Aplicação do algoritmo genético no mapeamento de genes epistáticos em cruzamentos controlados

Paulo Tadeu Meira e Silva de Oliveira

TESE DE DOUTORADO

A SER APRESENTADO

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO COMO PARTE DOS REQUISITOS NECESSÁRIOS

PARA OBTENÇÃO DO GRAU DE

DOUTOR EM CIÊNCIAS

Área de Concentração: **Estatística**

Orientadora: Profa. Júlia Maria Pavan Soler

São Paulo, agosto de 2008

Aplicação do algoritmo genético

no mapeamento de genes epistáticos

em cruzamentos controlados

Este exemplar corresponde à redação

final da tese devidamente corrigida

e defendida por Paulo Tadeu Meira e Silva de Oliveira

e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Profa. Dra. Júlia Maria Pavan Soler (orientadora) – IME-USP.
- Prof. Dr. Heleno Bolfarine – IME-USP.
- Prof. Dr. Luiz Lebensztajn – EP-USP.
- Profa. Dra. Roseli Aparecida Leandro – ESALQ-USP.
- Prof. Dr. Heyder Diniz Silva - UFU

Não é porque as coisas são difíceis que nós não nos atrevemos; é porque nós não nos atrevemos que elas se tornam difíceis. (Sêneca)

SER ESPECIAL É...

Ter um sonho que se realiza no meio de muitas tormentas, ter um encontro com a vida, quando ela está por te deixar, ter um momento de luz no meio da escuridão, ter humildade para voltar no caminho, ter sabedoria para escolher a melhor hora para seguir.

Ser especial é... Ser o encontro da eternidade com seu tempo, ter o encontro das almas, ter a essência jorrando em raios por todos os poros, ser o encontro das águas turvas, com toda a beleza do mar azul, ser o poder das forças que une os corpos.

Ser especial é... Ver que você pode seguir o caminho do meio, o caminho que te leva ao encontro do equilíbrio, o caminho que te deixa em paz com os teus, o caminho que te faz voltar para dentro, como se buscasse a luz, que tantas vezes te deixou na escuridão.

Ser especial é... Poder sentir o amor nas veias que pulsam, sempre chamando e dizendo: Viva!
É a vida que te chama sempre, aproveita esse momento e reflita.
Quanto você já fez por seus sonhos, para encontrar a sua vida?
Quantas lágrimas já derramou no seu caminho?
Quantas vezes caminhou sozinho?

Especial é ter luz, sentir a calma, deixar que a angústia não lhe derrube, ter forças para lutar.

Ser especial é ser como você. Alma pura, com sabedoria nas palavras, força nos braços, lágrimas sem dor.
É saber sorrir da tristeza quando ela te angustia, é saber caminhar sozinh o, sem muletas, é saber ouvir o silêncio, é saber calar na multidão.
Refletir sempre...
Sentir infinitamente...
Viver eternamente, e sonhar... sempre!!!

Resumo da Minha Vida.

Eu nasci foi triste
Ninguém pode negar
Contra a pobreza e deficiência,
Sempre tive de lutar

Todos diziam que estava condenado,
Meu destino estava traçado
Tinha de se conformar.

Até pelos médicos,
Fui desenganado
Meu caso podia ser tratado,
Mas pouco tinha de esperar.

Falou mais alto
A voz da insistência
Com coragem e persistência
Fui levado a tratar.

Foi assim, entra ano e sai ano,
Entre enganos e desenganos.
Até conseguir foi me formar.

Sei que tem muito a fazer
Para minha vida melhorar,
Enquanto tiver um sonho
Para que possa realizar
Contra tudo contra todos
Sempre hei de lutar.
Por mais árduo que seja a barreira
Que tenha de passar.

Desejo que essa pequena mensagem
Sirva também de imagem
Para quem esta a desanimar
Por mais obstáculos que tenha de ultrapassar
Não devemos desistir do nosso sonho realizar.

(Paulo Tadeu)

Gotas de Esperança

A sensação da solidão passa, assim como o inverno dá lugar à primavera. Você encontrará seu caminho para a felicidade, assim como uma planta encontra seu caminho em direção ao sol, crescendo, mudando e avançando.

No dia em que você nasceu, em algum lugar, uma planta floriu, o sol brilhou ainda mais forte, e enquanto o vento se movia pelo oceano, ele sussurrou seu nome. Você é uma pessoa especial e o mundo é melhor porque você existe.

Sempre que sentir o cheiro da chuva, ouvir o canto dos pássaros, ou ver uma borboleta ao sol saiba que alguém pensa em você e lhe deseja muita paz.

Quando nuvens de tempestade escurecerem seu mundo, lembre-se que a amizade lhe oferece um abrigo seguro onde você nunca está sozinho. Lá, encontrará bem-estar, apoio e compreensão.

Os dias frios cinza e solitários não duram para sempre. Os pássaros sabem disso, e é por isso que eles cantam. Não desista, não admita sentimentos de fracasso; dúvidas vão e vêm, assim como as estações do ano. Quando tudo o que é bom parece perdido, lembre-se que a vida é um círculo, e a esperança mora no horizonte.

Se você sente o calor do sol em seu rosto, o cheiro da terra, o canto do pardal, então saiba que você é parte da natureza, com a sua própria singularidade, beleza e razão de ser. A vida não é sempre radiante, mas se o sol pode brilhar depois da pior tempestade, nós também podemos. Não importa quão frio o vento, quão escuro o dia; há calor dentro de um coração repleto de amor e compreensão.

A natureza nos oferece o calor do sol, o perfume das flores e o canto dos pássaros para nos fazer lembrar que não importa quão difícil a vida é; haverá sempre momentos de bondade, paz e oportunidade de crescimento.

Alguns dias são melhores, outros, é melhor esquecer; mas assim que o anoitecer marcar o final do dia, pense nas coisas boas da vida, coisas inocentes e verdadeiras, e adormeça sonhando com a esperança que traz o amanhã.

Quando o vento frio soprar desânimo em seu coração e o mundo parecer rancoroso, seja paciente e perseverante, para que sempre voltem os momentos de bondade e amor. Busque o vento
para o seu sonho, depois deixe seu coração voar livremente. Com coragem, fé e firmeza, dê tudo de si.

Cada dia que amanhece traz esperança e oportunidade de fazer os sonhos se tornarem realidade. Existe mágica na passagem do dia para a noite. Assim como as cores que se desfazem no
crepúsculo, trazendo esperança para um amanhã radiante. Você tem dentro de si mesmo energia para vencer. Pode transformar um obstáculo em um degrau que o leve um passo adiante na realização de seu sonho. Se nós possuímos a habilidade de
sonhar, o potencial de realizar sonhos também é nosso.

Estabeleça um objetivo e mantenha-se em sua trilha; não deixe que os infortúnios da vida o desviem de seu caminho.
Persistência, assiduidade e trabalho árduo nunca ficam sem recompensa, e os sonhos realmente tornam-se realidade.

Para envelhecer bem, devemos continuar a sonhar, pois é a busca de sonhos que nos mantém jovens de coração. Reflita na vida sempre que possível; reserve tempo para recordar, e faça tempo para sonhar. Encontre seu próprio caminho; vá com
confiança e. Surpresas boas virão.

Lynne Gerard

Perguntas De Um Operário Que Lê.

Quem construiu Tebas, a das sete portas?
Nos livros vem o nome dos reis,
Mas foram os reis que transportaram as pedras?
Babilònia, tantas vezes destruída,
Quem outras tantas a reconstruiu? Em que casas
Da Lima Dourada moravam seus obreiros?
No dia em que ficou pronta a Muralha da China para onde
Foram os seus pedreiros? A grande Roma
Está cheia de arcos de triunfo. Quem os ergueu? Sobre quem
Triunfaram os Césares? A tão cantada Bizâncio
Só tinha palácios
Para os seus habitantes? Até a legendária Atlântida
Na noite em que o mar a engoliu
Viu afogados gritar por seus escravos.

O jovem Alexandre conquistou as Índias
Sòzinho?
César venceu os gauleses.
Nem sequer tinha um cozinheiro ao seu serviço?
Quando a sua armada se afundou Filipe de Espanha
Chorou. E ninguém mais?
Frederico II ganhou a guerra dos sete anos
Quem mais a ganhou?

Em cada página uma vitòria.
Quem cozinhava os festins?
Em cada década um grande homem.
Quem pagava as despesas?

Tantas histórias
Quantas perguntas

...

Bertold Breecht

Dedico esse trabalho:
Ao Silvino Neves Rodrigues (“Foi ai que tudo começou”)
Aos meus pais Francisco e Geralda,
que me acompanhou sempre
nos bons e maus momentos,
na alegria e na tristeza, na concórdia e na discórdia.

Agradecimentos

Em primeiro lugar, o mais sincero e profundo agradecimento a minha orientadora Professora Dra Júlia Maria Pavan Soler por ter acreditado em meu potencial, pela sugestão do tema, pelos seus conselhos e críticas, pelo seu apoio em todos os momentos que vão desde os mais fáceis aos mais difíceis na evolução deste trabalho, e acima de tudo, pela sua disposição e paciência *ad infinitum* durante a elaboração desta tese.

Agradecimento especial aos meus colegas Elmo, Iran, Gladys, Gilberto Matos, Wiliam, Juvêncio e Caio pelo apoio dado em disciplinas como Probabilidade Avançada I.

Outro agradecimento especial aminhas colegas Núbia Esteban pela assessoria a minha defesa e a Jacqueline pelas fotos.

Agradeço também ao Centro Acadêmico do Instituto de Matemática e Estatística pela permissão de utilizar o mural para a colocação de anúncios divulgando o meu trabalho de assessoria que possibilitou o meu sustento durante os meus estudos de doutorado.

Agradeço também o incentivo dos demais professores e colegas do departamento de estatística e sobretudo a minha família pelo apoio durante a realização desse trabalho.

Outro agradecimento aos médicos Dr. José Eduardo Krieger e Dr. Alexandre Pereira do Laboratório de Cardiologia e Genética Molecular do Instituto do Coração – INCOR do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo pela concessão do banco de dados reais dos ratos utilizados nesse estudo e ao Cesar Henrique Torres pelas orientações no uso do WinQTLCart e pelo banco de dados dos ratos em Excel.

Por fim, agradeço a Universidade de São Paulo pela oportunidade de poder melhorar minha formação acadêmica e pelas facilidades dispensadas.

Resumo

O mapeamento genético é constituído por procedimentos experimentais e estatísticos que buscam detectar genes associados à etiologia e regulação de doenças, além de estimar os efeitos genéticos e as localizações genômicas correspondentes. Considerando delineamentos experimentais que envolvem cruzamentos controlados de animais ou plantas, diferentes formulações de modelos de regressão podem ser adotados na identificação de QTL's (do inglês, *quantitative trait loci)*, incluindo seus efeitos principais e possíveis efeitos de interação (epistasia). A dificuldade nestes casos de mapeamento é a comparação de modelos que não necessariamente são encaixados e envolvem um espaço de busca de alta dimensão.

Para este trabalho, descrevemos um método geral para melhorar a eficiência computacional em mapeamento simultâneo de múltiplos QTL's e de seus efeitos de interação. A literatura tem usado métodos de busca exaustiva ou busca condicional. Propomos o uso do algoritmo genético para pesquisar o espaço multilocos, sendo este mais útil para genomas maiores e mapas densos de marcadores moleculares. Por meio de estudos de simulações mostramos que a busca baseada no algoritmo genético tem eficiência, em geral, mais alta que aquela de um método de busca condicional e que esta eficiência é comparável àquela de uma busca exaustiva. Na formalização do algoritmo genético pesquisamos o comportamento de parâmetros tais como: probabilidade de recombinação, probabilidade de mutação, tamanho amostral, quantidade de gerações, quantidade de soluções e tamanho do genoma, para diferentes funções objetivo: BIC (do inglês, *Bayesian Information Criterion*), AIC (do inglês, *Akaike Information Criterion*) e SSE, a soma de quadrados dos resíduos de um modelo ajustado. A aplicação das metodologias propostas é também considerada na análise de um conjunto de dados genotípicos e fenotípicos de ratos provenientes de um delineamento F_2.

Palavras chave: algoritmo genético, seleção de modelos genéticos, epistasia

.

Abstract

Genetic mapping is defined in terms of experimental and statistical procedures applied for detection and localization of genes associated to the etiology and regulation of diseases. Considering experimental designs in controlled crossings of animals or plants, different formulations of regression models can be adopted in the identification of QTL's (Quantitative Trait Loci) to the inclusion of the main and interaction effects between genes (epistasis). The difficulty in these approaches of gene mapping is the comparison of models that are not necessarily nested and involves a multiloci search space of high dimension.

In this work, we describe a general method to improve the computational efficiency in simultaneous mapping of multiples QTL's and their interactions effects. The literature has used methods of exhausting search or conditional search. We consider the genetic algorithm to search the multiloci space, looking for epistatics loci distributed on the genome. Compared to the others procedures, the advantage to use such algorithm increases more for set of genes bigger and dense maps of molecular markers. Simulation studies have shown that the search based on the genetic algorithm has efficiency, in general, higher than the conditional search and that its efficiency is comparable to that one of an exhausting search. For formalization of the genetic algorithm we consider different values of the parameters as recombination probability, mutation probability, sample size, number of generations, number of solutions and size of the set of genes. We evaluate different objective functions under the genetic algorithm: BIC, AIC and SSE. In addition, we used the sample phenotypic and genotypic data bank. Briefly, the study examined blood pressure variation before and after a salt loading experiment in an intercross (F_2) progeny.

Key words: genetic algorithms, model selection, epistasis

Sumário