RECENT PROGRESS IN PHYLOGENETIC COMBINATORICS

ANDREAS DRESS

CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, P.R. China and Max-Planck-Institut fuer Mathematik in den Naturwissenschaften, Inselstrasse 22-26, D-04103 Leipzig, Germany

1. Background

Phylogenetic combinatorics deals with the combinatorial aspects of phylogenetic tree reconstruction. A starting point was the following observation:

Given a metric $D : X \times X \to \mathbb{R}$ representing the approximative genetic distances between the members of a collection X of taxa, it was shown in Ref. 1 that the following assertions relating to the "object of desire", a phylogenetic-tree X-tree, all are equivalent:

(i) The space "tight span"

$$T_D := \{f \in \mathbb{R}^X : \forall_{x \in X} f(x) = \sup_{y \in X}(D(x,y) - f(y))\}$$

of D is an $\mathbb{R}$-*tree.*

(ii) There exists a tree (V, E) whose vertex set V contains X, and an edge weighting $\ell : E \to \mathbb{R}$ that assigns a positive length $\ell(e)$ to each edge e in E, such that D is the restriction of X to the shortest-path metric induced on V.

(iii) There exists a map $w : \mathcal{S}(X) \to \mathbb{R}_{\geq 0}$ from the set $\mathcal{S}(X)$ of all bi-partitions or *splits* of X into the set $\mathbb{R}_{\geq 0}$ of non-negative real numbers such that, given any two splits $S = \{A, B\}$ and $S' = \{A', B'\}$ in $\mathcal{S}(X)$ with $w(S), w(S') \neq 0$, at least one of the four intersections $A \cap A', B \cap A'$, $A \cap B'$, and $B \cap B'$ is empty and $D(x, y) = \sum_{S \in \mathcal{S}(X:x \leftrightarrow y)} w(S)$ holds where $\mathcal{S}(X : x \leftrightarrow y)$ denotes the set of splits $S = \{A, B\} \in \mathcal{S}(X)$ that *separate* x and y.

(iv) $D(x, y) + D(u, v) \leq \max\big(D(x, u) + D(y, v), D(x, v) + D(y, u)\big)$ holds for all $x, y, u, v \in X$

Moreover, the metric space T_D actually coincides in this case with the $\mathbb{R}$-tree that is canonically associated with a weighted X-tree (V, E, ℓ).

2. Discussion

This observation suggested to further investigate (1) the tight-span construction and (2) representations of metrics by weighted split systems with more or less specific properties, even if the metric in question would not satisfy the very special properties described above which investigations have, in turn, given rise to a full-fledged research program dealing with many diverse aspects of these two topics (see the list of references below).

In my lecture, I will focus on the rather new developments relating to *block decomposition* and *virtual cut points* of metric spaces reported, respectively, in References 2 and 3 that allow one to canonically decompose any given finite metric space into a sum of pairwise compatible *block metrics*, thus providing a far-reaching generalization of the result recalled above.

References

1. A. Dress, *AiM* **53**, 321 (1984).
2. A. Dress, K. Huber, J. Koolen and V. Moulton, Compatible decompositions and block realizations of finite metric spaces., submitted.
3. A. Dress, K. Huber, J. Koolen, V. Moulton and A. Spillner, A note on the metric cut point and the metric bridge partition problems., submitted.
4. J. Apresjan, *Mashinnyi perevod: prikladnaja lingvistika* **9**, 3 (1966).
5. E. Baake, *Math. Biosci.* **154**, 1 (1998).
6. H. J. Bandelt, *SIAM J. Disc. Math.* **3**, 1 (1990).
7. H. Bandelt and A. Dress, *Bul. Math. Biol.* **51**, 133 (1989).
8. H. Bandelt and A. Dress, *AiM* **92**, 47 (1992).
9. H. Bandelt and A. Dress, *Molecular Phylogenetics and Evolution* **1**, 242 (1992b).
10. H. Bandelt, V. Chepoi, A. Dress and J. Koolen, *Eur. J. Comb.* **27**, 669 (2006).
11. H. Bandelt and M. Steel, *SIAM. J. Disc. Math.* **8**, 517 (1995).
12. J.-P. Barthélemy and A. Guénoche, *Trees and proximity representations* (Wiley, 1991).
13. S. Böcker and A. Dress, *AiAM* **138**, 105 (1998).
14. B. Bowditch, *Notes on Gromov's hyperbolicity criterion for path metric spaces.*, in *Group theory from a geometric viewpoint*, eds. E. Ghys, A. Haefliger and A. Verjovsky (World Scientific, 1991), pp. 64–167.
15. P. Buneman, *The recovery of trees from measures of dissimilarity.*, in *Mathematics in the Archeological and Historical Sciences*, ed. F. H. et al. (Edinburgh University Press, 1971), pp. 387–395.
16. D. Bryant and V. Berry, *AiAM* **27**, 705 (2001).
17. D. Bryant and A. Dress, Linearly independent split systems., Europ. J. Comb., to appear.
18. A. Dress, *AiM* **74**, 163 (1989).
19. A. Dress, *Mathematical hierarchies and biology* , 271 (1996), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 37, Amer. Math. Soc., Providence, RI, 1997.
20. A. Dress, *AML* **15**, 995 (2002).
21. A. Dress, *Graphs and Metrics.*, in *Encyclopaedia of Genetics, Genomics, Proteomics and Bioinformatics*, ed. A. K. e. a. Shankar Subramaniam (John Wiley and Sons, 2005).
22. A. Dress, *The Category of X-nets.*, in *Networks: From Biology to Theory*, eds. J. Feng, J. Jost and M. Qian (Springer, 2006).

23. A. Dress, Phylogenetic analysis, split systems, and boolean functions., In: *Aspects of Mathematical Modelling*, Roger Hosking and Wolfgang Sprössig eds, Birkhäuser, 2006, to appear.
24. A. Dress, A note on group-valued split and set systems., *Contributions to Discrete Mathematics*, to appear.
25. A. Dress, Split decomposition over an abelian group, part i: Generalities., *AoC*, to appear.
26. A. Dress, Split decomposition over an abelian group, part ii: Group-valued split systems with weakly compatible support., *Discrete Applied Mathematics, Special Issue on Networks in Computational Biology*, to appear.
27. A. Dress, Split decomposition over an abelian group, part iii: Group-valued split systems with compatible support., submitted.
28. A. Dress, B. Holland, K. Huber, J. Koolen, V. Moulton and J. Weyer-Menkhoff, *Discrete Applied Mathematics* **146**, 51 (2005).
29. A. Dress, K. Huber, A. Lesser and V. Moulton, *AoC, Special Volume on Biomathematics* **10**, 63 (2006).
30. A. Dress, K. Huber, J. Koolen and V. Moulton, An algorithm for computing virtual cut points in finite metric spaces (2007).
31. A. Dress, K. Huber, J. Koolen and V. Moulton, Cut points in metric spaces, *AML*, in press.
32. A. Dress, K. Huber and V. Moulton, *AoC* **2**, 299 (1998).
33. A. Dress, K. Huber and V. Moulton, *AoC* **1**, 339 (1997).
34. A. Dress, K. Huber and V. Moulton, *AoC* **4**, 1 (2000).
35. A. Dress, K. Huber and V. Moulton, *AiM* **168**, 1 (2002).
36. A. Dress, K. Huber and V. Moulton, Some uses of the farris transform in mathematics and phylogenetics — a review., AoC, Special Volume on Biomathematics, to appear.
37. A. Dress, J. Koolen and V. Moulton, *AoC* **8**, 463 (2004).
38. A. Dress, V. Moulton and W. Terhalle, *Europ. J. Comb.* **17**, 161 (1996).
39. A. Dress and M. Steel, *AoC, Special Volume on Biomathematics* **10**, 77 (2006).
40. A. Dress and M. Steel, *AoC, Special Volume on Biomathematics* **10**, 77 (2006).
41. A. Dress and M. Steel, Phylogenetic diversity over an abelian group., AoC, Special Volume on Biomathematics, to appear.
42. J. Farris, *On the phylogenetic approach to vertebrate classification.*, in *Major patterns in vertebrate evolution*, eds. M. Hecht, P. Goody and B. Hecht (Plenum Press, 1976).
43. J. Farris, *Sys. Zool.* **28**, 200 (1979).
44. J. Farris, *Sys. Zool.* **28**, 483 (1979).
45. J. Farris, A. Kluge and M. Eckardt, *Sys. Zool.* **19**, 172 (1970).
46. M. Gromov, *Hyperbolic Groups,*, in *Essays in Group Theory, MSRI series, Vol. 8*, ed. S. Gersten (Spring-Verlag, 1988).
47. S. Grünewald, K. Forslund, A. Dress and V. Moulton, *Mol. Biol. Evol.* **24**, 532 (2007).
48. D. Huson, *Bioinformatics* **14**, 68 (1998), http://www-ab.informatik.uni-tuebingen.de/software/splits/welcome_en.html.
49. D. Huson and A. Dress, *ACM Transactions in Computational Biology and Bioinformatics* **1**, 109 (2004).
50. J. Isbell, *Comment. Math. Helv.* **39**, 65 (1964).
51. N. Jardine, *Biometrics* **25**, 609 (1969).
52. P. Lockhart, A. Meyer and D. Penny, *J. Mol. Evol.* **41**, 666 (1995).
53. A. Parker-Rhodes and R. Needham, *Information processing, Proceedings of the International Conference on Information Processing, Paris, 1960* , 321 (1960).
54. C. Semple and M. Steel, *AiAM* **23**, 300 (1999).

55. C. Semple and M. Steel, *Phylogenetics* (Oxford University Press, 2003).
56. M. Steel, *AML* **7**, 19 (1994).