



II Simpósio sobre Inovação e Criatividade Científica na Embrapa

Brasília, 28 a 30 de abril 2010

Detecção de Erros de Montagens em Regiões Gênicas

Autor(es): Roberto H. Herai, Poliana F. Giachetto, Fabio D. Vieira, Edgard H. dos Santos, Michel E. B. Yamagishi, Paula R. Kuser-Falcão

Unidade(s): Embrapa Informática Agropecuária, UNICAMP

Muitos genomas ainda apresentam erros de montagem. Tais erros são particularmente graves se ocorrerem em regiões gênicas, pois muitos trabalhos científicos se concentram exatamente nessas regiões por razões óbvias. Uma forma de tentar identificar erros de montagens em regiões gênicas é utilizar sequências adquiridas de forma independente, como, por exemplo bases de ESTs ou bases de Full length cDNA (FlcDNA). O foco deste trabalho é propor uma metodologia de Bioinformática que utiliza bibliotecas de FlcDNA para detectar tais erros. Como prova de conceito aplicamos a metodologia no genoma do *Bos taurus* UMD 3.0. Utilizando-se uma biblioteca de cDNA gerada pelo consórcio MGC (*Mammalian Gene Collection*), identificamos alguns erros, sendo um exemplo ilustrativo o que ocorre na região do gene HDHD2, haloacid dehalogenase-like hydrolase domain containing 2. Mapeando-se o transcrito BC102232 no genoma, parte do mapeamento ocorreu inteiramente no contig 965 em sentido *antisense*, e outra parte foi mapeada no contig 963 em sentido *sense*. Entretanto, se alinharmos este mesmo transcrito contra o genoma de referência *Bos taurus*, build 4.1, obtemos um alinhamento satisfatório em sentido *antisense*. A correção da montagem pode ser facilmente realizada, assumindo-se que o transcrito alinha-se em sentido *antisense*, o contig 963 deverá manter sua posição, devendo-se considerar sua sequência reversa e complementar correspondente. Essa metodologia pode ser aplicada nos genomas de interesse da Embrapa.