

NGHIÊN CỨU ĐÁNH GIÁ HIỆU QUẢ SỬ DỤNG CÁC LOẠI CHỈ THỊ PHÂN TỬ TRONG LẬP BẢN ĐỒ GIEN QUI ĐỊNH TÍNH TRẠNG SỐ LƯỢNG Ở BÒ

Mai Duy Minh¹, Goutam Sahana², Bernt Gulbrandsen² và Freddy B. Christiansen²

TÓM TẮT

Nghiên cứu mô phỏng nhằm so sánh hiệu quả của các mô hình sử dụng ba loại chỉ thị phân tử gồm nucleotit đa hình (SNP), haplotype (nhóm SNP liên kế trên đoạn nhiễm sắc thể NST) và block (nhóm SNP liên kế trên NST có tỉ lệ trao đổi chéo rất thấp) trong việc phát hiện gen qui định tính trạng số lượng (QTL). Phân tích ba bộ số liệu, mỗi bộ gồm có trình tự SNP của NST số 29 của bò sữa *Bos taurus*, trên đó có 5 QTL giá định. Đối với mỗi bộ số liệu, phân tích dữ liệu về kiểu gen SNP thực tế, hệ phá thực tế của 2351 cá thể bò đực và kiểu hình mô phỏng tương ứng của chúng (được lập lại 80 lần). Trong các mô hình, khi giá trị di truyền của các alen của chỉ thị được định nghĩa là biến ngẫu nhiên thì mô hình sử dụng SNP có khả năng tìm được QTL cao hơn và tỉ lệ công nhận QTL giá thấp hơn so với mô hình sử dụng haplotype hoặc block. Khả năng tìm được QTL giảm dần và tỉ lệ công nhận QTL giá tăng lên khi kích thước của haplotype hoặc block tăng lên. Kết quả nghiên cứu đề nghị sử dụng SNP riêng lẻ trong lập bản đồ QTL thay vì dùng các haplotype hoặc block.

Từ khóa: Chỉ thị phân tử, bản đồ gen, bò sữa.

1. MỞ ĐẦU

Trong nghiên cứu QTL, mô hình sử dụng các haplotype (kiểu đơn bội) (gồm hữu hạn các chỉ thị phân tử liên kế trên đoạn NST) được xem là cho hiệu quả tốt hơn so với sử dụng các chỉ thị phân tử đơn lẻ. Cơ sở khoa học của nhận định này là về mặt lý thuyết haplotype cho phép khai thác mức độ tổ hợp không ngẫu nhiên giữa các alen của gen nghiên cứu và alen của chỉ thị phân tử trong quần thể (gọi tắt là tổ hợp alen không ngẫu nhiên - linkage disequilibrium- LD) hiệu quả hơn nhờ đồng thời dựa vào các chỉ thị liên kế so với dựa vào từng chỉ thị riêng lẻ (Hayes, 2008). Nhiều nghiên cứu mô phỏng đã cho kết quả phù hợp với nhận định trên (Ví dụ Akey et al., 2001; Zaykin et al., 2002; De Bakker et al., 2005) và kích thước của haplotype (là số lượng của các chỉ thị trên đoạn NST cấu tạo haplotype) cho phép phân tích QTL hiệu quả cao nhất cũng đã được đề nghị (Grapes et al., 2006; Calus et al., 2009). Tuy vậy, gần đây một số kết quả trái ngược đã được ghi nhận, đó là mô hình sử dụng chỉ thị đơn cho hiệu quả cao hơn so với sử dụng haplotype (Zhao et al., 2007; Sahana et al., 2010). Trong các nghiên cứu này, haplotype được định nghĩa dựa trên việc mặc định số lượng các chỉ thị đơn liên kế thuộc một đoạn NST. Các nghiên cứu trên chưa kiểm chứng haplotype dạng block là đoạn NST

rất ngắn trên đó tần suất trao đổi chéo rất thấp vì mức LD giữa các đôi chỉ thị liên kế là rất cao (Wall & Pritchard, 2003). Ngoài ra kết quả thu được từ các nghiên cứu trên chủ yếu dựa vào phân tích các số liệu mô phỏng trên cơ sở lịch sử phát triển của quần thể nghiên cứu nên chưa hoàn toàn phản ánh được hiện trạng thực tế của quần thể.

Trong nghiên cứu này, tác giả đánh giá hiệu quả sử dụng các chỉ thị SNP, haplotype và block trong nghiên cứu QTL dựa trên phân tích số liệu kiểu gen mô phỏng, kiểu gen SNP thực tế và hệ phá thực tế của bò sữa.

II. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

1. Hệ phá và kiểu gen SNP

Nghiên cứu sử dụng kiểu gen SNP của NST 29 được xác định bằng chip Illumina Bovine SNP50 BeadChip (Matukumalli et al., 2009) của 2531 bò đực từ đàn bò sữa ở Đan Mạch có hệ phá rõ ràng. Các SNP được định vị trên NST theo phương pháp của Zimin et al. (2009) và pha liên kết (linkage phase) của haplotype được xác định bằng fastPHASE (Scheet & Stephens, 2006). Các SNP có tỉ lệ biến dị di truyền (có tần suất thấp hơn) cao hơn 0,05 đi lại và tổng số 764 SNP thuộc NST 29 được sử dụng cho nghiên cứu.

2. Giả định QTL và mô phỏng kiểu hình

Đối với mỗi bộ số liệu, dựa trên NST 29 QTL được giả định nhờ chọn ngẫu nhiên 11/764 SNP với điều kiện khoảng cách tối thiểu g

¹ Bộ Nông nghiệp và Phát triển nông thôn

² Trường Đại học Aarhus, Đan Mạch

TL là 5 cM. Ba bộ QTL được giả định tại các SNP vị trí lần lượt như sau: bộ thứ nhất là 120, 174, 376, 5, 673; thứ 2 là 29, 201, 414, 475, 671; và thứ 3 là 51, 6, 336, 492, 715. Vị trí của các QTL trên NST được in bày trong bảng 1. Các SNP được giả định là TL sẽ được loại bỏ khỏi bản đồ (NST 29) và vì thế chỉ có 759 SNP được phân tích. Tính phương sai di truyền của SNP theo công thức $2p(1-p)a^2$; trong đó a là ảnh hưởng di truyền chênh lệch giữa 2 alen của NP; p là tần suất của 1 trong 2 alen; a được hiệu chỉnh dựa trên tần suất của các alen của SNP trong quần thể sao cho mỗi QTL có phương sai = 0,02 tương đương sai kiểu hình. Giá trị di truyền của mỗi cá thể là tổng của giá trị di truyền của các QTL giả định và giá trị di truyền cộng gộp (polygenic effect). Giá trị di truyền cộng gộp của cá thể không rõ nguồn gốc được chọn ngẫu nhiên từ hàm phân phối chuẩn có $\mu=0$ và $\sigma^2=1$. Giá trị này của các cá thể ở thế hệ kế tiếp được tính bằng trung bình cộng của cá thể bố và mẹ cộng với giá trị hiệu chỉnh theo Mendel (Mendelian sampling term). Kiểu hình của mỗi bò được xác định là tổng của giá trị di truyền và phần không di truyền (random residual). Phương sai của phần không di truyền được xác định sao cho hệ số di truyền của tính trạng đạt 0,2. Trong mỗi bộ số liệu, giá trị kiểu hình được lặp lại 80 lần.

3. Thiết lập haplotype và block

Các haplotype được xây dựng dựa trên việc đếm số lượng SNP. Xuất phát từ trái qua phải NST 29, lần lượt từng SNP được chọn là chỉ thị đầu tiên để thiết lập haplotype. Sau khi SNP đầu tiên được chọn cho một haplotype, SNP liên kế bên phải được thêm vào cho đến khi đủ số lượng SNP cho haplotype theo yêu cầu. Có hai mức số lượng SNP được nghiên cứu là 2 và 4; tương ứng cho 2 loại haplotype được ký hiệu là Δp_2 và Δp_4 . Các block được thiết lập theo *Four Gamete Rule* dựa vào 2 SNP liên kế (Wang et al., 2002). Căn cứ vào việc tính toán tần số các alen kép alen được tạo thành từ cặp alen của 2 SNP liên kế và so sánh tần số này với ngưỡng nhất định để thiết lập các block. Trong nghiên cứu này có 2 ngưỡng được sử dụng là 0,01 và 0,1 tương ứng cho 2 loại block; được ký hiệu là BL01 và BL10.

4. Phân tích QTL

Mô hình: Thiết lập mô hình để phân tích trên cơ sở tham khảo Yu et al. (2006). Ảnh hưởng di truyền của các chỉ thị được định nghĩa là biến ngẫu nhiên.

Năm loại chỉ thị (SNP, Hap2, Hap4, BL01 và BL10) được sử dụng để phân tích theo mô hình như sau:

$$y_i = \mu + q_{in} + q_{mi} + s_i + e_i$$

Trong đó: y_i là giá trị kiểu hình của cá thể bò được thử i ; μ là giá trị trung bình quần thể; q_{in} và q_{mi} lần lượt là giá trị di truyền của alen chỉ thị có nguồn gốc từ bố và mẹ hiện diện trong cá thể bò được thử i và được giả định tuân theo hàm phân phối chuẩn $N(0, \sigma_q^2)$, trong đó σ_q^2 là phương sai di truyền của QTL giả định. Đồng phương sai di truyền của các alen khác nhau tại mỗi chỉ thị được giả định bằng 0; s_i là ảnh hưởng di truyền cộng gộp của cá thể bố của cá thể bò được thử i với $\mu=0$ và phương sai σ_s^2 ; e_i là phần không di truyền của cá thể được thử i .

Phân tích thống kê: Phân tích mô hình bằng phần mềm DMU (<http://gbi.agrsci.dk/dmu/>) dựa vào giá trị xác suất cực đại (Maximum likelihood) để kiểm định kết quả trên cơ sở so sánh giá trị này của mô hình có QTL ($\sigma_q^2 > 0$) và không có QTL ($\sigma_q^2 = 0$).

Kiểm soát sai số thống kê (FWER) theo luật Bonferroni ở đó $FWER_{\mu, n} = 1 - (1 - \alpha)^n \approx n\alpha$, trong đó α là mức kiểm định cho từng kiểm định trong số n kiểm định. Mức kiểm định chung cho mỗi mô hình là 0,05.

Định nghĩa QTL: Khi một SNP đạt mức ý nghĩa cao nhất được xác định, một QTL được ghi nhận tại vị trí của SNP đó. Khi một haplotype đạt mức ý nghĩa cao nhất được xác định thì một QTL được ghi nhận tại vị trí SNP đầu tiên và SNP thứ hai tương ứng cho mô hình Hap2 và Hap4. Khi xác định được block đạt mức ý nghĩa cao nhất một QTL được ghi nhận trên block đó. Đối với mô hình SNP hoặc haplotype, một QTL được công nhận nếu nó được tìm thấy trên đoạn NST tin cậy, bao gồm 10 SNP bên trái và 10 SNP bên phải vị trí QTL giả định. Đối với mô hình block, một QTL được công nhận nếu ít nhất một SNP thuộc block đó nằm trên đoạn NST tin cậy. QTL được ghi nhận không đáp ứng các tiêu chí kể trên được định nghĩa là QTL giả.

Tính hiệu quả tìm thấy QTL của mô hình theo công thức: $\frac{\#QTL}{1200} * 100$; trong đó 1200 (80 lần lặp lại

của kiểu hình x 3 bộ số liệu x 5 mô hình) là tổng số QTL có thể được phát hiện; $\#QTL$ là tổng số QTL được công nhận trong số $80 * 3 = 240$ lần phân tích.

Tính tỉ lệ công nhận QTL giả (*FDR*) theo công thức:

$$\frac{tsQTL}{tsQTL + nQTL} * 100$$
 trong đó *tsQTL* là tổng số QTL giả được ghi nhận từ 240 lần phân tích.

III. KẾT QUẢ VÀ THẢO LUẬN

1. Sự hiện diện của QTL giả định trên block

Các QTL giả định hiện diện ngẫu nhiên trên các block mà không phụ thuộc vào kích thước của block (bảng 1). Ở mức tần suất 0,01 (BL01) có 6 QTL không thuộc block nào. Tuy nhiên có 5, 3 và 1 QTL hiện diện trên các block có kích thước tương ứng là 1, 2 và 6 SNP. Ở mức tần suất 0,10 (BL10), có 2 QTL không thuộc block nào trong khi có 2, 3, 3 và 1 QTL hiện diện trên các block có kích thước tương ứng là 1, 2, 3 và 4 SNP. Ngoài ra có 4 QTL hiện diện trên các block có kích thước lớn hơn 4 SNP. Trong mô hình BL10, các QTL giả định hiện diện trên các block có kích thước lớn hơn (lớn nhất là 9) so với ở mô hình BL01 (lớn nhất là 6).

Bảng 1: Hiện diện của các QTL giả định trên các block

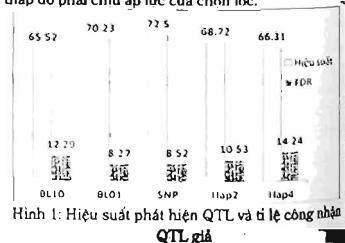
Bộ số liệu	QTL giả định		Kích thước block	
	SNP	Vị trí (base)	BL01	BL10
1	120	9446047	0	3
	174	12750459	1	2
	376	27510099	0	1
	515	36166067	0	0
	673	47146262	0	0
2	29	2015095	0	2
	201	15129552	1	3
	414	30200387	2	3
	475	33881797	1	1
	671	46927411	1	5
3	51	3769719	2	2
	206	15535268	0	5
	336	25186997	1	9
	492	34758033	2	4
	715	49207370	6	7

2. Khả năng tìm thấy QTL và mức độ công nhận QTL giả

Kết quả về khả năng tìm thấy QTL và mức độ công nhận QTL giả của 5 mô hình được trình bày trong hình 1. Xuất phát từ vị trí của mô hình SNP và quan sát về hướng trái hoặc phải (kích thước của các

haplotype hoặc block tăng lên) thì khả năng tìm thấy QTL giảm đi và tỉ lệ công nhận QTL giả tăng lên. Mô hình SNP có khả năng phát hiện ra QTL cao nhất (72,5%) và tỉ lệ công nhận QTL giả là thấp nhất (8,27%). Khả năng phát hiện ra QTL ở mô hình BL10 là thấp nhất, trong khi tỉ lệ công nhận QTL giả cao nhất ở mô hình Hap4.

Kết quả phân tích cho thấy mô hình sử dụng SNP đem lại hiệu quả cao hơn so với mô hình sử dụng haplotype hoặc block về khả năng phát hiện QTL và mức độ công nhận QTL giả. Phát hiện này phù hợp với kết quả đã được một số tác giả công bố (Zhao et al., 2007; Sahara et al., 2010) nhưng trái ngược với kết quả đã được công bố của nhiều công trình khác, đó là mô hình dùng haplotype cho hiệu quả tốt hơn dùng chỉ thị đơn (Akey et al., 2001; Grapes et al., 2006; Calus et al., 2009). Tính ưu việt và hạn chế của việc sử dụng haplotype trong phân tích QTL là cơ sở để giải thích cho kết quả trái ngược thu được từ các nghiên cứu khác nhau. Ưu thế về hiệu quả phân tích của mô hình sử dụng haplotype so với SNP phụ thuộc vào sự cân đối giữa hiệu quả thu được từ khai thác tính ưu việt và hạn chế do sự bất cập khi sử dụng haplotype và sự cân đối này phụ thuộc vào một số yếu tố khác nhau. Rõ ràng về lý thuyết dùng haplotype sẽ giúp khai thác được nhiều thông tin về LD giữa QTL và chỉ thị phân tử qua đó tăng khả năng phát hiện QTL và định vị nó chính xác hơn trên bản đồ gen (Hayes, 2008). Thêm vào đó, khả năng phát hiện QTL tăng lên nếu tần suất alen của chỉ thị tương tự như tần suất alen của QTL (Kaplan & Morris, 2001) và các haplotype thường có các alen ở tần suất thấp do đó sử dụng các haplotype có thể gia tăng khả năng phát hiện gen gây bệnh (Akey et al., 2001; De Bakker et al., 2005) vì các alen của gen gây bệnh thường có tần suất quần thể rất thấp do phải chịu áp lực của chọn lọc.



Hình 1: Hiệu suất phát hiện QTL và tỉ lệ công nhận QTL giả

Tuy nhiên việc sử dụng haplotype cũng bộc lộ một số hạn chế. Trong mô hình sử dụng haplotype, số lượng các alen của haplotype sẽ tăng lên khi kích thước của haplotype tăng lên. Trong trường hợp định nghĩa giá trị di truyền của các alen này là biến cố định thì mức độ chính xác của giá trị trung bình ước tính của các alen này sẽ giảm đi trong trường hợp kích thước mẫu phân tích không đủ lớn. Trong một quần thể xác định khi kích thước của haplotype tăng lên thì số loại alen tại haplotype đó tăng lên và số lượng của một loại alen của haplotype đó giảm đi, qua đó làm giảm độ chính xác trong ước tính giá trị trung bình của alen này. Điều này sẽ gia tăng khả năng công nhận QTL giả (Kolbehdari & Robinson, 2007). Trong trường hợp định nghĩa giá trị di truyền của các alen của haplotype là biến ngẫu nhiên và đồng phương sai của các alen của chi thị được so sánh với đồng phương sai của kiểu hình, nhờ đó giảm thiểu các tham số trong mô hình phân tích qua đó, làm tăng độ tin cậy của các kiểm định. Tuy vậy sử dụng các haplotype vẫn kèm theo những bất cập. Việc tính toán đồng phương sai của các alen khác nhau tại mỗi haplotype là khá phức tạp do liên quan đến nhiều chi thị phân tử. Trong trường hợp này, để đơn giản việc tính toán ở hầu hết các mô hình giá trị này thường được giả định bằng không. Giải pháp này có nguy cơ ảnh hưởng nhiều đến kết quả phân tích khi các alen khác nhau của một haplotype nhưng mang cùng 1 alen của QTL. Trong trường hợp này đồng phương sai chính là phương sai của QTL. Mức độ ảnh hưởng của yếu tố này tỉ lệ thuận với kích thước của haplotype. Trong nghiên cứu hiện tại ảnh hưởng của yếu tố này là khá lớn thể hiện qua kết quả phân tích, đó là khả năng phát hiện QTL giảm dần, tỉ lệ công nhận QTL giả tăng dần khi xem xét kết quả phân tích theo thứ tự mô hình SNP, Hap2 và Hap4 hoặc SNP, BL01 và BL10. Chính vì ảnh hưởng này nên ở các nghiên cứu trước đây (Grapes et al., 2006; Calus et al., 2009) hiệu quả xác định QTL chỉ tăng lên cho đến khi haplotype đạt kích thước nhất định, vượt quá mức đó hiệu quả phân tích giảm đi rõ rệt. Trong các mô hình sử dụng haplotype hoặc block ảnh hưởng của việc bỏ qua đồng phương sai của các alen tại mỗi haplotype vượt trội so với hiệu quả khai thác LD từ các SNP liên kề, do đó hiệu quả nghiên cứu của mô hình SNP vẫn cao hơn so với mô hình haplotype hoặc block.

Trong các nghiên cứu mô phỏng trước đây, tác giả sử dụng hầu hết là chỉ thị microsatellite (vi vệ

tinh), số liệu về kiểu gen, hệ phả và kiểu hình phục vụ cho phân tích QTL được xây dựng hoàn toàn do giả định. Nghiên cứu hiện tại sử dụng kiểu gen SNP thực tế và hệ phả thực tế của các quần thể bò sữa nên có thể mô phỏng hiện trạng gắn với điều kiện thực tế hơn so với các nghiên cứu trước đây dựa hoàn toàn vào các số liệu giả định.

IV. KẾT LUẬN

Trong nghiên cứu QTL, khi sử dụng số liệu về kiểu hình mô phỏng và QTL giả định kết hợp với kiểu gen SNP thực tế và hệ phả thực tế của bò, mô hình phân tích sử dụng SNP cho kết quả tốt hơn so với mô hình sử dụng haplotype hoặc block về khả năng phát hiện QTL và mức độ công nhận QTL giả. Từ kết quả nghiên cứu đã đạt được đề nghị sử dụng SNP đơn lẻ thay vì các haplotype trong lập bản đồ QTL ở bò.

Lời cảm ơn

Kết quả nghiên cứu thực hiện tại Trường Đại học Aarhus, Vương quốc Đan Mạch niên khoá 2007-2010, sử dụng số liệu từ các dự án do Chính phủ Việt Nam và Chính phủ Đan Mạch đồng tài trợ.

TÀI LIỆU THAM KHẢO

1. Akey J., L. Jin, and M. Xiong, 2001. Haplotypes vs. single marker linkage disequilibrium tests: what do we gain? *Eur. J. Hum. Genet.* 9:291-300.
2. Calus M. P. L., T. H. E. Meuwissen, J. J. Windig, E. F. Knol, C. Schrooten, A. L. J. Vereijken and R. F. Veerkamp, 2009. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet. Sel. Evol.* 41:11, doi:10.1186/1297-9686-41-11.
3. De Bakker P. I. W., R. Yelensky, I. Pe'er, S. B. Gabriel, M. J. Daly, and D. Altshuler, 2005. Efficiency and power in genetic association studies. *Nature Genetics* 37: 1217-1223.
4. Grapes L., M. Z. Firat, J. C. M. Dekkers, M. F. Rothschild, and R. L. Fernando, 2006. Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. *Genetics* 172: 1955-1965.
5. Hayes, B. J., 2008. QTL mapping, MAS, and genomic selection. A short-course organized by Department of Animal Science. Iowa State University. 120 pages.
6. Kaplan, N., and R. Morris, 2001. Issues concerning association studies for fine mapping a

susceptibility gene for a complex disease. *Genet. Epidemiol.* 20, 432-457.

7. Kolbehdari D. and J. A. B. Robinson, 2007. QTL Mapping Using Multiple Markers Simultaneously. *American Journal of Agricultural and Biological Science* 2:195-201.

8. Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. Smith, T. S. Sonstegard, and C. P. Van Tassell, 2009. Development and characterization of a high density SNP genotyping assay in cattle. *PLoS One* 4(4):e5350. Epub 2009 Apr 24.

9. Morris, R. W., and N. L. Kaplan, 2002. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genet. Epidemiol.* 23, 221-233.

10. Scheet, P., and M. Stephens, 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629-644.

11. Sahana G., B. Guldbrandtsen, L. Janss, and M. S. Lund, 2010. Comparison of association mapping methods in a complex pedigreed population. *Genet. Epidemiol.* [Accepted].

12. Wall J. D. and J. K. Pritchard, 2003. Assessing the Performance of the Haplotype Block Model of Linkage Disequilibrium. *Am. J. Hum. Genet.* 73: 502-515.

13. Wang N., J. M. Akey, K. Zhang, Chakraborty, and L. Jin, 2002. Distribution Recombination Crossovers and the Origin Haplotype Blocks: The Interplay of Population History, Recombination, and Mutation. *Am. J. Hum. Genet.* 71(5) pp. 1227 - 1234

14. Yu, J., G. Pressoir, W. H. Briggs, I. Vroh B. M. Yamasaki, J. F. Doebley, M. D. McMullen, B. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler, 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38:203-208.

15. Zaykin, D. V., Westfall, P. H., Young, S. S. Karnoub, M. A., Wagner, M. J., and M. G. Ehm, 2002. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum. Hered.* 53, 79-91.

16. Zhao H. H., R. L. Fernando, and J. C. M. Dekkers, 2007. Power and Precision of Alternative Methods for Linkage Disequilibrium Mapping of Quantitative Trait Loci. *Genetics* 175: 1975-1986.

17. Zhao H., D. Nettleton, M. Soller, and J. C. M. Dekkers, 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet. Res.* 86: 77-87.

18. Zimin, A., Delcher, A., Florea, L., et al., 2009. A whole-genome assembly of the domestic cow (*Bos taurus*). *Genome Biology* 10: R42.

THE EFFECTS OF INDIVIDUAL MARKERS VERSUS MARKER HAPLOTYPES ON POWER AND FALSE DISCOVERY RATE IN GENOME-WIDE ASSOCIATION STUDIES

Minh Duy Mai, Goutam Sahana, Bernt Guldbrandtsen, and Freddy B. Christiansen

Summary

A comparison of the effects of single nucleotide polymorphism (SNP) marker, overlapping fixed-size haplotype and block constructed on the basis of non-random association of markers by genome-wide association was done. Three datasets each contains five simulated quantitative trait loci (QTL) on SNP map of *Bos taurus* 29 and 80 replicates were analyzed. The SNP real genotype, real pedigree and simulated phenotypes of 2,531 bulls from Danish Holstein dairy cattle were used. When marker alleles were treated as random effect, the models using individual SNP gave better results than using overlapping fixed-size haplotype or block in terms of power and false discovery rate. The power decreased while the false discovery rate increased as the size of used haplotype or block increased. The results suggest use of individual SNP rather than SNP based haplotype or block in mapping genes causing genetic variation in quantitative traits.

Keywords: Molecular indicator, gene map, dairy cow.

Người phản biện: TS. Nguyễn Thị Thanh Thủy

Ngày nhận bài: 18/8/2012

Ngày thông qua phản biện: 17/9/2012

Ngày duyệt đăng: 3/10/2012