]$ for every inner vertex u and each of its children $v \in \text{child}_T(u)$.*

Proof We proceed by induction on $L := V(G)$. The statement trivially holds for $|L| = 1$. Hence, suppose that $|L| > 1$ and assume that the statement is true for every n -cBMG with less than $|L|$ vertices.

Let u be an inner vertex of T and $v \in \text{child}_T(u)$. We first show that $L(T(v))$ is connected in $[\mathcal{R}|_{L(T(u))}, L(T(u))]$, and then argue that there are no edges between $L(T(v))$ and $L(T(u)) \setminus L(T(v))$, i.e., that $L(T(v))$ forms a connected component.

If uv is an outer edge, i.e. v is a leaf, then $L(T(v))$ is trivially connected. Now suppose that uv is an inner edge of T . By Lemmas 3 and 4, $(G[L(T(v))], \sigma|_{L(T(v))})$ is explained by the least resolved tree $(T(v), \sigma|_{L(T(v))})$. By the induction hypothesis, $L(T(w))$ forms a connected component in $[\mathcal{R}|_{L(T(v))}, L(T(v))]$ for all children $w \in \text{child}_T(v)$. Together with $\mathcal{R}|_{L(T(v))} \subseteq \mathcal{R}|_{L(T(u))}$, this implies that the elements in $L(T(w))$ are also connected in $[\mathcal{R}|_{L(T(u))}, L(T(u))]$ for all $w \in \text{child}_T(v)$. Since uv is an inner edge of the least resolved tree (T, σ) , we can apply Corollary 1 to conclude that there is an informative triple $ab|b'$ in (G, σ) that distinguishes uv , i.e. $\text{lca}_T(a, b) = v$ and $b' \in L(T(u)) \setminus L(T(v))$ with color $\sigma(b') = \sigma(b)$. Hence, $ab|b'$ is also contained in $[\mathcal{R}|_{L(T(u))}, L(T(u))]$. In particular, there are children $w, w' \in \text{child}_T(v)$ such that $a \preceq_T w$ and $b \preceq_T w'$, and the edge ab connects $L(T(w))$ and $L(T(w'))$ in $[\mathcal{R}|_{L(T(u))}, L(T(u))]$.

Now suppose that there is an additional child $w'' \in \text{child}_T(v) \setminus \{w, w'\}$. We distinguish two cases. Either there is a leaf $b'' \preceq_T w''$ with $\sigma(b'') = \sigma(b)$ or no such leaf exists. If there is such a leaf b'' , then ab'' forms an arc in (G, σ) and $ab''|b'$ is an informative triple making $L(T(w))$ and $L(T(w''))$ connected in $[\mathcal{R}|_{L(T(u))}, L(T(u))]$. Otherwise, take an arbitrary leaf $c \preceq_T w''$. Since $\sigma(b) \notin \sigma(L(T(w'')))$, we have

$\sigma(c) \neq \sigma(b)$ and thus, there is an arc cb in (G, σ) . Since $\text{lca}_T(c, b') = u \succ_T v = \text{lca}_T(c, b)$, the arc cb' is not contained in (G, σ) . Hence, $cb|b'$ is an informative triple making $L(T(w'))$ and $L(T(w''))$ connected in $[\mathcal{R}_{|L(T(u))}, L(T(u))]$.

Therefore, the subgraph in $[\mathcal{R}_{|L(T(u))}, L(T(u))]$ induced by $L(T(v))$ must be connected.

It remains to show that $L(T(v))$ is a connected component in $[\mathcal{R}_{|L(T(u))}, L(T(u))]$ and thus, that there are no edges ab in $[\mathcal{R}_{|L(T(u))}, L(T(u))]$ with $a \in L(T(v))$ and $b \in L(T(u)) \setminus L(T(v))$. Assume, for contradiction, that there exists such an edge ab . Hence, this edge must be supported by an informative triple w.l.o.g. $ab|b'$ with $\sigma(a) \neq \sigma(b) = \sigma(b')$ and $b' \in L(T(u))$. Lemma 1 implies that $ab|b'$ must be displayed by T . However, $\text{lca}_T(a, b) = u = \text{lca}_T(a, b, b')$ implies that such a triple cannot exist. Thus, $L(T(v))$ is a connected component in $[\mathcal{R}_{|L(T(u))}, L(T(u))]$. \square

The least resolved tree of an n -cBMG therefore coincides with the Aho tree of its informative triples. In more detail, we have

Proposition 1 *If (G, σ) is an n -cBMG, then $(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma)$ is the unique least resolved tree for (G, σ) .*

Proof Since (G, σ) is an n -cBMG, Lemma 1 implies that there is a tree displaying all triples in $\mathcal{R}(G, \sigma)$. In particular, therefore, $\text{Aho}(\mathcal{R}(G, \sigma))$ exists. Moreover, there must be a least resolved tree (T^*, σ) for (G, σ) . To see this, consider an arbitrary tree (T, σ) that explains (G, σ) , and repeatedly identify and contract a redundant edge until no redundant edges remain. By definition, the resulting tree still explains (G, σ) and is least resolved. By Lemma 5 and by construction of $(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma)$, any least resolved tree (T^*, σ) for (G, σ) coincides with the latter. The uniqueness of $\text{Aho}(\mathcal{R}(G, \sigma))$ therefore implies that the least resolved tree is also unique. \square

We now have all the pieces in place to complete the proof of the main result:

Proof of Theorem 1 If (G, σ) is an n -cBMG, then Proposition 1 implies that $(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma)$ is its unique least resolved tree, and thus $G(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma) = (G, \sigma)$. Conversely, $G(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma)$ is an n -cBMG. \square

None of the intermediate results used to prove Theorem 9 in Geiß et al. (2019) is used below in our proof of Theorem 1. However, to the best of our knowledge, all results in Geiß et al. (2019) with the exception of the aforementioned Lemmas 9 and 11, and Thms. 4 and 9 are correct as stated. It is worth noting, furthermore, that Theorem 1 immediately implies Thms. 5, 6, and 7, as well as the existence of a unique least resolved tree in Thms. 2 and 8 of Geiß et al. (2019). In particular, Theorem 1 allows us to obtain the least resolved tree of an n -cBMG without the need to explicitly construct the least resolved trees of all its 2-colored induced subgraphs.

To prove the correctness of the amended version of Theorem 9, it only remains to show

Proposition 2 *If (G, σ) is a n -cBMG, then $\text{Aho}(\mathcal{R}(G, \sigma)) = \text{Aho}(R)$.*

Proof For brevity set $\mathcal{R} := \mathcal{R}(G, \sigma)$. From Eq. (1), i.e., $\mathcal{R} \subseteq R$, we immediately have $\mathcal{R}_{|L(T(u))} \subseteq R_{|L(T(u))}$ for every inner vertex u of T . Moreover, by Theorem 1, (T, σ) with $T := \text{Aho}(\mathcal{R})$ is the least resolved tree that explains (G, σ) .

Hence, we can apply the same arguments as in the proof of Lemma 5 to conclude that $L(T(v))$ forms a connected component in the Aho-graph $[R_{|L(T(u))}, L(T(u))]$ for every inner vertex u and each of its children $v \in \text{child}_T(u)$. More precisely, note that connectedness of any such $L(T(v))$ is guaranteed by the informative triples. Now assume, for contradiction, that there is an edge ab in $[R_{|L(T(u))}, L(T(u))]$ with $a \in L(T(v))$ and $b \in L(T(u)) \setminus L(T(v))$ connecting $L(T(v))$ and $L(T(v'))$ for some child $v' \in \text{child}_T(u) \setminus \{v\}$. In this case, there is a triple $ab|c \in R_{|L(T(u))}$ and thus, $a, b, c \in L(T(u))$ and $\text{lca}_T(a, b, c) = u$. By definition of R and Observation 4 in Geiß et al. (2019), $ab|c$ must be displayed by T . However, $a, b, c \in L(T(u))$ and $\text{lca}_T(a, b) = u = \text{lca}_T(a, b, c)$ imply that $ab|c$ is not displayed by T ; a contradiction. Therefore, $(T, \sigma) = (\text{Aho}(R), \sigma)$, which completes the proof. \square

For completeness, we show that conditions (i) and (ii) of Theorem 9 ensure that $G(\text{Aho}(R), \sigma)$ and $G(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma)$ are subgraphs of (G, σ) .

Proposition 3 *Let (G, σ) be a properly n -colored digraph with all 2-colored induced subgraphs being 2-cBMGs. Then the following two statements hold:*

- (1) *If $\mathcal{R}(G, \sigma)$ is consistent, then $G(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma) \subseteq (G, \sigma)$.*
- (2) *If R is consistent, then $G(\text{Aho}(R), \sigma) \subseteq (G, \sigma)$.*

Proof We set $(G', \sigma') := G(\text{Aho}(\mathcal{R}(G, \sigma)), \sigma)$. Since $\text{Aho}(\mathcal{R}(G, \sigma))$ is defined on $V(G)$, we have $V(G') = V(G)$ and $\sigma' = \sigma$. Now assume, for contradiction, that there is an arc $ab \in E(G')$ such that $ab \notin E(G)$. By assumption, the induced subgraph (G_{st}, σ_{st}) of (G, σ) , where $s = \sigma(a)$ and $t = \sigma(b)$, is a 2-cBMG and thus sink-free. Therefore, there must be a vertex b' of color $\sigma(b)$ with $ab' \in E(G)$. Hence, $ab'|b$ is informative for (G, σ) and contained in $\mathcal{R}(G, \sigma)$. In particular, $ab'|b$ must be displayed by $\text{Aho}(\mathcal{R}(G, \sigma))$; contradicting that ab is an arc in (G', σ') . Hence, statement (i) is true.

Statement (ii) can be shown using Eq. (1), i.e., $\mathcal{R}(G, \sigma) \subseteq R$, and arguments similar to the previous paragraph. \square

Consequences for the algorithms

Finally, we discuss the consequences of the corrections for the algorithmic aspects outlined in Section 5 of Geiß et al. (2019).

Algorithm 2 constructs the least resolved tree for 2-cBMGs based on Theorem 4 in Geiß et al. (2019). It therefore requires a sink-free graph as input, or needs to be amended to check that its input satisfies condition (N4). This can be done trivially in $O(|E|)$ time. The statements concerning its complexity, i.e., Lemmas 18 and 19, therefore are still correct.

Regarding the recognition of n -cBMGs, we have noted above that the consistency of the triple set R and the fact that all 2-colored induced subgraphs are 2-BMGs are not sufficient. Algorithm 1 of Geiß et al. (2019) therefore also needs to be corrected. By Theorem 1, it suffices to construct the tree $T := \text{Aho}(\mathcal{R}(G, \sigma))$ and to check whether $G(T, \sigma) = (G, \sigma)$. On the other hand, it is no longer necessary to require connectedness of the input graph. We therefore obtain a considerably simpler procedure, see Alg. 1.

Algorithm 1: Unique least resolved tree of n -cBMG.**Input:** Properly colored digraph $(G = (L, E), \sigma)$ **Output:** Least resolved tree (T, σ) if (G, σ) is an n -cBMG, false otherwise.

```

1  $T \leftarrow \text{Aho}(\mathcal{R}(G, \sigma))$ 
2 if  $G(T, \sigma) = (G, \sigma)$  then
3   | return  $(T, \sigma)$ 
4 else
5   | return false

```

The same arguments as in Geiß et al. (2019) show that $T = \text{Aho}(\mathcal{R}(G, \sigma))$ can be constructed in $O(|E||L| \log^2(|E||L|)) = O(|E||L| \log^2 |L|)$ time using the algorithm by Deng and Fernández-Baca (2018). The construction of $G(T, \sigma)$ can then be achieved in $O(|L|^2)$ time e.g. using Algorithm 1 of the Supplement of Geiß et al. (2020). The equality $G(T, \sigma) = (G, \sigma)$ can be checked in $O(|L|^2)$ operations. The total effort therefore remains dominated by the construction of the least resolved tree T .

We note that Algorithm 3 in Geiß et al. (2019) is essentially the simplified Algorithm 1 above with its input restricted to 2-colored connected digraphs. Its correctness therefore follows immediately from Theorem 1.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bryant D, Steel M (1995) Extension operations on sets of leaf-labeled trees. *Adv Appl Math* 16:425–453. <https://doi.org/10.1006/aama.1995.1020>
- Deng Y, Fernández-Baca D (2018) Fast compatibility testing for rooted phylogenetic trees. *Algorithmica* 80:2453–2477. <https://doi.org/10.1007/s00453-017-0330-4>
- Geiß M, Chávez E, González Laffitte M, López Sánchez A, Stadler BMR, Valdivia DI, Hellmuth M, Hernández Rosales M, Stadler PF (2019) Best match graphs. *J Math Biol* 78:2015–2057. <https://doi.org/10.1007/s00285-019-01332-9>
- Geiß M, González Laffitte ME, López Sánchez A, Valdivia DI, Hellmuth M, Hernández Rosales M, Stadler PF (2020) Best match graphs and reconciliation of gene trees with species trees. *J Math Biol* 80:1459–1495. <https://doi.org/10.1007/s00285-020-01469-y>
- Schaller D, Geiß M, Stadler PF, Hellmuth M (2021) Complete characterization of incorrect orthology assignments in best match graphs. *J Math Biol*. <https://doi.org/10.1007/s00285-021-01564-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

David Schaller¹ · Manuela Geiß² · Edgar Chávez³ · Marcos González Laffitte⁴ · Alitzel López Sánchez⁵ · Bärbel M. R. Stadler¹ · Dulce I. Valdivia⁶ · Marc Hellmuth⁷ · Maribel Hernández Rosales⁶ · Peter F. Stadler^{1,8,9,10,11,12} 

David Schaller
sdavid@bioinf.uni-leipzig.de

Manuela Geiß
manuela.geiss@scch.at

Edgar Chávez
echavezaparcio@gmail.com

Marcos González Laffitte
marcoslaffitte@im.unam.mx

Alitzel López Sánchez
Alitzel.Lopez.Sanchez@USherbrooke.ca

Bärbel M. R. Stadler
baer@bioinf.uni-leipzig.de

Dulce I. Valdivia
dulce.valdivia@cinvestav.mx

Marc Hellmuth
mhellmuth@mailbox.org

Maribel Hernández Rosales
maribel.hr@cinvestav.mx

- ¹ Max-Planck-Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany
- ² Software Competence Center Hagenberg GmbH, Softwarepark 21, 4232 Hagenberg, Austria
- ³ Centro de Física Aplicada y Tecnología Avanzada, UNAM, 76230 Juriquilla, QRO, México
- ⁴ Instituto de Matemáticas, UNAM Juriquilla, Blvd. Juriquilla 3001, 76230 Juriquilla, Querétaro, QRO, México
- ⁵ Department of Computer Science, Université de Sherbrooke, 2500 Boul. de l'Université, Sherbrooke J1K 2R1, Canada
- ⁶ Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV), Irapuato Unit, Libramiento Norte, Carretera Panamericana Irapuato-León Kilómetro 9.6, 36821 Irapuato, Guanajuato, México
- ⁷ Department of Mathematics, Faculty of Science, Stockholm University, SE - 106 91 Stockholm, Sweden
- ⁸ Bioinformatics Group, Department of Computer Science, Interdisciplinary Center of Bioinformatics, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Competence Center for Scalable Data Services and Solutions, Leipzig, Germany
- ⁹ Leipzig Research Center for Civilization Diseases, Leipzig University, Härtelstraße 16-18, 04107 Leipzig, Germany
- ¹⁰ Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Wien, Austria
- ¹¹ Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Colombia
- ¹² Santa Fe Institute, 1399 Hyde Park Rd. , Santa Fe, NM 87501, USA