

Zusammenfassung

Mutationsprozesse spielen eine fundamentale Rolle in den grundlegenden Mechanismen der molekularen Evolution. Sie erzeugen die notwendigen DNA-Veränderungen, aus denen genetische Variationen erst entstehen können. Im Vergleich zu Punktmutationen sind Insertionen und Deletionen von DNA-Segmenten weit weniger untersucht. Ziel dieser Arbeit war es, Charakteristika, Herkunft und Auswirkungen kurzer DNA Insertionen und Deletionen in der Genomevolution zu analysieren.

Wir untersuchten multiple Alignments von Mensch, Schimpanse und Rhesus, um genomweit kurze Insertionen und Deletionen zu identifizieren, die sich in der menschlichen Stammlinie seit ihrer Speziation vom letzten gemeinsamen Vorfahren mit dem Schimpansen ereignet haben. Wir konnten zeigen, dass die meisten Insertionen Tandemduplikationen direkt angrenzender Sequenzsegmente sind. Diese Eigenschaft hätte man für Insertionen in Mikro- und Minisatelliten durchaus erwartet. Wir stellten jedoch überraschenderweise fest, dass Insertionen auch in nicht-repetitiven Bereichen des Genoms vorwiegend als Tandemduplikationen identifiziert werden konnten. Nur ein kleiner Teil aller Insertionen und Deletionen hatte sich überhaupt in repetitiven Bereichen ereignet.

Die Auswirkungen dieser Beobachtungen sind vielschichtig. Wir konnten zum Beispiel zeigen, dass sich die Sequenzcharakteristika von Insertionen und Deletionen und ihrer direkten genomischen Umgebung deutlich von jenen Signaturen unterscheiden, die man für ungleiches Crossing over oder "Ausrutscher" der DNA-Polymerase während der Replikation erwarten würde. Beiden Mechanismen werden üblicherweise als die vorwiegenden Prozesse zur Generierung kurzer Insertionen und Deletionen vermutet. Wir stellten daher die Hypothese auf, dass ein anderer Mechanismus, die nicht-homologe End-zu-End-Verknüpfung zur Reparatur von Doppelstrangbrüchen, in diesem Zusammenhang möglicherweise ebenfalls eine wichtige Rolle spielt.

Tandemduplikationen könnten sich auch als die Quelle so genannter langreichweitiger Korrelationen in der Nukleotid-Zusammensetzung eukaryotischer Genome erweisen. Trotz ihrer Allgegenwart werden die zugrundeliegenden Entstehungsmechanismen dieser Korrelationen seit ihrer Entdeckung vor mehr als zehn Jahren kontrovers diskutiert. Wir konnten unter Verwendung formaler Methoden zur Analyse stochastischer Prozesse und nichtlinearer dynamische Systeme zeigen, dass Modelle lokaler Sequenzevolution, die auch Tandemduplikationsprozesse miteinbeziehen, zu einer Universalitätsklasse eindimensionaler Expansions-Randomisierungs Systeme mit langreichweitigen Korrelationen gehören.

Der die Skalierungseigenschaften von Korrelationsfunktionen in diesen Systemen definierende charakteristische Exponent wurde explizit für mehrere Evolutionsdynamiken der Universitätsklasse berechnet. Er wird von nur zwei effektiven Raten des jeweiligen dynamischen Modells bestimmt. Auch komplexere Evolutionsszenarien, in denen die Raten der Prozesse zeitlich und räumlich entlang der Sequenz variieren, wurden von uns untersucht. Wir schlussfolgerten aus unserer Analyse, dass die beobachteten langreichweitigen Korrelationen in der DNA durch Tandemduplikationsprozesse in der Genomevolution generiert worden sein können.

Der dominierende Anteil an Tandemduplikationen unter DNA Insertionen hat weitgehende Konsequenzen für statistische Aussagen bioinformatischer Methoden zur Sequenzanalyse, da sie zur Berechnung von Signifikanzwerten ihrer Vorhersagen oft auf ein Hintergrundmodell für DNA Sequenzen angewiesen sind. Wir konnten zum Beispiel zeigen, dass sich die p-Werte für Ähnlichkeitsscores von Sequenzalignments deutlich ändern, wenn man anstelle des Standard iid-Hintergrundmodells eines mit langreichweitigen Korrelationen von genomischer Stärke verwendet.

Das Konzept duplikationsgetriebener Evolution von Genen ist seit längerem bekannt. Es ist daher vielleicht gar nicht so überraschend, dass Tandemduplikationen auch bei Mutationsprozessen auf kürzeren Längenskalen eine wichtige Rolle spielen. Tatsächlich konnten wir zeigen, dass Duplikationen und Deletionen in protein-kodierenden Bereichen des menschlichen Genoms weniger schädlich im Vergleich zu nicht-synonymen Substitutionen sind, solange sie nicht das Leseraster des Proteins verschieben. Tandemduplikationen könnten sich auch auf kürzeren Längenskalen als wichtiger Motor zur schnellen Erzeugung neuer genetischer Variation und Funktion erweisen, so wie es für Gene und längere genomische Bereiche bereits allgemein anerkannt ist.