

KAIKOURA 2014 CHALLENGES

SOME QUESTIONS: POSTED FEB 1, 2014 (MIKE STEEL)

- (1) Is the problem of determining whether a collection of binary characters on X has a ‘persistent’ phylogeny polynomial-time or not? This problem was left open in [1].
- (1a) What is the smallest number k_n for which each binary phylogenetic tree on X is the only ‘persistent’ phylogeny for k_n carefully chosen binary characters?
- (2) A collection $\mathcal{C}$ of subsets of a finite set X is a *bureaucracy* if (i) $\mathcal{C} \neq \emptyset$ and $\emptyset \notin \mathcal{C}$, and (ii) every hierarchy $H \subseteq \mathcal{C}$ can be extended to a maximum hierarchy H' such that $H \subseteq H' \subseteq \mathcal{C}$. Is there an algorithm for determining whether or not an arbitrary $\mathcal{C}$ is a bureaucracy which is polynomial in $|\mathcal{C}|$ and $|X|$. This problem was left open in [2].
- (3) Suppose $\mathcal{Q}$ is a set of quartet trees on leaves drawn from a set X , and the semi-dyadic closure of $\mathcal{Q}$ is the set of all induced quartet trees of some binary phylogenetic X -tree. Then does $\mathcal{Q}$ contain a subset $\mathcal{Q}'$ of size $n-3$ that defines T ? This question was left open from [3] (Remark 4), where the converse result was established in Theorem 3.
- (4) [‘Something old, something new’] A collection $\mathcal{C}$ of subsets of a finite set X has an *evolutionary ordering* if the sets in $\mathcal{C}$ can be ordered $S_1, S_2, \dots$ so that (i) each set contains at least one element of X which occurs in an earlier set in the sequence, and (ii) each set contains at least one element that is absent from any earlier set in the sequence. This notion was studied in [4]. Is the problem of determining whether $\mathcal{C}$ has an evolutionary ordering NP-complete?
- (5) The above questions are purely combinatorial. This one is probabilistic. Consider a simple model of discrete-state evolution (e.g. the symmetric 2-state model) on an unrooted binary phylogenetic X -tree T . Let b be the (vector of) branch lengths of T , and let $s(T, b)$ be the expected frequency of site patterns on X generated by this model. For a given vector v of site pattern frequencies, let $L(v, b)$ be the log-likelihood function defined by $-\sum_i v_i \log s(T, b)_i$ where the sum is over all site patterns. Now suppose that $v_0 = s(T, b_0)$ for some fixed positive branch length assignment b_0 . It is well known (and easily shown) that the function $b \mapsto L(v_0, b)$ has a unique (and interior) global minimum (i.e. a global ‘maximum likelihood solution’), namely $b = b_0$. Question: Is this the only local minimum of this function (for every binary T and choice of b_0 strictly positive).

REFERENCES

- [1] Bonizzoni, P, Braghin, C., Dondic, R. and Trucco, G. (2012). The binary perfect phylogeny with persistent characters, *Theoretical Computer Science* 454 51–63.
- [2] Bryant, D. and Steel, M. (2012). ‘Bureaucratic’ set systems, and their role in phylogenetics. *Applied Math. Letters*. 25: 1148–1152.
- [3] Bryant, D., Böcker, S., Dress, A.W.M., and Steel, M. (2000). Algorithmic aspects of tree amalgamation. *Journal of Algorithms* 37: 522–537.
- [4] Little, C.H.C., Campbell, A.E (2000).: Evolutionary Families of Sets. *Electr. J. Comb.* 7.