

Un nuovo algoritmo per la ricerca di tandem repeat nel genoma

De Fonzo V, Aluffi-Pentini F, Bersani E, Parisi V.

EuroBioPark c/o Parco Scientifico Tor Vergata - via della Ricerca Scientifica 1 - 00133 Roma

Il nostro gruppo si occupa da anni di studiare teoricamente i fenomeni legati a VNTR (variable number of tandem repeat); i risultati delle ricerche condotte ci hanno portato ad introdurre il concetto di "genetica dinamica" - "dynamical genetics" mutuato da Goldschmidt (1938) - per spiegare l'insorgenza di varie malattie e piu' in generale vari fenomeni legati alla dinamica del DNA [De Fonzo et al, 2000]. In particolare, rientrano nella genetica dinamica, oltre alle note malattie dovute a espansione di triplette (ad es. morbo di Huntington e sindrome da cromosoma X fragile [Sutherland & Richards, 1995]), anche altre dovute a non triplette (ad es. il diabete mellito e l'epilessia mioclonica [Virtaneva et al, 1997]) e fondamentali fenomeni quali i tumori, l'apoptosi e l'integrazione virale.

Abbiamo, quindi, implementato un programma di bioinformatica in grado di identificare i TR (tandem repeat) e di rappresentarli in una maniera facilmente leggibile.

In un articolo [De Fonzo et al, 1998] abbiamo introdotto i primi accenni di genetica dinamica e descritto una versione preliminare dell'algoritmo utilizzato basato su quelli noti di Needleman & Wunsch (1970) e Smith and Waterman (1981), a loro volta basati sulla programmazione dinamica (Bellman, 1957). Inoltre, abbiamo stabilito un criterio obiettivo per assegnare i punteggi di somiglianza tra le basi confrontate tra una ripetizione e l'altra.

L'algoritmo e' stato implementato in linguaggio C e il software e' stato a lungo testato, con risultati soddisfacenti, su tutte le sequenze contenenti VNTR associate a malattie umane, disponibili nella banca dati GenBank.

Stiamo per pubblicare una versione migliorata, che fornisce anche il consenso (limitazione principale del precedente) e in seguito produrremo una versione piu' potente in grado anche di ricostruire la storia delle successive mutazioni.

Referenze:

- Bellman, R. E. (1957). Dynamic programming. Princeton University Press. Serial T-267. Princeton, NJ.
- De Fonzo V., Bersani E., Aluffi-Pentini F., Castrignano' T., Parisi V. (1998). Are only repeated triplets guilty? *J. Theor. Biol.* 1994, 125-142.
- De Fonzo, V., Bersani, E., Aluffi-Pentini, F., and Parisi, V. (2000). A new look at the challenging world of tandem repeats. In press on *Medical Hypotheses*: Received 26 April 1999; Accepted 20 July 1999. Preprint ROM2F/99/027, 23 August 1999.
- Goldschmidt, R. (1938). *Physiological genetics*. McGraw-Hill. New-York.
- Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J.Mol. Biol.* 48, 443-453.
- Smith, T. F. & Waterman, M. S. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195-197.
- Sutherland, G. R. & Richards, R. I. (1995). Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. USA* 92, 3636-3641.
- Virtaneva, K., D'Amato, E., Miao, J., Koskiniemi, M., Norio, R., Avanzini, G., Franceschetti, S., Michelucci, R., Tassinari, C. A., Omer, S., Pennacchio, L. A., Myers, R. M., Dieguez-Lucena, J. L., Krahe, R., de la Chapelle, A. & Lehesjoki, A.-E. (1997). Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nature Genet.* 15, 393-396.