

GWAS: TIỀM NĂNG ỨNG DỤNG VÀ NHỮNG THÁCH THỨC TRONG NGHIÊN CỨU CHỌN TẠO GIỐNG LÚA

Tạ Kim Nhung, Khổng Ngân Giang, Phùng Thị Phương Nhung, Lê Huy Hàm, Đỗ Năng Vịnh

Viện Di truyền nông nghiệp

Stephane Jouannic

Viện Nghiên cứu phát triển của Pháp (IRD);

Phòng thí nghiệm liên kết Việt - Pháp (LMI RICE), Hanoi, Vietnam

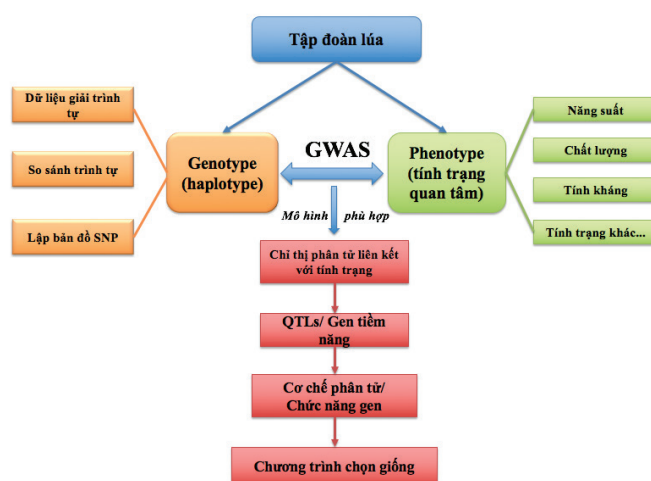
Trong một thập kỷ gần đây, cùng với sự phát triển của khoa học và công nghệ, việc giải trình tự gen và xây dựng bản đồ điểm đa hình đơn nucleotide (SNP) có độ phân giải cao đã làm sáng tỏ nhiều yếu tố di truyền ở nhiều loại cây trồng, đặc biệt là cây lúa (*Oryza spp.*). Đối với các tính trạng nông học phức tạp như năng suất, chất lượng, khả năng chống chịu của một quần thể lúa thì nghiên cứu liên kết trên toàn hệ gen (GWAS - Genome Wide Association Study) là công cụ vô cùng hữu hiệu. GWAS cung cấp cái nhìn sâu sắc về các tính trạng nông học trong mối tương quan với kiểu gen, qua đó cung cấp một số lượng lớn các locus tính trạng số lượng (QTL) và gen tiềm năng cho các nghiên cứu tiếp theo. Tuy nhiên, khi ứng dụng GWAS trong các nghiên cứu tìm kiếm QTL và gen tiềm năng cũng gặp phải một số thách thức nhất định.

Mở đầu

Là một trong những cây lương thực quan trọng bậc nhất trên thế giới, những hiểu biết về cơ chế di truyền liên quan đến sự sinh trưởng, phát triển, chống chịu với các điều kiện bất lợi của môi trường và đa dạng hình thái của cây lúa châu Á (*Oryza sativa*) có ý nghĩa lớn trong việc bảo đảm an ninh lương thực. Trong số các ứng dụng công nghệ sinh học được áp dụng trong nhiều thập kỷ qua (nuôi cấy bao phấn, ưu thế lai, gây đột biến, cây trồng chuyển gen...), lập bản đồ QTL dựa trên quần thể bố mẹ (bi-parental population) vẫn là kỹ thuật phổ biến nhất trên thế giới được ứng dụng trong các chương trình chọn tạo giống.

Mặc dù gặt hái được nhiều thành công, nhưng QTL rõ ràng không phải là phương pháp tối ưu để khai thác nguồn gen to lớn với kiểu hình đa dạng của hơn 120.000 giống lúa, nguyên nhân là do: 1) Giới hạn số lượng allen quan tâm vì e ngại sự phân ly giữa bố mẹ từ đời F_2 của các dòng tái tổ hợp; 2) Bản đồ QTL có mức độ phân giải bị giới hạn, vì vậy khoảng tin cậy của QTL thường có kích thước rất lớn. Trong khi đó, GWAS với sự trợ giúp của thể hệ giải trình tự mới cung cấp bộ marker bao phủ toàn hệ gen với độ phân giải cao, hoàn toàn không chịu những giới hạn trên, đã trở thành công cụ mạnh mẽ trong việc nghiên cứu đa dạng di truyền, đặc biệt là ở những tính trạng nông học phức tạp. Nguyên tắc cơ bản của GWAS là đánh giá

mối tương quan giữa mỗi chỉ thị di truyền với tính trạng quan tâm trong một quần thể cùng loài. GWAS cung cấp cái nhìn sâu sắc về đặc tính di truyền của các tính trạng trên, cho phép lựa chọn các cặp bố mẹ tốt nhất để phân tích QTL, cũng như các gen tiềm năng quy định tính trạng quan tâm (hình 1).



Hình 1: phương pháp phân tích GWAS

GWAS được sử dụng lần đầu tiên hơn 10 năm trước trong nghiên cứu di truyền ở người, đến nay đã có hơn 1.500 công bố ở người, các sinh vật mô hình cũng như cây

trồng, đặc biệt là cây lúa. Với diện tích phân bố rộng khắp thế giới với hơn 120.000 giống khác nhau (theo Ngân hàng lúa quốc tế), cây lúa thể hiện sự đa dạng về nguồn gen và trở thành nguồn vật liệu lý tưởng cho các nghiên cứu bằng GWAS.

Một số kết quả nghiên cứu GWAS ở cây lúa

Để đáp ứng nhu cầu lương thực ngày càng tăng do bùng nổ dân số cũng như quá trình đô thị hóa và khí hậu thay đổi, cải thiện, duy trì ổn định năng suất lúa gạo cũng như tăng cường tính kháng luôn là yêu cầu cấp bách đối với nền nông nghiệp của các quốc gia. Đối với tính trạng năng suất, đây là một tính trạng nông học phức tạp, chịu ảnh hưởng trực tiếp bởi số bông/cây, số hạt/bông, trọng lượng hạt, và chịu ảnh hưởng gián tiếp của các tính trạng nông học khác như chiều cao cây, thời gian ra hoa, số nhánh/cây, hình thái, màu sắc và số lượng lá/cây... Nếu các phương pháp truyền thống thường chỉ có thể tập trung vào một vài tính trạng để nghiên cứu thì GWAS lại có thể áp dụng nghiên cứu trên cả tập đoàn cùng lúc nhiều tính trạng. Trong vài năm gần đây, số lượng lớn dữ liệu về nguồn gen được công bố không ngừng tăng lên ở cả số lượng trình tự gen và bản đồ SNP có độ phân giải cao.

Trên thế giới

Hiện nay có hơn 400 QTL liên quan đến tính trạng hình thái hạt được phát hiện bằng nhiều phương pháp lập bản đồ khác nhau (<http://www.gramene.org/qtl>) nhưng chỉ có khoảng 20 gen liên quan đến kích thước và trọng lượng hạt được phân lập bằng phương pháp cloning. Bằng cách áp dụng GWAS trên tập đoàn lúa *Indica* Trung Quốc, Huang và cộng sự (2011) tìm được 32 loci mới liên quan đến thời gian ra hoa và 10 loci liên quan đến kích thước hạt. Feng và cộng sự (2016) đã công bố 27 loci và 424 gen tiềm năng liên quan trực tiếp đến hình thái hạt, trong đó có 16/27 loci trên nằm gần với các QTL quy định hình thái hạt đã được công bố trước đây, đáng chú ý là 2 QTL quy định chiều dài (GS3) và chiều rộng hạt (qSW5). Trong khi đó, trên tập đoàn lúa Trung Quốc, Si và cộng sự (2016) phát hiện một QTL lớn trên nhiễm sắc thể số 7 chứa gen *OsSPL13* quy định kích thước hạt và cấu trúc bông thông qua việc điều hoà quá trình phân chia tế bào. Không chỉ các giống bản địa, tập đoàn các giống lúa lai cũng trở thành nguồn vật liệu quan trọng phục vụ nghiên cứu GWAS. Với tập đoàn 369 giống lúa lai ưu tú (IRRI), Begum và cộng sự (2015) công bố 52 QTL mới liên quan đến 11 tính trạng nông học quan trọng gồm nhiều QTL lớn quy định thời gian ra hoa, cao cây, năng suất lúa, chiều dài, chiều rộng và tỷ lệ dài/rộng ở hạt.

Các phân tích GWAS trên tính trạng cấu trúc bông cũng bùng nổ trong một vài năm gần đây, không chỉ tìm lại các QTL từng được công bố như một minh chứng cho tính

hiệu quả của phương pháp, mà còn tìm ra rất nhiều QTL mới. Từ các nghiên cứu GWAS trên các tính trạng ảnh hưởng trực tiếp đến năng suất như trọng lượng 1.000 hạt, số lượng bông/cây, số gié sơ cấp/bông, số gié thứ cấp/bông, dài trục bông, số nhánh/bông, số hạt/bông đến các nghiên cứu trên một số tính trạng ảnh hưởng gián tiếp đến năng suất như chiều cao cây, thời gian ra hoa, kích thước, số lượng, màu sắc lá, hàm lượng diệp lục, chất lượng hạt... Những thành tựu này phản ánh sự phổ biến, mức độ đáng tin cậy cũng như tiềm năng to lớn của các nghiên cứu bằng GWAS trên các tính trạng nông học ở lúa.

Bên cạnh tính trạng năng suất, tăng cường tính kháng ở lúa cũng là một thách thức lớn đối với các nhà chọn giống. Đối với tính kháng mặn, đến nay mới chỉ có vùng QTL *Saltol* chứa 3 gen (*SKC1*, *Salt* và *pectinesterase*) được nghiên cứu chức năng. Năm 2015, nghiên cứu GWAS đầu tiên ở tính trạng chịu mặn được công bố bởi Kumar và cộng sự, kết quả tìm thấy một QTL bao phủ cả vùng *Saltol* và một số QTL khác có ảnh hưởng ở mức độ nhỏ và trung bình đến khả năng kháng mặn. Đối với tính chống chịu hạn, hai nghiên cứu đầu tiên bằng GWAS được công bố năm 2013.

Ở Việt Nam

Trong hai năm 2011-2012, thông qua dự án hợp tác giữa Viện Di truyền nông nghiệp với Trung tâm Nghiên cứu John Innes và Trung tâm Phân tích genome, 36 giống lúa bản địa ưu tú của Việt Nam được giải trình tự illumina 7x, dự kiến đến năm 2017 dự án sẽ hoàn thành việc giải mã trên 300 giống lúa Việt Nam.

Bên cạnh đó, thông qua dự án hợp tác Việt - Pháp, Phùng và cộng sự đã xây dựng một quần thể các giống lúa bản địa đại diện cho các hệ sinh thái canh tác đa dạng của Việt Nam gồm 182 giống được phân tích với gần 30.000 SNP bằng kỹ thuật genotyping by sequencing (GBS) tạo tiền đề cho các nghiên cứu GWAS đối với các tính trạng như cấu trúc bộ rễ, cấu trúc bông... Dựa trên phân tích kiểu gen của 182 giống lúa (115 *indica* và 64 *japonica*) thu thập từ khắp các tỉnh/thành phố trong cả nước, các nhà khoa học Việt Nam đã tiến hành phân tích GWAS trên các tính trạng liên quan đến cấu trúc bộ rễ như: dài rễ, trọng lượng rễ ở các độ sâu khác nhau, độ dày rễ, số lượng rễ bên, tỷ lệ rễ/chồi và một số tính trạng khác ảnh hưởng đến khả năng đâm sâu và hấp thụ nước, qua đó tăng cường khả năng chịu hạn. Kết quả nghiên cứu đã tìm ra 2 QTL quan trọng và một số gen tiềm năng trên nhiễm sắc thể số 2 và 11 ảnh hưởng lớn đến tính trạng độ dày của rễ và số lượng rễ bên.

Thành công của các dự án giải trình tự hệ gen lúa tại Việt Nam nói riêng và trên thế giới nói chung đã cung cấp nguồn dữ liệu vô cùng phong phú cho các nghiên cứu

GWAS trên cây lúa.

Thách thức trong nghiên cứu GWAS ở Việt Nam

Tuy chỉ mới xuất hiện trong vài năm gần đây, nhưng các nghiên cứu GWAS đã được áp dụng trên hầu hết các tính trạng nông học quan trọng ảnh hưởng đến năng suất, chất lượng cũng như tính kháng ở nhiều tập đoàn lúa. Các phân tích GWAS không chỉ tìm ra các QTL đã được biết đến mà còn đưa ra giả thiết về sự tồn tại của nhiều QTL và gen tiềm năng mới đáng quan tâm. Bằng cách này, các nhà chọn giống có thể thu hẹp khoảng cách từ việc xác định QTL quan trọng đến xác định gen tiềm năng đáp ứng cho các chương trình chọn giống tăng tính kháng cũng như tối ưu hoá năng suất và chất lượng lúa. Mặc dù gặt hái được nhiều thành công trong vài năm trở lại đây, song các nghiên cứu GWAS trên thế giới nói chung, ở Việt Nam nói riêng cũng gặp phải một số thách thức sau:

Một là, để tiến hành nghiên cứu GWAS cần 4 yếu tố quan trọng: (1) Tập đoàn đủ lớn đảm bảo độ đa dạng và phù hợp với yêu cầu nghiên cứu; (2) Phân tích kiểu gen với SNP bao phủ toàn bộ hệ gen; (3) Phân tích kiểu hình và (4) Chọn được mô hình phân tích phù hợp. Đối với yếu tố thứ nhất, tập đoàn sử dụng trong nghiên cứu phải lớn, đảm bảo kết quả phân tích thống kê đủ mạnh để tìm được những mối liên kết dù nhỏ giữa kiểu gen và kiểu hình. Tuy nhiên nguyên tắc “càng nhiều, càng tốt” không phải luôn đúng, bởi sự đa dạng di truyền cũng như mối quan hệ giữa các cá thể trong tập đoàn mới là nhân tố ảnh hưởng nhiều đến kết quả của GWAS. Đối với yếu tố thứ hai, số lượng SNP bao phủ hệ gen hay các dữ liệu giải trình tự thường xuyên được cập nhật đều có thể dễ dàng truy cập từ các website như gramene, oryzasnp, aws, msu... Tuy nhiên, nếu muốn phân tích GWAS trên một tập đoàn mới, việc giải trình tự cũng như lưu trữ dữ liệu cần nguồn kinh phí lớn và các chuyên gia về tin sinh học. Vấn đề thứ 3 - phân tích kiểu hình - công việc quan trọng bậc nhất trong nghiên cứu. Việc đánh giá kiểu hình của hàng trăm đến hàng ngàn cá thể yêu cầu quy chuẩn đảm bảo độ đồng nhất, nhất là với các thí nghiệm đồng ruộng là không hề đơn giản. Để đảm bảo cho chất lượng của các dữ liệu về kiểu hình, thí nghiệm cần được lặp lại và nên được lặp lại trên một vài điều kiện môi trường khác nhau. Cuối cùng là phải lựa chọn được mô hình phân tích phù hợp thông qua các thuật toán mô tả mối quan hệ giữa các cá thể trong quần thể. Các mô hình tính toán như tuyến tính hỗn hợp (mixed linear), hỗn hợp nhiều locus (multi-locus mixed) và hỗn hợp nhiều tính trạng (multi-trait mixed) cần được xây dựng, phát triển và tối ưu hoá để giải quyết vấn đề trên. Hiện nay, mô hình được sử dụng nhiều nhất là tuyến tính hỗn hợp kết hợp với ma trận mô tả quan hệ của các cá thể trong quần thể. Các mô hình nghiên cứu vẫn được hoàn thiện từng ngày để giúp tăng thêm sức mạnh của GWAS.

Hai là, khi tiến hành phân tích các tính trạng nông học bằng GWAS là cung cấp thông tin và kết quả có thể sử dụng ngay cho các chương trình chọn tạo giống (ví dụ như cung cấp các chỉ thị phân tử liên kết với tính trạng quan tâm). Tuy nhiên, kết quả của GWAS dựa trên mô hình thống kê, do vậy QTL và gen tiềm năng cần được nghiên cứu và chứng minh trước khi đưa vào các chương trình chọn giống. Mặc dù GWAS giúp giới hạn đoạn QTL từ vài mega base xuống còn 50-100 kb thì đây vẫn là kích thước lớn có thể chứa hàng chục gen. Tin tốt là với nhiều tiến bộ trong nghiên cứu chức năng gen như chuyển gen, sàng lọc đột biến từ ngân hàng đột biến T-ADN, bất hoạt ARN hay chỉnh sửa hệ gen (genome editing) hỗ trợ rất nhiều cho việc nghiên cứu chức năng gen.

Bất chấp một số khó khăn, thách thức nêu trên, với thành công bước đầu trong nghiên cứu GWAS trên bộ rễ của Phụng và cộng sự năm 2016 đã mở ra hướng đi mới bên cạnh các phương pháp truyền thống. GWAS không chỉ giúp khai thác hiệu quả nguồn gen phong phú của cây lúa Việt Nam mà còn hứa hẹn cung cấp những chỉ thị phân tử mới phù hợp cho các chương trình chọn tạo giống lúa ở nước ta.

Tài liệu tham khảo

1. H. Begum, J.E. Spindel, A. Lalusin, T. Borromeo, G. Gregorio, J. Hernandez, P. Virk, S.R. Collard, McCouch (2015), “Genome - Wide Association Mapping for Yield and Other Agronomic Traits in an Elite Breeding Population of Tropical Rice (*Oryza sativa*)”, *PLOS ONE*, **10**, pp.1-19.
2. Y. Feng, Q. Lu, R. Zhai, M. Zhang, Q. Xu, Y. Yang, S. Wang, X. Yuan, H. Yu, Y. Wang, et al (2016), “Genome wide association mapping for grain shape traits in indica rice”, *Planta*, pp.1-12.
3. X. Huang, Y. Zhao, X. Wei, C. Li, A. Wang, Q. Zhao, W. Li, et al (2011), “Genome - wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm”, *Nat. Genet.*, **44**, pp.32-39.
4. V. Kumar, A. Singh, S.V.A. Mithra, S.L. Krishnamurthy, et al (2015), “Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*)”, *DNA Res.*, **22**, pp.133-145.
5. S.R. McCouch, M.H. Wright, C.W. Tung, L.G. Maron, McNally, et al (2016), “Open access resources for genome-wide association mapping in rice”, *Nat. Commun.*, **7**, p.10532.
6. N.T.P. Phung, C.D. Mai, P. Mournet, J. Frouin, G. Droc, N.K. Ta, S. Jouannic, L.T. Le, V.N. Do, P. Gantet, et al (2014), “Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes”, *BMC Plant Biol.*, **14**, p.371.
7. N.T.P. Phung, C.D. Mai, G.T. Truong Hoang, et al (2016), “Genome-wide association mapping for root traits in a panel of rice accessions from Vietnam”, *BMC Plant Biol.*, **16**, p.64.
8. L. Si, J. Chen, X. Huang, H. Gong, J. Luo, Q. Hou, T. Zhou, T. Lu, Zhu, et al (2016), “OsSPL13 controls grain size in cultivated rice”, *Nat. Genet.*, **48**, pp.447-456.
9. <http://www.gramene.org/qlt>.
10. <http://irri.org/our-work/research/genetic-diversity/international-rice-genebank>.