Forum Math. **24** (2012), 495–505
DOI 10.1515/FORM.2011.070
Forum Mathematicum

Phylogenetic analysis and mod 2 homology

Andreas W. M. Dress and Yaokun Wu

Communicated by Andrew Ranicki

Abstract. Given a finite simplicial complex $\mathcal{K}$, we construct a chain-complex isomorphism from the simplicial chain complex of $\mathcal{K}$ over $\mathbb{F}_2$ endowed with the standard *boundary operator* $\partial_{\mathcal{K}}$ (that maps any simplex $A \in \mathcal{K}$ onto the sum of all of its maximal subsets) to that same complex endowed with the *incidence operator* $\mathcal{I}_{\mathcal{K}}$ that maps any simplex $A \in \mathcal{K}$ to the sum of just *all* of its proper subsets. We also indicate why this became of interest in the context of the theory of phylogenetic diversity.

Keywords. Simplicial complexes, mod 2 homology, boundary operator, incidence operator, phylogenetic diversity.

2010 Mathematics Subject Classification. 05E45, 15A24.

1 Introduction

Given a collection X of extant species, a *clade* C in X is a subset of X that consists of all species in X that are offspring of a single ancestral species while none of the species in the complement $X \setminus C$ of C have evolved from this ancestral species. One of the most basic tasks in phylogenetic analysis is, given any set X of species, to identify the collection of all clades in X (of which, by the way, there can be at most $|X|-2$ distinct non-trivial ones). Yet, as *Charles Darwin* put it in his treatise *The descent of man, and selection in relation to sex*: "As we have no record of the lines of descent, the pedigree can be discovered only by the degrees of resemblance between the beings which are to be classed." That is, all that we commonly can rely on to identify the collection of all clades in X is information about how distinct, or how similar, the present-day species are that make up the set X.

Consequently, a standard assumption in phylogenetic analysis is that, together with a finite set X of species or, more generally, of any kind of taxonomic units (for short, *taxa*), we are given a metric D defined on X that quantifies that *degree of resemblance between* the taxa contained in X. In other words, one assumes that

The work of the first author was supported by CAS, MPG, and BMBF; the work of the second author was supported by STCSM (No. 08QA14036, 09XD1402500), NNSFC (No. 10871128), Chinese Ministry of Education (No. 108056) and MIS MPG. The corresponding author is Y. Wu.

one is given a map

$$D : X \times X \to \mathbb{R} : (x, y) \mapsto D(x, y)$$

from $X \times X$ into the set $\mathbb{R}$ of real numbers for which $D(x, x) = 0$ and $D(x, y) \leq D(x, z) + D(y, z)$ holds for any three taxa x, y, z under consideration. And the task one has to address can then be described as that of designing methods for deriving, from these data, a *phylogenetic X-tree* $T = T(D)$ that – at least approximately – represents the map D. That is, one wants to find a finite edge-weighted and X-labeled tree $T = (V, E, \ell; \varphi)$ consisting of

- a vertex set V and an edge set $E \subseteq \binom{V}{2}$,
- a *weight map* $\ell : E \to \mathbb{R}_{>0}$ from E into the set $\mathbb{R}_{>0}$ of positive real numbers,
- and a *labeling map* $\varphi : X \to V$ whose image $\varphi(X)$ contains – at least – all vertices in V of degree 1 or 2

such that the distance $D(x, y)$ of any two taxa x, y in X coincides – at least approximately – with the length $\ell_T\big(\varphi(x), \varphi(y)\big)$ of the unique path in T from $\varphi(x)$ to $\varphi(y)$ relative to ℓ (cf. for example [23] for a thorough discussion of this concept that was, and still is, one of the focal points for all conceptual development in computational phylogenetics).

Remarkably, denoting the set consisting of all *splits* $A|B$ of X – i.e. the set of all 2-element subsets $\{A, B\}$ of the power set $\mathcal{P}(X)$ of X for which $A \cup B = X$ and $A \cap B = \emptyset$ holds – by $\mathcal{S}(X)$, this task is simply equivalent to finding a map Σ from $\mathcal{S}(X)$ into the set $\mathbb{R}_{\geq 0}$ of non-negative real numbers such that

(i) the distance $D(x, y)$ of x and y coincides – again at least approximately – with the sum $\Sigma(x : y) := \sum_{A|B \in \mathcal{S}(X):\, x \in A,\, y \in B} \Sigma(A|B)$ over all splits $A|B$ of X that "separate" x from y,

(ii) $\Sigma(\{\emptyset, X\}) = 0$ holds, and

(iii) any two splits in the support $\operatorname{supp}(\Sigma) := \big\{A|B \in \mathcal{S}(X) : \Sigma(A|B) \neq 0\big\}$ of Σ are *compatible* – i.e. one of the four intersections $A \cap A'$, $A \cap B'$, $B \cap A'$, $B \cap B'$ is empty for any two splits $A|B$ and $A'|B'$ in $\operatorname{supp}(\Sigma)$.

This fact, also discussed *in extenso* in [23], was probably folklore already in the mid-twentieth century, in one or the other disguise; it was stated explicitly, more or less just as stated above, by Buneman around 1970 (cf. for instance [7]); and it has also been another one of those fundamental insights on which much further development of computational phylogenetics was based.

As observed in [8], one can further re-interpret this task in the context of a rather naturally defined injective $\mathbb{R}$-linear map $\mathbf{D}$ from the $\mathbb{R}$-vectorspace $\mathbb{R}^{\mathcal{S}(X)}$

consisting of all $\mathbb{R}$-*weighted split systems over* X into the $\mathbb{R}$-vectorspace $\mathbb{R}^{\mathcal{P}(X)}$ consisting of all $\mathbb{R}$-*weighted set systems over* X, i.e., all maps from the power set $\mathcal{P}(X)$ of X into $\mathbb{R}$ that maps any given $\mathbb{R}$-weighted split system Σ to the $\mathbb{R}$-weighted set system $\Sigma_\bullet$ that associates, to any subset Y of X, the "weight"

$$\Sigma_\bullet(Y) := \sum_{A \in \mathcal{P}(X \setminus Y)} \Sigma(\{Y \cup A, X \setminus (Y \cup A)\}).$$

Note that $D_\Sigma(x, y)$ apparently coincides, for all $x, y \in X$, with the difference between the "total weight" $|\Sigma| := \sum_{S \in \mathcal{S}(X)} \Sigma(S)$ of Σ and the value $\Sigma_\bullet(\{x, y\})$ that $\Sigma_\bullet$ attains at the subset $\{x, y\}$ of X.

More generally (see also [8]–[16]), defining the *phylogenetic diversity* $\Sigma^\bullet(Y)$ of a subset Y of X relative to Σ in case X is a collection of species (that is, the "ecological value" retained when all species in X except those in Y became extinct) by

$$\Sigma^\bullet(Y) := \sum_{A|B \in \mathcal{S}(X), Y \cap A \neq \emptyset \neq B \cap Y} \Sigma(A|B)$$

relative to an $\mathbb{R}$-weighted split system $\Sigma : \mathcal{S}(X) \to \mathbb{R}$, we have, just as above, $\Sigma^\bullet(Y) + \Sigma_\bullet(Y) = |\Sigma|$, implying that there is a close correlation between the map $\mathbf{D}$ and the concept of phylogenetic diversity.

In view of these facts, it seemed to be of some interest to abstractly characterize the image of $\mathbf{D}$. To this end, recall first from [8] that the above set-up allows to replace $\mathbb{R}$ by any field F in all definitions and assertions above (as long, of course, as they do not refer to $\mathbb{R}_{\geq 0}$), attaching the letter F as an upper index wherever required to indicate reference to the field F. Replacing $\mathbb{R}$ wherever possible by F from now on, one may then also recall from [8] that there exists a canonically defined F-linear involution τ of $F^{\mathcal{P}(X)}$ that maps any F-valued set system $\Pi \in F^{\mathcal{P}(X)}$ onto the F-valued set system $\widehat{\Pi}$ which associates, to any given subset A of X, the alternating sum

$$\widehat{\Pi}(A) := \sum_{A' \in \mathcal{P}(A)} (-1)^{|A'|}\, \Pi(A')$$

whose fixed-point space $(F^{\mathcal{P}(X)})^\tau$ coincides with the image $\mathbf{D}(F^{\mathcal{S}(X)})$ of the map $\mathbf{D} = \mathbf{D}^F$, and also with the set of all F-valued set systems of the form $\Phi = \Pi + \tau(\Pi)$ for some F-valued set system $\Pi \in F^{\mathcal{P}(X)}$.

Moreover, given a *simplicial complex* over X, say $\mathcal{K} \subseteq \mathcal{P}(X)$, i.e., a non-empty subset of the power set $\mathcal{P}(X)$ of X such that $B \in \mathcal{K}$ holds for every subset B of a set A in $\mathcal{K}$, one may consider the subspace $F[\mathcal{K}]$ of $F^{\mathcal{P}(X)}$ consisting of all maps $\Pi \in F^{\mathcal{P}(X)}$ whose support $\mathrm{supp}(\Pi)$ is contained in $\mathcal{K}$. It is obvious that

$\tau(\Pi) \in F[\mathcal{K}]$ holds for all $\Pi \in F[\mathcal{K}]$, implying that τ induces also an involution $\tau_{\mathcal{K}} = \tau_{\mathcal{K}}^F$ when restricted to $F[\mathcal{K}]$.

So, to develop a better understanding of the map $\mathbf{D}$ and the involution τ, it seemed a sensible task to study the family of involutions $\tau_{\mathcal{K}} = \tau_{\mathcal{K}}^F$ indexed by the simplicial complexes $\mathcal{K} \subseteq \mathcal{P}(X)$ over X (and the field F), and the associated chain complex $F[\mathcal{K}] \to F[\mathcal{K}] \to F[\mathcal{K}]$ where the first map is given by the operator $\mathcal{D}_{\mathcal{K}} = \mathcal{D}_{\mathcal{K}}^F := \tau_{\mathcal{K}}^F - \mathrm{Id}_{F[\mathcal{K}]}$, and the second one by the map $\tau_{\mathcal{K}} + \mathrm{Id}_{F[\mathcal{K}]}$. The kernel of the first map is the fixed-point space of $\tau_{\mathcal{K}}^F$ and, hence, the image of $\mathbf{D}^F$. However, neither the image of $\mathcal{D}_{\mathcal{K}}$, nor the kernel or the image of $\tau_{\mathcal{K}} + \mathrm{Id}_{F[\mathcal{K}]}$, can be described that easily unless we assume that the characteristic char F of F is distinct from 2 in which case the image of $\mathcal{D}_{\mathcal{K}}$ must coincide with the kernel of $\tau_{\mathcal{K}} + \mathrm{Id}_{F[\mathcal{K}]}$, that is, the fixed-point space of the involution $-\tau_{\mathcal{K}}$, as in this case $\mathcal{D}_{\mathcal{K}}(\frac{-\Pi}{2}) = \tau_{\mathcal{K}}(\frac{-\Pi}{2}) - (\frac{-\Pi}{2}) = \Pi$ must hold for every F-valued set system Π in the fixed-point space of $-\tau_{\mathcal{K}}$.

However, while the two operators $\mathcal{D}_{\mathcal{K}}$ and $\tau_{\mathcal{K}} + \mathrm{Id}_{F[\mathcal{K}]}$ coincide in case $F = \mathbb{F}_2$ (or, more generally, whenever char $F = 2$ holds), the image $\mathrm{Im}(\mathcal{D}_{\mathcal{K}})$ of $\mathcal{D}_{\mathcal{K}}$ (that is always contained in its kernel $\mathrm{Ker}(\mathcal{D}_{\mathcal{K}})$) may, in general, not coincide with that kernel in case $F = \mathbb{F}_2$. Instead, and that seemed actually quite surprising to us when we began to conjecture this fact based on some small examples, the following turned out to be true:

Theorem 1.1. *For any finite simplicial complex $\mathcal{K}$, the image $\mathrm{Im}(\mathcal{D}_{\mathcal{K}})$ of $\mathcal{D}_{\mathcal{K}}$ and its kernel $\mathrm{Ker}(\mathcal{D}_{\mathcal{K}})$ coincide if and only if $\mathcal{K}$ is acyclic over $\mathbb{F}_2$. More precisely, the quotient $R(\mathcal{K}) := \mathrm{Ker}(\mathcal{D}_{\mathcal{K}})/\mathrm{Im}(\mathcal{D}_{\mathcal{K}})$ is isomorphic – as a vector space over $\mathbb{F}_2$ – to the direct sum $H^*(\mathcal{K}, \mathbb{F}_2) := \bigoplus_i H^i(\mathcal{K}, \mathbb{F}_2)$ of all the cohomology groups $H^i(\mathcal{K}, \mathbb{F}_2)$ of $\mathcal{K}$ over $\mathbb{F}_2$.*

In this note, we will establish this fact by explicitly constructing an appropriate chain complex isomorphism (depending on an arbitrarily chosen linear order of X) between the corresponding dual complexes, i.e., the pair $(\mathbb{F}_2[\mathcal{K}], \partial_{\mathcal{K}})$ and the pair $(\mathbb{F}_2[\mathcal{K}], \mathcal{I}_{\mathcal{K}})$ where $\partial_{\mathcal{K}}$ is the standard boundary operator that maps every single simplex $A \in \mathcal{K}$ onto the sum of all of its maximal subsets and $\mathcal{I}_{\mathcal{K}}$ is the $\mathbb{F}_2$-linear operator that maps every single simplex $A \in \mathcal{K}$ onto the sum of all of its proper subsets.

We would like to point out, however, that we do not expect this fact to be of any immediate relevance in phylogenetic analysis (even though there may be some interesting applications in the context of the work presented in [8] and [15], two papers where also mod 2 computations with splits turn out to be of some interest). Thus, the rather lengthy and perhaps a bit circumstantial introduction above should not be misunderstood as a (necessarily vain) attempt to bolster our results by claiming any biological significance – it is meant exclusively to give a historical account of how we were actually lead to stumble over and then conjecture (and finally establish) our result. Yet, we consider, after all, the result (as well as its proof) to be of some independent interest in its own right – in particular as it offers the option of reconsidering cap products and (co-)homology operations in the setting defined by the operator $\mathcal{D}_{\mathcal{K}}$ – work that we hope either us or somebody else to find the time to do in the foreseeable future.

2 Notations, definitions, and preliminary results

Let X be a finite set and $\mathcal{K} \subseteq \mathcal{P}(X)$ be a simplicial complex. Given a subset A of X, let δ_A denote the map from $\mathcal{P}(X)$ into $\mathbb{F}_2$ that maps any $B \subseteq X$ to $\delta_{A,B}$ where δ here stands for the Kronecker delta function, and recall that $\mathbb{F}_2[\mathcal{K}]$ can be considered as a commutative ring, the *Stanley–Reisner ring* associated with $\mathcal{K}$ with coefficients in $\mathbb{F}_2$, whose multiplication "$*$" given by their convolution (or "Faltung")

$$\Pi * \Psi := \sum_{A \in \mathcal{K}} \sum_{B \subseteq A} \Pi(A \setminus B)\, \Psi(B)\, \delta_A$$

for all $\Pi, \Psi \in \mathbb{F}_2[\mathcal{K}]$. Clearly, $\delta_\emptyset$ is the multiplicative unit $\mathbf{1}_{\mathcal{K}} = \mathbf{1}_{\mathbb{F}_2[\mathcal{K}]}$ of $\mathbb{F}_2[\mathcal{K}]$ and, of course, the "all-zero map" $\mathbf{0} = \mathbf{0}_{\mathcal{K}}$ is the additive unit. The set $\{\delta_A : A \in \mathcal{K}\}$ forms a canonical basis of $\mathbb{F}_2[\mathcal{K}]$ and we have $\Pi = \sum_{A \in \mathcal{K}} \Pi(A)\delta_A$ for every $\Pi \in \mathbb{F}_2[\mathcal{K}]$. To simplify notation, we will often identify a subset A of X with the associated map $\delta_A \in \mathbb{F}_2[\mathcal{K}]$ in case $A \in \mathcal{K}$ holds, and with $\mathbf{0}$ in case A is not contained in $\mathcal{K}$ so that $\mathcal{K}$ is identified with the basis $\{\delta_A : A \in \mathcal{K}\}$ of $\mathbb{F}_2[\mathcal{K}]$. Observe that $A * B = A \cup B$ holds for all $A, B \in \mathcal{K}$ with $A \cap B = \emptyset$, and $A * B = \mathbf{0}$ else.

Now, recall (cf. [19]) that the *boundary operator* $\partial_{\mathcal{K}} : \mathbb{F}_2[\mathcal{K}] \rightarrow \mathbb{F}_2[\mathcal{K}]$ is the linear map from $\mathbb{F}_2[\mathcal{K}]$ into itself that maps any $\Pi \in \mathbb{F}_2[\mathcal{K}]$ to the map $\partial_{\mathcal{K}}\Pi$ defined by putting $\partial_{\mathcal{K}}\Pi\,(A) := \sum_{b \in X \setminus A} \Pi(A \cup b)$ for every subset $A \subseteq X$ (where we use the convention of writing x rather than $\{x\}$ for every singleton set $\{x\}$ of X) and, hence, every basis element $A \in \mathcal{K}$ of $\mathbb{F}_2[\mathcal{K}]$ to the sum $\partial_{\mathcal{K}}(A) = \sum_{b \in A}(A \setminus b)$. Note that $\partial_{\mathcal{K}}(\emptyset) = \mathbf{0}$.

Here, we will compare the boundary operator $\partial_{\mathcal{K}}$ with the *incidence operator $\mathcal{I}_{\mathcal{K}}$*, another map from $\mathbb{F}_2[\mathcal{K}]$ into itself that maps every set in $\mathcal{K}$ onto the sum of all of its proper subsets, i.e., it maps every basis element $A \in \mathcal{K}$ of $\mathbb{F}_2[\mathcal{K}]$ to the sum $\mathcal{I}_{\mathcal{K}}(A) = \sum_{B \subsetneq A} B$ and, hence, any $\Pi \in \mathbb{F}_2[\mathcal{K}]$ to the map $\mathcal{I}_{\mathcal{K}}\Pi \in \mathbb{F}_2[\mathcal{K}]$ defined by putting $(\mathcal{I}_{\mathcal{K}}\Pi)\,(A) := \sum_{\emptyset \neq B \subseteq X \setminus A} \Pi(A \cup B)$ for every subset $A \subseteq X$.

It is well known and easy to see that

$$\partial_{\mathcal{K}}^2 = 0 \tag{2.1}$$

holds. In view of $\operatorname{char}\mathbb{F}_2 = 2$ and the fact that $\tau_{\mathcal{K}}^{\mathbb{F}_2} := \tau_{\mathcal{K}}$ is an involution, we see that also

$$\mathcal{D}_{\mathcal{K}}^2 = (\tau_{\mathcal{K}} - \mathrm{Id}_{\mathbb{F}_2[\mathcal{K}]}) \circ (\tau_{\mathcal{K}} - \mathrm{Id}_{\mathbb{F}_2[\mathcal{K}]}) = {\tau_{\mathcal{K}}}^2 + \mathrm{Id}_{\mathbb{F}_2[\mathcal{K}]}^2 = 0$$

holds. Since $\mathcal{I}_{\mathcal{K}}$ is clearly the adjoint of $\mathcal{D}_{\mathcal{K}}$, meaning that the matrix representation for them with respect to the canonical basis of $\mathbb{F}_2[\mathcal{K}]$ are the transpose to each other, it follows immediately that also

$$\mathcal{I}_{\mathcal{K}}^2 = 0 \tag{2.2}$$

holds – a fact that can also be established very easily by direct verification.

The following fact follows immediately from our definitions:

Lemma 2.1. *Given any two disjoint subsets $A, B \subseteq X$ with $A \cup B \in \mathcal{K}$, we have*

$$\mathcal{I}_{\mathcal{K}}(A * B) = \sum_{C \subsetneq A \cup B} C = A * \mathcal{I}_{\mathcal{K}}(B) + \mathcal{I}_{\mathcal{K}}(A) * B + \mathcal{I}_{\mathcal{K}}(A) * \mathcal{I}_{\mathcal{K}}(B).$$

In particular, we have $\mathcal{I}_{\mathcal{K}}(A * b) = A + \mathcal{I}_{\mathcal{K}}(A) * (\emptyset + b)$ for every $A \subsetneq X$ and $b \in X \setminus A$ with $A \cup b \in \mathcal{K}$ and, hence, by linearity also

$$\mathcal{I}_{\mathcal{K}}(\Pi * b) = \Pi + \mathcal{I}_{\mathcal{K}}(\Pi) * (\emptyset + b)$$

for every $\Pi \in \mathbb{F}_2[\mathcal{K}]$ and $b \in X$ for which $A \in \operatorname{supp}(\Pi)$ implies $b \notin A$ and $A \cup b \in \mathcal{K}$.

In the next section, we will show that the two chain-complexes $(\mathbb{F}_2[\mathcal{K}], \partial_{\mathcal{K}})$ and $(\mathbb{F}_2[\mathcal{K}], \mathcal{I}_{\mathcal{K}})$ are actually isomorphic.

3 A chain-complex isomorphism from $(\mathbb{F}_2[\mathcal{K}], \partial_{\mathcal{K}})$ to $(\mathbb{F}_2[\mathcal{K}], \mathcal{I}_{\mathcal{K}})$

Consider an arbitrary but fixed linear ordering "$<$" of the elements in X, and denote by $\max(A)$, for any non-empty subset A of X, the unique maximal element of A relative to that order, and by $\widehat{A}$ the "decapitated" set $A \setminus \max(A)$. We define a linear map $\omega = \omega_{\mathcal{K}}$ from $\mathbb{F}_2[\mathcal{K}]$ into itself recursively by putting

$$\omega(A) := \begin{cases} \mathbf{0}, & \text{if } A = \emptyset, \\ \widehat{A} + \omega(\widehat{A}) * \big(\emptyset + \max(A)\big), & \text{if } |A| \text{ is odd}, \\ \omega(\widehat{A}) * \max(A), & \text{otherwise}, \end{cases}$$

for every $A \in \mathcal{K}$. Then, we have the following theorem.

Theorem 3.1. *The identity*

$$(\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(A) = \omega\big(\partial_{\mathcal{K}}(A)\big) + \mathcal{I}_{\mathcal{K}}\big(\omega(A)\big) = (\omega \circ \partial_{\mathcal{K}} + \mathcal{I}_{\mathcal{K}} \circ \omega)(A)$$

holds for every $A \in \mathcal{K}$.

Proof. We proceed by induction on $n := |A|$. If A is the empty set, both sides clearly vanish. Now assume that $n = |A| > 0$ holds. Put $a := \max(A)$. Note that $\mathcal{I}_{\mathcal{K}}(A) + \widehat{A} = \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \mathcal{I}_{\mathcal{K}}(\widehat{A}) * a$ as well as $\partial_{\mathcal{K}}(A) + \widehat{A} = \partial_{\mathcal{K}}(\widehat{A}) * a$ and, therefore, also

$$(\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(A) = \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \mathcal{I}_{\mathcal{K}}(\widehat{A}) * a + \partial_{\mathcal{K}}(\widehat{A}) * a \tag{3.1}$$

as well as

$$\omega(\partial_{\mathcal{K}}(A) + \widehat{A}) = \omega(\partial_{\mathcal{K}}(\widehat{A}) * a) \tag{3.2}$$

hold.

If n is even, equation (3.2) yields

$$\begin{aligned}
\omega\big(\partial_{\mathcal{K}}(A)\big) + \omega(\widehat{A}) = \omega\big(\partial_{\mathcal{K}}(\widehat{A}) * a\big) &= \omega\Bigg(\sum_{b \in \widehat{A}} (\widehat{A} \setminus b) * a\Bigg) \\
&= \sum_{b \in \widehat{A}} \omega\big((\widehat{A} \setminus b) * a\big) \\
&= \sum_{b \in \widehat{A}} \big((\widehat{A} \setminus b) + \omega(\widehat{A} \setminus b) * (\emptyset + a)\big) \\
&= \sum_{b \in \widehat{A}} (\widehat{A} \setminus b) + \omega\Bigg(\sum_{b \in \widehat{A}} (\widehat{A} \setminus b)\Bigg) * (\emptyset + a)
\end{aligned}$$

and, therefore, also $(\omega \circ \partial_{\mathcal{K}})(A) = \partial_{\mathcal{K}}(\widehat{A}) + \omega(\partial_{\mathcal{K}}(\widehat{A})) * (\emptyset + a) + \omega(\widehat{A})$. Further, we have $(\mathcal{I}_{\mathcal{K}} \circ \omega)(A) = \mathcal{I}_{\mathcal{K}}(\omega(\widehat{A}) * a) = \omega(\widehat{A}) + \mathcal{I}_{\mathcal{K}}\big(\omega(\widehat{A})\big) * (\emptyset + a)$ by the definition of ω and Lemma 2.1. Thus, invoking our induction hypothesis $(\omega \circ \partial_{\mathcal{K}} + \mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A}) = (\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(\widehat{A})$, this implies our claim:

$$\begin{aligned}
(\omega \circ \partial_{\mathcal{K}} + \mathcal{I}_{\mathcal{K}} \circ \omega)(A) &= \big(\partial_{\mathcal{K}}(\widehat{A}) + \omega(\partial_{\mathcal{K}}(\widehat{A})) * (\emptyset + a) + \omega(\widehat{A})\big) \\
&\qquad + \big(\omega(\widehat{A}) + \mathcal{I}_{\mathcal{K}}\big(\omega(\widehat{A})\big) * (\emptyset + a)\big) \\
&= \partial_{\mathcal{K}}(\widehat{A}) + \big((\omega \circ \partial_{\mathcal{K}})(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A})\big) * (\emptyset + a) \\
&= \partial_{\mathcal{K}}(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(\widehat{A}) * (\emptyset + a) \\
&= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \mathcal{I}_{\mathcal{K}}(\widehat{A}) * a + \partial_{\mathcal{K}}(\widehat{A}) * a \\
&= (\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(A) \quad \text{(by equation (3.1))}.
\end{aligned}$$

If n is odd, we have

$$(\omega \circ \partial_{\mathcal{K}})(A) = \omega(\partial_{\mathcal{K}}(\widehat{A}) * a) + \omega(\widehat{A}) = \omega\big(\partial_{\mathcal{K}}(\widehat{A})\big) * a + \omega(\widehat{A}).$$

In addition, $\omega(A) := \widehat{A} + \omega(\widehat{A}) * \big(\emptyset + a\big)$ together with Lemma 2.1, applied to $f := \omega(\widehat{A})$ and $b := a$, gives

$$\begin{aligned}(\mathcal{I}_{\mathcal{K}} \circ \omega)(A) &= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \mathcal{I}_{\mathcal{K}}\big(\omega(\widehat{A}) * (\emptyset + a)\big)\\ &= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A}) + \mathcal{I}_{\mathcal{K}}\big(\omega(\widehat{A}) * a\big)\\ &= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A}) + \omega(\widehat{A}) + \mathcal{I}_{\mathcal{K}}\big(\omega(\widehat{A})\big) * (\emptyset + a)\\ &= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \omega(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A}) * a,\end{aligned}$$

and our induction hypothesis still reads $(\omega \circ \partial_{\mathcal{K}} + \mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A}) = (\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(\widehat{A})$. So, putting things together, we get

$$\begin{aligned}(\omega \circ \partial_{\mathcal{K}} + \mathcal{I}_{\mathcal{K}} \circ \omega)(A) &= \big(\omega\big(\partial_{\mathcal{K}}(\widehat{A})\big) * a + \omega(\widehat{A})\big)\\ &\qquad + \big(\mathcal{I}_{\mathcal{K}}(\widehat{A}) + \omega(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A}) * a\big)\\ &= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \big((\omega \circ \partial_{\mathcal{K}})(\widehat{A}) + (\mathcal{I}_{\mathcal{K}} \circ \omega)(\widehat{A})\big) * a\\ &= \mathcal{I}_{\mathcal{K}}(\widehat{A}) + \big(\mathcal{I}_{\mathcal{K}}(\widehat{A}) + \partial_{\mathcal{K}}(\widehat{A})\big) * a\\ &= (\mathcal{I}_{\mathcal{K}} + \partial_{\mathcal{K}})(A) \quad \text{(by equation (3.1))},\end{aligned}$$

just as claimed above. □

Put $\alpha_{\mathcal{K}} := \mathrm{Id}_{\mathbb{F}_2[\mathcal{K}]} + \omega_{\mathcal{K}}$. Due to the fact that $\omega_{\mathcal{K}}$ is nilpotent, the $\mathbb{F}_2$-linear endomorphism $\alpha_{\mathcal{K}}$ of $\mathbb{F}_2[\mathcal{K}]$ must be an $\mathbb{F}_2$-linear isomorphism whose inverse is given by $\beta_{\mathcal{K}} = \alpha_{\mathcal{K}}^{-1} := \mathrm{Id}_{\mathbb{F}_2[\mathcal{K}]} + \omega_{\mathcal{K}} + \omega_{\mathcal{K}}^2 + \omega_{\mathcal{K}}^3 + \cdots$. We remark that Theorem 3.1 says nothing but

$$\alpha_{\mathcal{K}} \circ \partial_{\mathcal{K}} = \mathcal{I}_{\mathcal{K}} \circ \alpha_{\mathcal{K}}. \tag{3.3}$$

Let $\mathcal{K}' \subseteq \mathcal{P}(X)$ be a simplicial complex over X that is a subcomplex of $\mathcal{K}$. When restricting to $\mathbb{F}_2[\mathcal{K}']$, the actions of $\alpha_{\mathcal{K}}$, $\mathcal{I}_{\mathcal{K}}$ and $\partial_{\mathcal{K}}$ coincide with $\alpha_{\mathcal{K}'}$, $\mathcal{I}_{\mathcal{K}'}$ and $\partial_{\mathcal{K}'}$, respectively. We hence obtain obvious induced linear maps $\alpha_{\mathcal{K}/\mathcal{K}'}$, $\mathcal{I}_{\mathcal{K}/\mathcal{K}'}$ and $\partial_{\mathcal{K}/\mathcal{K}'}$ from $\mathbb{F}_2[\mathcal{K}/\mathcal{K}'] := \mathbb{F}_2[\mathcal{K}]/\mathbb{F}_2[\mathcal{K}']$ into itself, where $\alpha_{\mathcal{K}/\mathcal{K}'}$ is an $\mathbb{F}_2$-linear isomorphism of $\mathbb{F}_2[\mathcal{K}/\mathcal{K}']$. It follows from equation (3.3) that

$$\alpha_{\mathcal{K}/\mathcal{K}'} \circ \partial_{\mathcal{K}/\mathcal{K}'} = \mathcal{I}_{\mathcal{K}/\mathcal{K}'} \circ \alpha_{\mathcal{K}/\mathcal{K}'}. \tag{3.4}$$

As an immediate consequence of equations (2.1), (2.2) and (3.4), we come to our main result:

Theorem 3.2. *The endomorphism* $\alpha_{\mathcal{K}/\mathcal{K}'} : \mathbb{F}_2[\mathcal{K}/\mathcal{K}'] \to \mathbb{F}_2[\mathcal{K}/\mathcal{K}']$ *is a chain-complex isomorphism from* $(\mathbb{F}_2[\mathcal{K}/\mathcal{K}'], \partial_{\mathcal{K}/\mathcal{K}'})$ *to* $(\mathbb{F}_2[\mathcal{K}/\mathcal{K}'], \mathcal{J}_{\mathcal{K}/\mathcal{K}'})$ *and induces therefore a canonical isomorphism*

$$H_{\mathbb{F}_2}(\mathcal{K}/\mathcal{K}' \mid \partial_{\mathcal{K}/\mathcal{K}'}) \cong H_{\mathbb{F}_2}(\mathcal{K}/\mathcal{K}' \mid \mathcal{J}_{\mathcal{K}/\mathcal{K}'})$$

between the homology group

$$H_{\mathbb{F}_2}(\mathcal{K},\ \mathcal{K}' \mid \partial_{\mathcal{K}}) := \mathrm{Ker}(\partial_{\mathcal{K}/\mathcal{K}'}) / \mathrm{Im}(\partial_{\mathcal{K}/\mathcal{K}'})$$

of the relative simplicial complex $\mathcal{K}/\mathcal{K}'$ *with coefficients in* $\mathbb{F}_2$ *relative to the boundary operator* $\partial_{\mathcal{K}}$ *and the corresponding homology group*

$$H_{\mathbb{F}_2}(\mathcal{K},\ \mathcal{K}' \mid \mathcal{J}_{\mathcal{K}}) := \mathrm{Ker}(\mathcal{J}_{\mathcal{K}/\mathcal{K}'}) / \mathrm{Im}(\mathcal{J}_{\mathcal{K}/\mathcal{K}'})$$

relative to the incidence operator $\mathcal{J}_{\mathcal{K}}$.

To see that Theorem 3.2 implies Theorem 1.1, we only need to take $\mathcal{K}' = \emptyset$ and notice that $\mathrm{Ker}(\mathcal{J}_{\mathcal{K}}) / \mathrm{Im}(\mathcal{J}_{\mathcal{K}})$ vanishes if and only if so does $\mathrm{Ker}(\mathcal{D}_{\mathcal{K}}) / \mathrm{Im}(\mathcal{D}_{\mathcal{K}})$.

In further work, we would like to discuss potential consequences of Theorem 3.2 relating e.g. to the canonical grading of $\mathbb{F}_2(\mathcal{K})$ and $R(\mathcal{K})$ in terms of the subgroups generated by the subsets $A \in \mathcal{K}$ of cardinality bounded from above by a given number k, the dual (co-homology) setting, cap products, and (co-)homology operations.

Bibliography

[1] J. Backelin and S. Linusson, Parity splits of X-trees, *Annals of Combinatorics* **10** (2006), 1–18.

[2] H.-J. Bandelt, Recognition of tree metrics, *SIAM Journal on Discrete Mathematics* **3** (1990), 1–6.

[3] H.-J. Bandelt and A. W. M. Dress, A canonical split decomposition theory for metrics on a finite set, *Advances in Mathematics* **92** (1992), 47–105.

[4] H.-J. Bandelt and A. W. M. Dress, Split decomposition: A new and useful approach to phylogenetic analysis of distance data, *Molecular Phylogenetics and Evolution* **1** (1992), 242–252.

[5] H.-J. Bandelt and M. A. Steel, Symmetric matrices representable by weighted trees over a cancellative abelian monoid, *SIAM Journal on Discrete Mathematics* **8** (1995), 517–525.

[6] G. M. Barker, Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation, *Biological Journal of the Linnean Society* **76** (2002), 165–194.

[7] P. Buneman, The recovery of trees from measures of dissimilarity, in: *Mathematics in the Archaeological and Historical Sciences*, pp. 387–395, edited by F. R. Hodson, D. G. Kendall and P. Tautu, Edinburgh University Press, Edinburgh, 1971.

[8] A. W. M. Dress, Phylogenetic analysis, split systems and Boolean functions, in: *Aspects of Mathematical Modelling*, pp. 235–248, edited by R. J. Hosking and E. Venturino, Mathematics and Biosciences in Interaction, Birkhäuser-Verlag, 2008.

[9] A. W. M. Dress, Even set systems, *Annals of Combinatorics* **13** (2009), 193–198.

[10] A. W. M. Dress, Split decomposition over an Abelian group. Part I: Generalities, *Annals of Combinatorics* **13** (2009), 199–232.

[11] A. W. M. Dress, Split decomposition over an Abelian group. Part II: Group-valued split systems with weakly compatible support, *Discrete Applied Mathematics* **157** (2009), no. 10, 2349–2360.

[12] A. W. M. Dress, A note on group-valued split and set systems, to appear in *Contributions to Discrete Mathematics*.

[13] A. W. M. Dress, Split decomposition over an Abelian group. Part III: Group-valued split systems with compatible support, manuscript, Shanghai.

[14] A. W. M. Dress, K. Huber and V. Moulton, Some uses of the Farris Transform in Mathematics and Phylogenetics – A Review, *Annals of Combinatorics* **11** (2007), 1–37.

[15] A. W. M. Dress and M. A. Steel, Mapping edge sets to splits in trees: the path index and parsimony, *Annals of Combinatorics* **10** (2006), 77–96 [Special Volume on Biomathematics].

[16] A. W. M. Dress and M. A. Steel, Phylogenetic diversity over an abelian group, *Annals of Combinatorics* **11** (2007), 143–160 [Special Volume on Biomathematics].

[17] S. N. Evans and T. P. Speed, Invariants of some probability models used in phylogenetic inference, *Annals of Statistics* **21** (1993), 355–377.

[18] D. P. Faith, Conservation evaluation and phylogenetic diversity, *Biological Conservation* **61** (1992), 1–10.

[19] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002. Also available at `http://www.math.cornell.edu/~hatcher/AT/ATpage.html`.

[20] W. J. Heiser and M. Bennani, Triadic distance models: aximomatization and least squares representation, *Journal of Mathematical Psychology* **41** (1997), 189–206.

[21] S. Joly and G. Le Calvé, Three-way distances, *Journal of Classification* **12** (1995), 191–205.

[22] L. Pachter and D. Speyer, Reconstructing trees from subtree weights, *Applied Mathematics Letters* **17** (2004), no. 6, 615–621.

[23] C. Semple and M. A. Steel, *Phylogenetics*, Oxford University Press, 2003.

[24] M. A. Steel, Phylogenetic diversity and the greedy algorithm, *Systematic Biology* **54** (2005), 527–529.

Received May 25, 2009; revised March 7, 2010.

Author information

Andreas W. M. Dress, CAS-MPG Partner Institute and Key Laboratory for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, P. R. China; and Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany.
E-mail: dress@sibs.ac.cn

Yaokun Wu, Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.
E-mail: ykwu@sjtu.edu.cn