

Resumo

Neste trabalho desenvolvemos uma plataforma de gerenciamento de dados e projetos, provenientes de estudos de genética de populações e epidemiologia genética, a DIVERGENOME. A plataforma apresenta dois componentes funcionais: A) uma base de dados relacional, o DIVERGENOMEdb, desenvolvida com o objetivo de armazenar os dados de forma segura e organizada e integrar diferentes fontes de informação (disponíveis em repositórios públicos e gerados localmente), dados genéticos (genótipos e haplótipos, provenientes de diferentes tipos de polimorfismos) e informações epidemiológicas (fenótipos, constituídos de variáveis qualitativas e quantitativas); e B) um conjunto de scripts para manipulação de formatos de arquivos, o DIVERGENOMETools, com o objetivo de otimizar a tarefa de conversão de formatos para análises em diferentes software, tarefa comprovadamente árdua e fonte de grande número de erros evidenciados nos resultados finais das análises. Nossa plataforma apresenta uma nova metodologia para a integração de diferentes scripts permitindo maior número possível de conversões e facilitando sua extensão. Uma primeira versão da ferramenta pode ser acessada em (pggenetica.icb.ufmg.br/pipelinedgh/). Os diferentes componentes da plataforma foram utilizados na condução dos trabalhos sobre a ação dos fatores evolutivos que moldam a diversidade genética apresentados na dissertação, mostrando-se eficientes às suas propostas. Para garantir a o acesso de forma rápida e ampla utilização de nossa plataforma pela comunidade científica desenvolvemos ainda uma interface web dessa forma não exigindo do usuário conhecimentos prévios de programação e gerenciamento de bancos de dados.

Palavras chave: genética de populações, epidemiologia genética, banco de dados, ferramentas bioinformáticas