

QDD: um software para identificação de regiões microssatélites e desenho de *primers* usando dados de sequenciamento de nova geração

Ueric José Borges de Souza

Programa de Pós-Graduação em Genética e Melhoramento de Plantas,
Universidade Federal de Goiás, Goiânia

Autor para correspondência - uericjose@gmail.com

Palavras-chave: marcadores moleculares, SSRs, NGS, bioinformática



Os primeiros métodos para isolamento e obtenção de seqüências microssatélites de um organismo foram realizados por meio de técnicas laboriosas que envolviam a construção de bibliotecas genômicas enriquecidas com regiões repetitivas ou bibliotecas de cDNA (DNA complementar). Entretanto, o desenvolvimento e constante aprimoramento das tecnologias de sequenciamento de nova geração (NGS - *Next-Generation Sequencing*) e dos programas de bioinformática para análise de dados têm possibilitado a identificação de centenas de milhares de regiões microssatélites, com posterior desenvolvimento de marcadores moleculares, para uso tanto em espécies modelos como nas não modelos.

Dentre os diversos programas de bioinformática existentes para identificar regiões microssatélites e realizar o desenho de *primers*, pode-se citar o QDD. O QDD é um programa desenvolvido em linguagem Perl, por um grupo de pesquisadores de instituições francesas (*Universidade Aix-Marseille, Montpellier SupAgro e Institut National*

de la Recherche Agronomique), disponível nas plataformas Windows e Linux, e que pode ser usado por meio de linhas de comando (`perl pipe1.pl -input_file arquivo.fasta`; Figura 1) ou pela plataforma Galaxy (`qddGalaxy`; Figura 2). Atualmente, o programa é distribuído gratuitamente no endereço eletrônico: <http://net.imbe.fr/~emeglecq/qdd.html>.

```

Terminal
qdd@qdd-VirtualBox: ~/galaxy-dist/tools/qdd
qdd@qdd-VirtualBox:~/galaxy-dist/tools/qdd$ ls
cut_vector_galaxy.pl      ncbi_taxonomy.pm      QDD_pipe1.xml          set_qdd_default.ini
data                     ncbi_tax.pl           QDD_pipe2.xml          set_qdd_default.ini.save
find_and_mask_ms_galaxy.pl pipe1.pl               QDD_pipe3.xml          sort_tag_galaxy.pl
Glycine_soja.fasta       pipe2.pl               QDD_pipe4.xml          subprogramQDD.pm
motifs_transmot_hach1_6_corr.txt pipe3.pl               QDD.pl
MS_extract.pl            pipe4.pl               QDDupdate.txt
ncbi_tax.db              qdd_output             remote_blast.pl
qdd@qdd-VirtualBox:~/galaxy-dist/tools/qdd$ perl pipe1.pl -input_file Glycine_soja.fasta
*****
QDD version3.1.2 July 2014
Emese Meglecz, Aix-Marseille University, Marseille, France
emese.meglecz@imbe.fr
http://net.imbe.fr/~emeglecq/qdd.html
*****
SCRIPT pipe1.pl
Input file with sequences: Glycine_soja.fasta
Selecting sequences with microsatellites

See log_file for summary:
/home/qdd/galaxy-dist/tools/qdd/qdd_output/Glycine_soja_pipe1_log_v1.txt
qdd@qdd-VirtualBox:~/galaxy-dist/tools/qdd$

```

Banded Columns

Figura 1. Imagem do uso do software QDD na plataforma Linux.

O programa funciona através de um processamento sequencial dividido em quatro etapas (*pipelines*), onde os resultados gerados pelo processamento de uma etapa (*outputs* ou arquivos de saída), servem como entrada (*inputs* ou arquivos de entrada) para a próxima etapa de execução.

O primeiro pipeline (PIPE1) é para identificação das regiões microssatélites. O programa usa como arquivo de entrada

os *reads* gerados pelo sequenciamento ou *contigs* e *scaffolds* obtidos pela montagem dos *reads*. Como resultado, é obtido um *output* em formato *fasta* (*pipe1_for_pipe2.fas*) que contém as regiões microssatélites identificadas para o organismo em estudo, e deve ser usado como *input* para a próxima etapa.

O segundo *pipeline* (PIPE2) realiza uma análise de similaridade entre os micros-

satélites identificados por BLAST de modo a manter somente sequências únicas (*singletons*), potenciais para o desenho de iniciadores ou *primers*. O terceiro *pipeline* (PIPE3) usa o programa Primer3 (<http://primer3.sourceforge.net/>) para o desenho de *primers*. Os parâmetros para o Primer3 podem ser selecionados pelo usuário e os resultados são sumarizados em uma tabela (*pipe3_primers.tabular*). Essa tabela é então usada como arquivo de entrada para a quarta etapa (PIPE4) que verifica se os microssatélites identificados possuem similaridade com elementos transponíveis (TEs) conhecidos utilizando o programa RepeatMasker.

O QDD pode ser utilizado para diversas atividades, tanto em pesquisa, quanto na sala de aula, para estudantes de graduação e pós-graduação. O site do programa contém diversos exemplos do seu uso, tanto por meio de linhas de comando como no qddGalaxy (http://net.imbe.fr/~emeglec/z/qdd_run.html#examples). Recomenda-se o seu uso em aulas práticas sobre microssatélites para que os estudantes adquiram familiaridade com este programa e suas aplicabilidades. Além disso, o site também tem uma série de sugestões para orientar pesquisadores na escolha de *primers* para a pesquisa genética (<http://net.imbe.fr/~emeglec/z/qdd.html#-choice>).

The image shows a screenshot of the Galaxy web interface. The main panel displays the configuration for the tool 'QDD_pipe1 (version 3.1.2)'. The 'Input fasta file' is set to '1: genome300_clean_chloromito.fasta'. The 'Input sequences are already assembled (contigs/scaffolds /chromosomes):' option is set to 'yes'. The 'Flanking region length:' is set to '200'. The 'Sequence length limit:' is set to '80'. Below the configuration is an 'Execute' button. The right-hand panel shows the 'History' section, listing previous runs with their names and sizes. The top of the history list shows '22: Table with primers, RepeatMasker and NCBI BLAST hit info' (14.0 MB). Below it are '21: pipe4 Log file', '19: Table with primers', '18: Sequences with primers', '17: pipe3 Log file', '15: Input for pipe3', '8: pipe2 Log file', '3: Input for pipe2', '2: pipe1 Log file', and '1: genome300_clean_chloromito.fasta'.

Figura 2.

Imagem do software QDD na plataforma Galaxy.