

15.06.2001 - Phylogenetische Bäume

Phylogenetische Bäume versuchen genetische Verwandtschaftsbeziehungen zwischen Lebensformen darzustellen, indem sie die genetischen Distanzen zwischen Ihnen, berechnet aus paarweisen Alignments in Pfaden abbilden.

Es wird angenommen, daß alle Spezies von Ahnen abstammen und sich durch Mutationen aus diesen entwickelt haben. Ziel ist es einen Baum zu konstruieren, der heutige Spezies an den Blättern trägt und hypothetische Ahnen den inneren Knoten zuweist. Dieser Baum soll die durch paarweise Alignments bestimmten genetischen Distanzen durch die Entfernung zwischen zwei Blättern repräsentieren, man sagt, er sei kantengewichtet. Man muß auch von jedem beliebigen Blatt zum anderen wandern, auch über Knoten, und dabei genau die entsprechende Distanz zurücklegen können, man sagt, die Distanzen seien additiv.

A B C D E Das Ergebnis der Alignments ist eine
A 0 2 7 4 7 sogenannte Distanzmatrix, die die
B 0 7 4 7 Entfernung aller zu behandelnden
C 0 7 6 Genomsequenzen zueinander enthält.
D 0 7 Wir spielten gestern mit einer Bei-
E 0 spielmatrix der Genome A bis E, die

Ihr links abgebildet seht. Die angegebenen Distanzen sind additiv, bilden eine sogenannte additive Metrik.

Zum Bau des Baumes beginnt man damit, aus einem Genom-Vergleichspaar den ersten Pfad zu bilden. Beginnen wir mit A und B, so dieser Pfad zwei Einheiten lang. Nun ist ein drittes Objekt C einzufügen, irgendwo zwischen A und B, an einer Stelle, nennen wir sie x. Dann gilt:

$d_{AB} = d_{Ax} + d_{Bx}$ \ was sich umformulieren lässt zu
 $d_{AC} = d_{Ax} + d_{Cx}$ |
 $d_{BC} = d_{Ax} + d_{Cx}$ / 3 Gleichungen,

die uns helfen, $d_{Ax} = (d_{AB} + d_{AC} - d_{BC}) / 2$
das neue Objekt $d_{Bx} = (d_{AB} + d_{BC} - d_{AC}) / 2$
einzufügen. $d_{Cx} = (d_{AC} + d_{BC} - d_{AB}) / 2$

Die Entfernungen zwischen A, B und C entnehmen wir aus der Distanzmatrix und setzen sie ein. So erhalten wir $d_{Ax} = (2 + 7 - 7) / 2 = 1$ und d_{Bx} muß auch gleich eins sein, sodaß der Knoten in der Mitte zwischen A und B einzufügen ist. Die von diesem Knoten neue abzweigende Kante nach C ist $d_{Cx} = (7 + 7 - 2) / 2 = 6$ Einheiten lang.

Um fortzufahren, wählen wir zufällig ein Blätterpaar aus dem bisher konstruierten Baum und berechnen den Punkt, wo das Objekt eingefügt werden soll. Kommt das Objekt auf keinem Knoten zu liegen, so fügen wir es an diese Stelle ein und machen mit dem nächsten Objekt weiter. Wenn man aber zu dem Schluß kommt, daß eine neue Kante in einem Punkt abzweigen soll, der schon ein Knoten ist, also schon eine abzweigenden Teilbaum besitzt, so müssen wir den Einfüangepunkt nochmals berechnen, für ein neues Blätterpaar, das aus einem der beiden vorigen Blätter und einem des Teilbaums besteht, da das neue

Objekt offensichtlich irgendwo als Blatt des Teilbaumes eingeordnet werden soll.

Erklärt man einen Punkt des additiven Baumes zur Wurzel, so nennt man ihn ultrametrischen Baum, der biologischen Stammbäumen entspricht.

Eine Spezielle Methode, genauere Phylogenetische Bäume zu erhalten, ist, die Bedeutung des zunächst berechneten Baumes in die Alignment- und Bewertungsverfahren rückfließen zu lassen und so iterativ zu besseren Ergebnissen zu gelangen.

Probleme bei der Bildung phylogenetischer Bäume:

Die Mutationsrate ist nicht unbedingt linear. - Im Genom konnte es zu plötzlichen großen Veränderungen kommen, die zu großen Distanzergebnissen führen, obwohl nahe Verwandtschaft besteht

Mutation ist nicht unbedingt gerichtet, Rückmutationen sind möglich. - Das führt zu geringer Distanz, obwohl mehr Ahnen dazwischenliegen.

Mutationen konvergieren, Fledermäuse wie auch Vögel haben Flügel - Geringere Distanz.

-- Samuel