

Structural Parsimony: Reductions in Sequence Space

Roberto Blanco

Department of Computer Science and Systems Engineering (DIIS), Aragon Institute of Engineering Research (I3A)
Universidad de Zaragoza
Zaragoza, Spain
robertob@unizar.es

Abstract—Computational phylogenetics has historically neglected strict theoretical approaches that exploit the mathematical models beneath which it abstracts away the nuances of evolution. In particular, parsimony is conceptually simple and amenable to rigorous treatment, and has a clear analogue in graph theory, the Steiner tree. We present and refine the notion of sequence space as the soil from which all graph-theoretical methods arise, studying its structural properties and complexity with an eye on maximum parsimony. We therefrom introduce a basic set of very efficient implicit reductions that discard information with a fixed effect on the optimality of the solution, and show how it can be applied to large, real datasets.

Keywords—computational phylogenetics; maximum parsimony; reduction rule; sequence space; Steiner tree

I. INTRODUCTION

The objective of phylogenetics is the study of inheritance relationships between living organisms. Many mathematical models can be devised in an attempt to mimic the intricate forces of evolution, the simplest of which applies the *principle of parsimony* to phylogenetic reconstruction. It intuitively seeks to interpret evolution through Ockham’s razor: a preference for solutions that entail the least possible overall change (i.e., mutations).

The attractiveness of parsimony lies in its simplicity: making few assumptions, it is rid of sources of error that may disrupt more sophisticated models; in particular, it is mostly unburdened by the otherwise ever-present *model selection problem*, since, distinctively, the method is self-contained in its ability to score potential solutions. Even more, it is purely discrete in nature, and characterization of inner nodes, i.e., ancestral taxa, is integral to the method. However, algorithms scale poorly and quickly become unworkable as inputs grow.

Parsimony is intimately related to the *Steiner tree problem* [1], whose rich algorithmic corpus transcends conventional, more or less blind, search methods and their costly statistical sampling. The connection was probably first noted by [2], albeit in a restricted form, and discussed at length in [3]. Both expanded the state space to solve the derived Steiner problem explicitly; this was recently translated to Euclidean space, where it can be computed implicitly [4].

The analogy opens up further avenues of research, among which *input reduction* may be one of the most exciting. In this paper, we concern ourselves with the study of

dependence between input components to achieve optimal solutions by solving smaller, though equivalent, problem instances. By treating inputs implicitly, we will avoid the combinatorial explosion of the space in which they are embedded, and preserve the universality of our proposal.

II. THEORETICAL FOUNDATIONS

A. Problem Definitions

The Steiner tree problem, STP, over a certain metric space (or graph) and a given set \mathcal{T} of special terminal points (vertices) consists in the computation of a connecting tree \mathcal{S} of minimum combined length. This is equivalent to the *maximum parsimony problem* in phylogenetics [5] when the metric space from which the distance function $d(\cdot, \cdot)$ derives is a Hamming space [6]:

$$\min_{\mathcal{S}=(\mathcal{V} \supseteq \mathcal{T}, \mathcal{E})} \sum_{(x,y) \in \mathcal{E}} d(x,y) : d(t_i, t_j) \neq \infty, \forall t_i, t_j \in \mathcal{T} \quad (1)$$

Several remarks are in order. Firstly, this definition is restricted to the case of *unordered characters*, referred to as *Fitch parsimony* [7]. Secondly, Steiner trees need neither be binary nor confine terminals to the leaves, as is customary in phylogenetic trees; however, empty branches (i.e., connecting duplicates of the same point) can be added to account for these differences. And thirdly, the departure from binary alphabets grants the space some remarkable topological twists, to be explored under the next heading.

Input is in the form of a *multiple sequence alignment* matrix $A \in \mathcal{M}_{s \times l}$, where s (rows) is the number of sequences and l (columns) the number of inferred characters. Although columns with gaps are commonly suppressed, it is not especially unreasonable to preserve them as long as they represent legitimate absent features. Our inputs will be constrained to fully resolved sequences; unresolved descriptions would represent sets of acceptable sequences wherefrom one such that allows a combinatorially optimal solution must be selected.

The following typographic conventions are observed: minuscule script for scalar values and points, capital for matrices, calligraphic for sets, Greek for alphabets (lowercase for symbols, with σ reserved for unknowns), blackletter for tuples and structures (e.g., graphs), and blackboard for spaces. Proofs will be omitted due to space constraints.

B. Time and Space Complexity

Our discourse will henceforth concentrate on Fitch parsimony, the most general and widespread paradigm. We will examine the *sequence space* composed of all potentially feasible Steiner vertices, first established by [8] in the context of natural selection and fitness landscapes, and more formally in [9]. Before we undertake its study in a new light, some underlying concepts are required.

Definition. Let $\mathcal{W} = \{w_1, \dots, w_n\} : w_i \in \Sigma^s$ be an ordered set of words. The i -th component subalphabet is defined as $\Sigma_{\mathcal{W}}^i = \{\sigma \in \Sigma : w_{i\sigma} > 0\}$, where $w_\sigma : \sigma \in \Sigma$ denotes the number of occurrences of σ in a word w .

Alignments can be conceived as ordered sets of columns. Thus, an alignment A induces an l -dimensional Hamming space \mathbb{H}_A , whose j -th dimension, \mathbb{H}_A^j , spans the symbols contained in Σ_A^j . Its structure is that of the *generalized hypercube* [10]. The defining quality of unordered parsimony is full connectivity within each dimension, so any sequence is adjoined to $\sum_{j=1}^l (|\Sigma_A^j| - 1)$ others.

Typical alphabets are $\Sigma_{DNA} = \{A, C, G, T\}$ and $\Sigma_{AA} = \{Ala, \dots, Val\}$, possibly augmented with the gap $\{-\}$; they can be inferred inductively as well: $\Sigma_A = \bigcup_{j=1}^l \Sigma_A^j$, being closely related to sequence space size as follows.

Definition. Given a sequence alignment A , the unconstrained size of its associated Hamming space equals $|\mathbb{H}_A| = |\Sigma_A|^l$. Its constrained size equals $\|\mathbb{H}_A\| = \prod_{j=1}^l |\Sigma_A^j|$.

Size is obviously exponential in l , though conservation of functional sites can mitigate this trend to a large extent, especially for sequences over large alphabets. Regarding computational complexity, this variation of the Steiner problem is known to be, like most, NP-complete [11]. Furthermore, even the reoptimization problem has recently been shown to be NP-hard [12]. Consequently, achievements in peeling down the alignment will produce substantial speedups.

It suffices to solve the Steiner problem in a “restricted” sequence space where only direct connections are valid. In consequence, a branch of non-unitary length is a composite of single point mutations. Three types of vertices can be distinguished: *leaves* ($\deg(v) = 1$) and inner nodes, *degenerate* ($\deg(v) = 2$) or *regular* ($\deg(v) > 2$). We will denote paths by their endpoints \overline{vw} or the node sequence $\overline{v_1 \dots v_n}$.

C. Reduction Rules

An input reduction rule is an operation that may allow the optimal resolution of a problem through a smaller instance of itself. Rules should be fast and efficient, so that: $t_{red}(I \rightarrow I', R) + t_{sol}(I' \rightarrow O') + t_{aug}(O', R \rightarrow O) \ll t_{sol}(I \rightarrow O)$ holds.

Relevant work in the field is generally limited to the *Steiner problem in graphs* (SPG) and variations thereof since its inception in [13]. The discipline is firmly founded upon [14], which remains its major reference to this day, but

is ineffective for our purposes: the regularity of sequence space (limited connectivity, universal edge length) precludes isolated or suboptimal paths, bottlenecks and swaps.

Here we shall present implicit rules that perform a direct attack on terminal points, reducing either their number or complexity, and show their effects in sequence space.

III. NEW DEVELOPMENTS

A. Reduction Rules in Columns

The parsimony model assumes columns are independent for all theoretical purposes. Though a significant departure from reality, this property is the cornerstone of the first and simplest group of reduction rules. Herein we reconsider the body of results developed by [15] under the prism of reductions in sequence space. The Fitch score of a tree under a column of a compatible alignment will be marked with $\$F(A_{*j}, \mathfrak{T})$. Obviously, $\$F(A, \mathfrak{T}) = \sum_{j=1}^l \$F(A_{*j}, \mathfrak{T})$.

1) *Deterministic Contributions:* A first form of exact reduction over columns removes those characters whose contribution to tree score is fixed regardless of tree topology. The order relation determined by the scoring function in the set of possible solutions (alignment-compatible leaf-labeled trees) is therefore not altered by the exclusion of said columns. It will be advantageous to define the general case in terms of the natural concept of dominance.

Definition. Let w be a word over an alphabet Σ . w is said to be k -invariant, $k \in [0, |w| - 1]$, iff $\exists \alpha \in \Sigma : w_\alpha = |w| - k \wedge \nexists \beta \in \Sigma : w_\beta > w_\alpha$.

This property establishes a dominant symbol α in a word and the number k of exceptions to its hegemony; note that if $w_\alpha \geq |w|/2$ (conservation of 50% or higher) there cannot exist more than one dominant. This concept can be applied to cladistic characters interpreted as words; coupled with alphabet restrictions, we can reach a full characterization of “parsimonious uninformaton”.

Theorem 1. The contribution of any k -invariant column A_{*j} such that $|\Sigma_A^j| = k + 1$ to the score of any tree \mathfrak{T} is fixed: $\$F(A_{*j}, \mathfrak{T}) = k$.

Corollary. If the conditions of Theorem 1 are met, the Steiner problem can be solved for the reduced alignment matrix $A' = [A_{*1} \dots A_{*(j-1)} A_{*(j+1)} \dots A_{*l}]$. The constrained size of its sequence space is such that: $\|\mathbb{H}_{A'}\| = \|\mathbb{H}_A\| / (k + 1)$.

This theorem defines parsimoniously informative characters by exclusion: no further topology-independent reductions can be effected. It can be seen that all 0-, 1- and $(s - 1)$ -invariant columns are, clearly, uninformative.

2) *Parsimony Equivalence Classes:* A second compaction arises from the observation that all columns must be scored separately for each considered tree, yet Fitch parsimony makes no distinction between symbols: only their

(in)equalities matter. This corresponds to the concept of isomorphism as sketched by [15], which, though certainly useful, has not received widespread adoption to date.

Formally speaking, given an ordered, canonical alphabet of sufficient size, $\Sigma_K = \{\alpha_1, \dots, \alpha_k\}$, any word can be translated into canonical form as follows: all occurrences of the first symbol that appears in the sequence are replaced by the first canonical symbol, and so on. When applied to sequence alignments, $|\Sigma_K| \geq |\Sigma_A|$ suffices. The next property emerges naturally.

Theorem 2. *Let A be an alignment matrix and $A_{*j}, A_{*j'} \in \Sigma_A^s$: $j \neq j'$ two of its columns. If the canonical translations of both columns are equal, $k_j = k_{j'}$, then $\mathcal{S}_F(A_{*j}, \mathcal{T}) = \mathcal{S}_F(A_{*j'}, \mathcal{T})$ for every tree \mathcal{T} .*

Corollary. *In accord with Theorem 2, the columns of an alignment A can be classified in $l' \leq l$ isomorphic sets: $\{\mathcal{K}_1, \dots, \mathcal{K}_{l'}\}$, each represented by a canonical word $k_i \in \Sigma_K^s$ of length s . The alignment can be replaced by a (possibly) reduced alignment $A' = [k_1 \cdots k_{l'}] \in \mathcal{M}_{s \times l'}$ and a class weight vector $W = [|\mathcal{K}_1| \cdots |\mathcal{K}_{l'}|]$, and trees be equivalently scored per the revised formula: $\mathcal{S}_F(A, W, \mathcal{T}) = \sum_{j=1}^{l'} W_j \mathcal{S}_F(A_{*j}, \mathcal{T})$. The sizes of both spaces are related by: $\|\mathbb{H}_{(A', W)}\| = \|\mathbb{H}_A\| / \prod_{j=1}^{l'} \max\{1, (W_j - 1) \Sigma_{A'}^j\}$.*

Column symmetries produce folds in sequence space, projecting sets of dimensions together so they become optimally dependent upon one another, and compacted “edges” have uniform lengths equal to the cardinality of their class dimension. Because dimensions remain independent otherwise, graph reductions continue to be ineffective.

B. Reduction Rules in Rows

We have hitherto considered the discard of columns under certain conditions of single-symbol dominance. Decreasing the number of terminals, as opposed to their length, requires a global search for optimal subtrees that can be treated as atomic partial solutions. Here we demonstrate some fundamental conditions for these, noting the trivial case of terminal equality before extending it to near-equality.

Theorem 3. *Any two identical sequences can be unified.*

Corollary. *A Steiner tree can be computed for the set of distinct terminal sequences, i.e., unique points in space. Once the result has been converted into a standard phylogenetic tree, terminal nodes acting in stead of sets of sequences can be expanded as polytomies of the elements of each set.*

1) *Outer Isolation:* The first group of reductions that operate directly on full terminal sequences shares a *condition of isolation* that presently permits the removal of a secluded terminal without any further alterations to the input set. Firstly, we introduce an auxiliary property which will be of assistance in several results by identifying the potential leaf nature of certain terminals based on local features alone.

Lemma 1. *Let $w \in \Sigma^s$ be a word and $\omega \in \Sigma$ be a symbol such that $w_\omega = 1$: we call this a singleton symbol. If w represents the j -th column of a sequence alignment A_{*j} , then there exist Steiner trees over \mathbb{H}_A that feature the containing sequence of the singleton symbol as a leaf.*

In essence, the lemma claims that the possibility always exists of a *single generation point* for ω in the j -th dimension. From this, we will conclude that terminals thus differentiated are themselves isolated and expendable.

Theorem 4. *Let $a \in \mathcal{T}$ be a terminal. If there exists $b \in \mathcal{T}$ such that $d_{\mathbb{H}}(a, b) = 1 \wedge d_{\mathbb{H}}(a, t) \geq d_{\mathbb{H}}(b, t), \forall t \in \mathcal{T} - \{a, b\}$, then there exists a Steiner tree $\mathcal{S} = (\mathcal{V}, \mathcal{E})$ such that $(a, b) \in \mathcal{E} \wedge \nexists (a, v) \in \mathcal{E} : v \neq b$.*

Corollary. *If the conditions of Theorem 4 are met, we may solve for $\mathcal{T}' = \mathcal{T} - \{a\}$, yielding $\mathcal{S}' = (\mathcal{V}', \mathcal{E}')$, and augment the solution as follows: $\mathcal{V}^+ = \mathcal{V}' \cup \{a\}$, $\mathcal{E}^+ = \mathcal{E}' \cup \{(a, b)\}$. The constrained size of the modified sequence space is expressed by: $\|\mathbb{H}_{A'}\| = \|\mathbb{H}_A\| (|\Sigma_{A'}^1| - 1) / |\Sigma_{A'}^1|$.*

For economy of notation, here and in what follows A', A'' denote the canonical alignments of $\mathcal{T}, \mathcal{T}'$, where singleton columns of interest occupy the first positions in the matrix.

More distant isolation between terminals becomes not quite so straightforward to treat and is beyond the extent of this paper. We will instead aim our attention at a more stringent form of isolation revolving around singleton symbols that generalizes the column structure of the last theorem.

Theorem 5. *Let $a \in \mathcal{T}$ be a terminal with k singleton symbols. If there exists $b \in \mathcal{T}$ such that $d_{\mathbb{H}}(a, b) = k$, then there exists a Steiner tree $\mathcal{S} = (\mathcal{V}, \mathcal{E})$ such that $\{(a, v_1), \dots, (v_i, v_{i+1}), \dots, (v_{k-1}, b)\} \subseteq \mathcal{E} \wedge \deg(a) = 1 \wedge \deg(v_i) = 2, \forall i$.*

Corollary. *If the conditions of Theorem 5 are met, we may solve for $\mathcal{T}' = \mathcal{T} - \{a\}$, yielding $\mathcal{S}' = (\mathcal{V}', \mathcal{E}')$, and augment the solution as follows: $\mathcal{V}^+ = \mathcal{V}' \cup \{a, v_1, \dots, v_{k-1}\}$, $\mathcal{E}^+ = \mathcal{E}' \cup \{(a, v_1), \dots, (v_i, v_{i+1}), \dots, (v_{k-1}, b)\}$, for any fitting ordering of the intermediate mutations. The size of the reduced space is: $\|\mathbb{H}_{A'}\| = \|\mathbb{H}_A\| \prod_{j=1}^k ((|\Sigma_{A'}^j| - 1) / |\Sigma_{A'}^j|)$.*

2) *Inner Fusion:* Throughout the last heading, we have shown how estranged terminals can sometimes be ignored. There are cases where a similar optimal substructure is visible though it cannot be isolated. Regardless, if uninterrupted optimal paths may exist, they can be simplified likewise.

To pave the way for the basic result of terminal fusion, we need to introduce the view of sequences as subspaces, and a compact representation thereof.

Definition. *A hypersequence of length l over an alphabet Σ is a sequence of the same length over the power set of Σ , minus the empty set: $w \in (\mathcal{P}(\Sigma) - \emptyset)^l$. If necessary, a sequence may have its single point (if applicable) or subspace nature stressed by marking it as \dot{w} or \ddot{w} , respectively.*

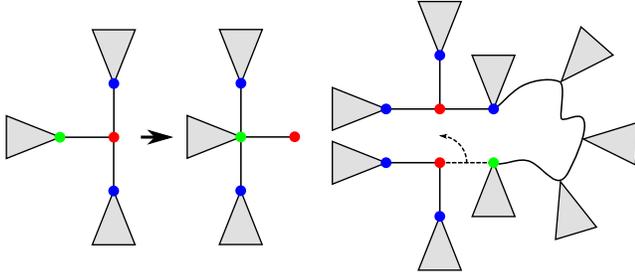


Figure 1. Graphic proof of Lemma 1 (left) and Theorem 6 (right).

A hypersequence promotes non-empty sets of alphabet symbols as components of a new alphabet, and extends the string concatenation operator as the Cartesian product of sets. The dual nature of single points as trivial subspaces is expressed by: $\dot{w} = \sigma_1 \cdots \sigma_l \Leftrightarrow \dot{w} = \{\sigma_1\} \times \cdots \times \{\sigma_l\}$.

The rationale for fusion reductions lies in the amalgamation of terminals that accrues from our final theorem. It can be conceived as a relaxation of Theorem 4 where no restrictions are imposed on external terminals.

Theorem 6. *Let $a, b \in \mathcal{T}$ be terminals. If $d_{\mathbb{H}}(a, b) = 1$, then there exists a Steiner tree $\mathfrak{S} = (\mathcal{V}, \mathcal{E})$ such that $(a, b) \in \mathcal{E}$.*

Corollary. *If the conditions of Theorem 6 are met, we may solve for $\mathcal{T}' = (\mathcal{T} - \{a, b\}) \cup \{\tilde{c}\} : \tilde{c} = \{a, b\}$, yielding $\mathfrak{S}' = (\mathcal{V}', \mathcal{E}')$; no augmentations are needed. The new space is related to its predecessor as: $\|\mathbb{H}_{A'}\| = \|\mathbb{H}_A\| - 1$.*

Although the corollary is universal for arbitrary sets of connected terminals, the independent nature of each dimension mandates that hypersequences stand for full subspaces unless additional information is provided, which surpasses the scope of this paper: suffice it to say for now that single-dimension hypersequences can be treated as they stand.

Most importantly, connections to a “terminal subtree” can occur equivalently at any point; Hamming distance must be extended as: $d_{\mathbb{H}}(\tilde{x}, \tilde{y}) = \min \{d_{\mathbb{H}}(\hat{x}, \hat{y}) : \hat{x} \in \tilde{x} \wedge \hat{y} \in \tilde{y}\}$. The overlap between hypersequences and ambiguous sequences now becomes apparent; it is tempting to allow the latter and treat them as single “points”, though the interpretation of the resulting trees may suffer subtle alterations.

It remains to consider how former reduction rules are affected by the introduction of hypersequences. Fusion operates as expected insofar as all starting terminals stand for singular points in space, whereas isolation requires extension to whole related regions of space. Having augmented matrix elements to sets of symbols, it is the compatibility of these with a reducible alignment pattern that becomes important.

IV. PRACTICAL MATTERS

A. Implementation

The practicalities of SPG reduction are best illustrated by [16], where the problem of arriving at the most reduced

dataset possible is discussed in terms of the potential benefits of preprocessing strategies balanced by their computational load. General tests, too costly for practical use, are restricted or worked into on-demand computation schemes.

Implicit reductions are important because, unlike graph reductions, they are purely polynomial in time and space, so often they will make the difference between tractability and intractability. Moreover, any algorithm, exact or heuristic — including traditional parsimony search —, will greatly benefit from simpler datasets. Recurring items can be updated through dynamic recomputation to further decrease costs.

Let h be the maximum number of primary symbols per hypersymbol. Column tests involve informative checks and parsimony classes, both taking $O(sh)$ plus access times; the (cost-multiplicative) number of candidates is $O(l)$ and $O(l^2)$. Row tests always have $O(s^2)$ potential candidates; calculating the number of singletons in a sequence takes $O(slh)$. Individually, Hamming distances can be computed in $O(lh)$; row and column reductions depend on the data structure: bidimensionally-linked lists allow deletions in $O(l)$ and $O(s)$, and simple matrices take up to $O(sl)$.

Row and column reductions interplay by diminishing diversity in the complementary dimension and triggering simplifying chain-reactions. Row fusion as it stands is unlikely to isolate previously unclustered terminals or straightforwardly degrade columns, while being sensitive to order of candidate selection. All this leads to a simple and efficient preprocessing algorithm, where column rules are applied first due to their locality and speed:

repeat

apply Theorem 1; apply Theorem 2; apply Theorem 5

until dataset is irreducible

apply Theorem 6

B. Case Studies

Three genes from the mitochondrial genome encoding a variety of structures —a tRNA (MT-TL1), an rRNA (MT-RNR1), and a protein (MT-CO1) in DNA-base and amino acid forms— have been extracted from complete human genomes and reference mammalian genomes in GenBank, with incomplete sequences removed prior to alignment and application of the proposed reduction procedure.

Whereas more diverse datasets undergo less radical transformations, densely sampled taxa exhibit a higher proportion of closely related sequences and a more compact proper size s_1 ; generally, the genetic code makes amino acid translations more amenable to simplification. On the other hand, while the informative column bound proves quite resistant to improvement, the number of distinct rows (the main source of exponentiality in Steiner algorithms) decreases substantially, so the effect of the improvements should not be downplayed.

Overall, sequence spaces are compressed by many orders of magnitude; the impact of reduction is roughly proportional to the raw dimensions of each problem. Executions

Table I

EXPERIMENT RESULTS (s_0, l_0 : BASE DIMENSIONS; s_1, l_1 : STANDARD REDUCTIONS; s_2, l_2 : FINAL DIMENSIONS; r : REDUCTION FACTOR).

Dataset	s_0	l_0	s_1	l_1	s_2	l_2	r
Hum. tRNA	6785	77	18	6	5	5	7.68e2
Hum. rRNA	6762	969	416	119	215	116	1.29e240
Hum. gene	6778	1554	986	314	578	294	2.03e359
Hum. prot.	6776	517	157	28	44	26	1.00e27
Mam. tRNA	304	95	238	66	202	56	1.81e9
Mam. rRNA	301	1263	299	870	298	851	5.71e95
Mam. gene	303	1565	301	779	298	773	1.31e40
Mam. prot.	303	518	268	160	238	159	2.20e30

last mere seconds and so can be considered negligible, even if no subsequent profit were to be gained from their application.

V. CONCLUDING REMARKS

In the foregoing we have shed new light on some of the lesser known faces of parsimony, offering as a novelty the implicit reduction of alignments over a rigorously defined sequence space. Our rules are extremely efficient and very effective in practice: biological data are far from unstructured and there are many opportunities to learn and employ their properties to our advantage. Whenever there is an underlying structure, it is possible to exploit it through implicit reduction, irrespective of subsequent processing.

Certainly, deterministic optimality may impose certain structures that restrict the set of reachable solutions to a subset of the original, though it has a clear parsimonious interpretation. In any case, the alternative (untreated inputs) may be too costly for exact algorithms and too complicated for heuristic techniques. Nevertheless, there is still much room for improvement, most clearly regarding extensions of these essential properties to learn structural constraints of divergent data; the versatility of terminal fusion is greater than might be surmised from its basic form, as well.

Beyond exact operations, structural features may also bring *approximate reductions* with well-founded degradation bounds, which could be used to engage even bigger inputs with certain guarantees; the STP in Euclidean space is the closest implicit analogue otherwise. Related subjects of ongoing study include alternative parsimony models and provisions to incorporate ambiguous or heteroplasmic sequences cleanly. On the algorithmic front, rule selection and prioritization according to expectation of success, as well as cost prediction and optimization, are all matters of concern.

Let us conclude by heeding Box's dictum: all models are wrong and parsimony is no exception, but useful it is indeed, and much can be learned from it.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Science and Innovation (MICINN) [TIN2008-06582-C03-02]; and the Spanish Ministry of Education [AP2008-03447].

REFERENCES

- [1] F. K. Hwang, D. S. Richards, and P. Winter, *The Steiner tree problem*. Amsterdam: North-Holland, 1992.
- [2] L. Nastansky, S. M. Selkow, and N. F. Stewart, "Cost-minimal trees in directed acyclic graphs," *Math. Method. Oper. Res.*, vol. 18, pp. 59–67, Feb. 1974.
- [3] L. R. Foulds, M. D. Hendy, and D. Penny, "A graph theoretic approach to the development of minimal phylogenetic trees," *J. Mol. Evol.*, vol. 13, pp. 127–149, Jun. 1979.
- [4] M. Brazil, B. K. Nielsen, D. A. Thomas, P. Winter, C. Wulff-Nilsen, and M. Zachariassen, "A novel approach to phylogenetic trees: d -dimensional geometric Steiner trees," *Networks*, vol. 53, pp. 104–111, Mar. 2009.
- [5] J. H. Camin and R. R. Sokal, "A method for deducing branching sequences in phylogeny," *Evolution*, vol. 19, pp. 311–326, Sep. 1965.
- [6] R. W. Hamming, "Error detecting and error correcting codes," *AT&T Tech. J.*, vol. 29, pp. 147–160, Apr. 1950.
- [7] W. M. Fitch, "Toward defining the course of evolution: minimum change for a specific tree topology," *Syst. Zool.*, vol. 20, pp. 406–416, Dec. 1971.
- [8] J. Maynard Smith, "Natural selection and the concept of a protein space," *Nature*, vol. 225, pp. 563–564, Feb. 1970.
- [9] M. Eigen, R. Winkler-Oswatitsch, and A. Dress, "Statistical geometry in sequence space: a method of quantitative comparative sequence analysis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 85, pp. 5913–5917, Aug. 1988.
- [10] L. M. Bhuyan and D. P. Agrawal, "Generalized hypercube and hyperbus structures for a computer network," *IEEE Trans. Comput.*, vol. C-33, pp. 323–333, Apr. 1984.
- [11] L. R. Foulds and R. L. Graham, "The Steiner problem in phylogeny is NP-complete," *Adv. Appl. Math.*, vol. 3, pp. 43–49, Mar. 1982.
- [12] H.-J. Böckenhauer, J. Hromkovič, R. Krlovič, T. Mömke, and P. Rossmanith, "Reoptimization of Steiner trees: changing the terminal set," *Theor. Comput. Sci.*, vol. 410, pp. 3428–3435, Aug. 2009.
- [13] J. E. Beasley, "An algorithm for the Steiner problem in graphs," *Networks*, vol. 14, pp. 147–159, Spring 1984.
- [14] C. W. Duin and A. Volgenant, "Reduction tests for the Steiner problem in graphs," *Networks*, vol. 19, pp. 549–567, Aug. 1989.
- [15] W. M. Fitch, "On the problem of discovering the most parsimonious tree," *Am. Nat.*, vol. 111, pp. 223–257, Mar./Apr. 1977.
- [16] T. Koch and A. Martin, "Solving Steiner tree problems in graphs to optimality," *Networks*, vol. 32, pp. 207–232, Oct. 1998.