

MATHEMATIKKOLLOQUIUM

Das Institut für Mathematik lädt zu folgendem Vortrag ein:

Hans-Jürgen Bandelt

Universität Hamburg

Was sind und sollen phylogenetische Netzwerke?

In der Biologie versteht man unter einem Netzwerk meist einen ungerichteten Graphen, der im Gegensatz zu einem Baum, Kreise enthalten kann und dessen Ecken und Kanten mit Taxa und Mutationen teilweise markiert sind. Hybridisierung und Rekombination erfordern letztendlich eine vernetzte Darstellung der evolutionären Ereignisse. In der Datenanalyse molekularer Daten würden aber auch, wenn das Endergebnis einen Baum wiedergeben soll, in Zwischenschritten Netzwerke zum Ansatz kommen, um konkurrierende Signale aus den Daten darzustellen und einer geeigneten Interpretierung zuzuführen. Falls schließlich optimale Baumlösungen nicht eindeutig sind und Majoritätskonsensus als zu uninformativ eingeschätzt wird, gibt es auch die Möglichkeit auf Konsensusoptionen zuzugreifen, die Bäume teilweise überlagern und damit Netzwerke produzieren.

Die derzeitige Situation in der Phylogenetik und Phylogeographie ist aber, was Netzwerke betrifft, die eines völligen Wildwuchses, da oftmals nur auf beliebige ad hoc Verfahren zurückgegriffen wird, deren Beschreibung bisweilen sogar völlig im Dunkeln bleiben. Eine mathematische Grundlegung der sinnvollen Netzwerkstrukturen hat in der Biologie noch nicht stattgefunden. Da aber in vielen Ansätzen letztendlich entweder ein Sparsamkeitskriterium im Sequenzraum oder analog kürzeste Länge bei Distanzbenutzung angestrebt wird, reicht es aus, grundsätzlich nur Netzwerke zuzulassen, die kürzeste Baumlösungen garantieren.

Mit ein wenig mathematischer Theorie braucht man dann als potentielle Netzwerkstrukturen von vornherein nur entweder (1) quasi-mediane Netzwerke (als Retrakte von Hamminggraphen, die direkt den Sequenzraum kodifizieren), oder (2) injektive metrische Räume (bei uneingeschränkter Distanzverwendung), oder (3) Helly-Graphen (wenn nur ganzzahlige Distanzen verwendet werden sollen) zu betrachten. In jedem Falle werden dabei innere Knoten, die nicht durch beobachtete Taxa markiert sind, durch spezielle Sequenzen von (1) Nukleotiden, bzw. (2) reellen Zahlen, bzw. (3) ganzen Zahlen koordinatisiert. Ob eine Koordinatensequenz für einen gegebenen Datensatz zulässig ist, kann dann jeweils in polynomialer Zeit entschieden werden. Dieses erlaubt, nur solche Netzwerkverfahren zu betrachten, die schrittweise immer nur zulässige Koordinatensequenzen erzeugen. Somit gilt auch hier das wohlbekannte Zitat, daß eine gute Theorie die beste Anwendung ist.

Zeit: Dienstag, den 3. Oktober 2006 um 17¹⁵ Uhr

Ort: Viktor-Franz-Hess-Haus, Technikerstrasse 25, HS D

Arne Dür

Gäste sind herzlich willkommen!