

BÀI TỔNG QUAN

CÁC ỨNG DỤNG CỦA KỸ THUẬT LAI HUỖNH QUANG TẠI CHỖ (FISH) TRONG NGHIÊN CỨU HỆ GEN

Hoàng Thị Như Phương^{1,3}, Huỳnh Thị Thu Huệ², Cao Xuân Hiếu^{1,4,✉}

¹Viện Di truyền thực vật và Nghiên cứu cây trồng (IPK), Gatersleben, Cộng hòa liên bang Đức

²Viện Nghiên cứu hệ gen (IGR), Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Trường Đại học Đà Lạt, Việt Nam

⁴Viện Sinh học, Trường Đại học Martin-Luther tại Halle-Wittenberg, Cộng hòa liên bang Đức

✉ Người chịu trách nhiệm liên lạc. E-mail: xuan.cao@biologie.uni-halle.de

Ngày nhận bài: 09.12.2018

Ngày nhận đăng: 20.4.2019

TÓM TẮT

Lai huỳnh quang tại chỗ (FISH) là kỹ thuật xác định vị trí DNA đích (ví dụ trên tiêu bản kỳ giữa nhiễm sắc thể -NST). Bài tổng quan này tập trung vào các hướng ứng dụng của FISH trong nghiên cứu hệ gen, bao gồm việc xác nhận và chỉnh sửa kết quả lắp ráp hệ gen từ công nghệ giải trình tự gen hiện đại (next-generation sequencing, NGS) hay còn được biết đến là các công nghệ giải trình tự thế hệ thứ hai. Các đầu dò đặc hiệu cho từng vùng DNA có thể được tạo ra từ sản phẩm PCR đặc hiệu hoặc tách dòng từ vector khác nhau. FISH hiển thị vùng NST mang đầu dò, từ đó cho phép kiểm tra tính chính xác của kết quả lắp ráp hệ gen. Bài báo mô tả những cải tiến của FISH bao gồm FISH đa màu cùng với quy trình rửa bỏ và lai mới đầu dò trên cùng tiêu bản. Bản đồ di truyền tế bào từ FISH và tích hợp với trình tự NGS giúp cho việc lắp ráp trình tự toàn bộ NST của hệ gen trở nên hoàn chỉnh hơn, đồng thời mở ra các hướng nghiên cứu mới về cấu trúc và tiến hóa hệ gen. Để nghiên cứu mối quan hệ tiến hóa ở các loài chưa được giải mã, người ta có thể sử dụng kỹ thuật sơn màu NST so sánh. Bên cạnh đó, FISH với đoạn dò cho một nhóm trình tự lặp lại nhất định hoặc cho các cấu trúc cơ bản của NST (DNA tâm động, đầu mút NST, DNA mã hóa rRNA ...) có thể dùng để nghiên cứu cấu trúc hệ gen và những yếu tố biệt hóa bộ NST. Bài báo cũng thảo luận những giới hạn và triển vọng ứng dụng tương lai của FISH.

Từ khóa: công nghệ giải trình tự gen hiện đại, di truyền tế bào học, hệ gen học, giải mã hệ gen, kỹ thuật lai huỳnh quang tại chỗ

MỞ ĐẦU

Một bộ gen được giải mã hoàn chỉnh, có chất lượng cao sẽ là nguồn thông tin tham khảo rất quan trọng cho các nghiên cứu sâu thêm về cơ chế sinh học phân tử, đặc biệt đối với các loài không phải là sinh vật mô hình. Nhờ vào các thành tựu của công nghệ giải trình tự gen hiện đại (*next-generation sequencing*, NGS) hay còn được biết đến là công nghệ giải trình tự thế hệ thứ hai mà tốc độ giải mã hệ gen ngày càng nhanh chóng với chi phí thấp và có thể sử dụng trên rất nhiều đối tượng nghiên cứu khác nhau. Tuy nhiên việc lắp ráp một cách chính xác các đoạn giải trình tự để cho ra một hệ gen hoàn chỉnh ở các loài sinh vật nhân chuẩn có hệ gen lớn, phức tạp và nhiều vùng lặp lại hiện vẫn là thách thức cho các nhà hệ gen học. Kỹ thuật lai huỳnh quang tại chỗ

(*fluorescence in situ hybridization*, FISH) là một kỹ thuật có thể góp phần giải quyết được những khó khăn này và giúp xác định mối liên kết giữa các contig, scaffold cũng như phát hiện các sai sót của kết quả lắp ráp trình tự. Ngoài ra, với sự hỗ trợ của các công cụ tin sinh học, FISH có thể mô tả đặc tính của phần trình tự lặp lại và qua đó cung cấp bức tranh toàn cảnh về cách thức tổ chức hệ gene. Trong bài tổng quan này, nhóm tác giả sẽ sơ lược một số hệ thống giải mã gen và đề cập đến nguyên lý, ứng dụng cũng như vai trò của kỹ thuật FISH trong các nghiên cứu giải mã hệ gen.

CÁC CHIẾN LƯỢC GIẢI MÃ HỆ GEN MỚI HOÀN TOÀN (*de novo*)

Việc giải trình tự phân tử DNA được bắt đầu vào những năm 1970 với phương pháp hóa học Maxam-

Gillbert, sau đó là phương pháp enzyme của Sanger. Các thành công trong việc cải tiến phương pháp giải trình tự cũng như ứng dụng các chương trình máy tính đã giúp việc giải trình tự gen nhanh chóng và tiết kiệm hơn. Tuy nhiên, để giải mã toàn bộ hệ gen của 1 sinh vật là vô cùng khó khăn và mất thời gian. Dự án giải mã hệ gen người (3.2 Gbp, 23 cặp nhiễm sắc thể) được khởi động từ năm 1990 với kinh phí dự kiến là 3 tỷ dollar đã được công bố vào 2003 (sớm hơn 2 năm so với dự kiến) tuy vẫn còn một vài khoảng trống chưa hoàn thiện với tổng chi phí là 2,7 tỷ dollar (International Human Genome Sequencing Consortium 2004). Cuộc đua giải trình tự hệ gen người trở nên nóng hơn từ năm 1998 khi có sự tham gia của Celera Genomics – một công ty tư nhân. Công ty này đã sử dụng chiến thuật giải trình tự gen “Whole genome shotgun sequencing” (1, phân tách hệ gen thành các đoạn nhỏ ngẫu nhiên; 2, giải trình tự hàng loạt các đoạn này bằng phương pháp Sanger; 3, dùng thuật toán lắp ráp các đoạn ngắn dựa trên vùng trình tự chồng lắp để cho ra những trình tự liên tục) nhằm rút ngắn thời gian và giảm chi phí, kết quả là họ đã hoàn thành dự án (Istrail *et al.*, 2004) trong vòng 3 năm với kinh phí chỉ 300 triệu dollar (so với 13 năm và 2,7 tỷ dollar của dự án giải mã hệ gen người). Tuy nhiên các kỹ thuật này vẫn tốn nhiều thời gian để hoàn thành và chúng được gọi là kỹ thuật giải trình tự thế hệ thứ nhất “First-generation”. Đối với các loài thực vật, *Arabidopsis thaliana* là hệ gen mô hình (157 Mbp, 5 cặp nhiễm sắc thể) đầu tiên được giải mã hoàn chỉnh vào năm 2000 (Arabidopsis Genome Initiative 2000), sau đó là cây lúa (*Oryza sativa*, 430 Mbp, Yu *et al.*, 2002).

Các hệ thống NGS được giới thiệu trong thập niên vừa qua cho phép phân tích đồng thời hàng vạn trình tự gen một cách nhanh chóng. Ví dụ, bằng hệ thống HiSeq X™ Ten có thể giải mã hệ gen của 45 người trong vòng một ngày với chi phí khoảng 1.000 dollar/ người (Illumina 2016). Sự ra đời của các hệ thống thiết bị giải trình tự gen với tốc độ nhanh và chi phí thấp đã đưa việc giải mã gen trở thành một kỹ thuật thông dụng trong sinh học phân tử. Các kỹ thuật mới của giải mã hệ gen ngày càng được ứng dụng rộng rãi hơn không chỉ trong các nghiên cứu mà còn ở cả các thử nghiệm lâm sàng (ví dụ chẩn đoán nhanh chóng cơ sở di truyền của bệnh, Knoppers *et al.*, 2015). Có 4 phương pháp chính được sử dụng trong các hệ thống NGS đó là: pyrosequencing (ví dụ 454, Roche, nay đã dừng ứng dụng), giải trình tự bằng con đường tổng hợp (sequencing by synthesis, ví dụ Illumina), con đường lai (sequencing by ligation, ví dụ SOLID) và ion bán

dẫn (ion semiconductor sequencing, ví dụ Ion Torrent). Mặc dù có ưu điểm là giải mã nhanh chóng 1 lượng lớn trình tự với chi phí thấp, các phương pháp này đều chỉ tạo ra các trình tự tương đối ngắn (< 700 bp) và gặp khó khăn rất lớn trong việc giải mã mới hoàn toàn (*de novo*) hệ gen lớn (ví dụ > 700 Mbp) và phức tạp (ví dụ nhiều vùng lặp lại). Gần đây, công nghệ giải trình tự đơn phân tử theo thời gian thực (single molecular, real time, SMRT của Pacific Biosciences) cho phép giải mã đoạn trình tự với độ dài có thể lên đến hơn 20,000 bp (với độ dài trung bình khoảng 3,000 bp) nghĩa là gấp 30 đến 200 lần so với các công nghệ NGS khác (Chin *et al.*, 2013). Ngoài ra, công nghệ giải trình tự lỗ nano (nanopore sequencing, MinION) mặc dù đang ở trong giai đoạn phát triển nhưng cũng có nhiều hứa hẹn (Deamer *et al.*, 2016).

Lắp ráp hệ gen *de novo* là thuật toán lắp ráp các trình tự giải mã ngắn thành contig (trình tự liên tục do lắp ráp), các contig thành scaffold (trình tự tạo thành do lắp ráp có thể không liên tục) để cho ra trình tự hệ gen của đối tượng nghiên cứu. Nhiều công cụ khác nhau được dùng trong lắp ráp hệ gen *de novo* như Velvet (Zerbino, Birney 2008), SOAPdenovo (Luo *et al.*, 2012), AbySS (Simpson *et al.*, 2009), MaSuRCA (Zimin *et al.*, 2013) và ALLPATHS-LG (Butler *et al.*, 2008). Tuy nhiên, không một công cụ nào là toàn năng để có thể lắp ráp được hệ gen hoàn chỉnh của tất cả các đối tượng (Gilchrist *et al.*, 2015). Về tổng quan các thuật toán lắp ráp trong NGS có thể xem thêm ở (Miller *et al.*, 2010; Escalona *et al.*, 2016; Goodwin *et al.*, 2016). Quá trình phân tích và lắp ráp hệ gen cung cấp cho chúng ta một số thông tin di truyền quan trọng của đối tượng nghiên cứu như số lượng gen mã hóa protein, cấu trúc hệ gen được giải mã và lắp ráp, vị trí các gen trên nhiễm sắc thể, lịch sử hình thành hệ gen (ví dụ thông qua sự kiện nhân đôi toàn bộ hệ gen, whole-genome duplications WGDs). Giải mã hệ gen hoàn chỉnh của các loài thực vật là những nhiệm vụ đặc biệt khó khăn do kích thước hệ gen lớn (lên đến 150 Gbp ở *Paris japonica*) và thành phần các đoạn lặp lại nhiều và phức tạp. Bảng 1 tóm tắt kết quả giải mã hệ gen của các loài thực vật trong 10 năm gần đây (2006 – 2016).

Qua bảng tóm tắt trong bảng 1 có thể thấy rằng trong khoảng 5 năm trở lại đây các kỹ thuật NGS đã dần thay thế phương pháp giải mã Sanger trong các dự án giải mã hệ gen, dù rằng một lượng nhỏ trình tự có thể vẫn giải mã bằng Sanger để làm cơ sở đối chiếu chất lượng của trình tự NGS. Các kỹ thuật

NGS đều có những ưu và nhược điểm riêng của mình, vì thế trong một số dự án, các nhà nghiên cứu sử dụng kết hợp các kỹ thuật này để giải mã hệ gen nhanh chóng và độ chính xác cao nhất có thể.

Bảng 1. Danh mục một số loài thực vật được giải mã hệ gen.

TT	Đối tượng	Kích thước (Mbp)	Kỹ thuật giải mã*	Năm	Nhóm tác giả
1	<i>Arabidopsis thaliana</i>	125	Sa	2000	(Arabidopsis Genome Initiative 2000)
2	Lúa (<i>Oryza sativa</i>)	430	Sa	2002	(Yu <i>et al.</i> , 2002)
3	Dương (<i>Populus trichocarpa</i>)	485	Sa	2006	(Tuskan <i>et al.</i> , 2006)
4	Nho (<i>Vitis vinifera</i>)	475	Sa	2007	(Jaillon <i>et al.</i> , 2007)
5	Rêu (<i>Physcomitrella patens</i>)	510	Sa	2008	(Rensing <i>et al.</i> , 2008)
6	Đu đủ (<i>Carica papaya</i>)	372	Sa	2008	(Ming <i>et al.</i> , 2008)
7	Dưa leo (<i>Cucumis sativus</i>)	367	Sa, I	2009	(Huang <i>et al.</i> , 2009)
8	Ngô (<i>Zea mays</i>)	2300	Sa	2009	(Schnable <i>et al.</i> , 2009)
9	Đậu nành (<i>Glycine max</i>)	1115	Sa	2010	(Schmutz <i>et al.</i> , 2010)
10	Táo (<i>Malus x domestica</i>)	742	Sa, 4	2010	(Velasco <i>et al.</i> , 2010)
11	Ca cao (<i>Theobroma cacao</i>)	430	Sa, 4, I	2011	(Argout <i>et al.</i> , 2011)
12	Dâu tây (<i>Fragaria vesca</i>)	240	4, S, I	2011	(Shulaev <i>et al.</i> , 2011)
13	Khoai tây (<i>Solanum tuberosum</i>)	844	Sa, 4, I	2011	(Xu <i>et al.</i> , 2011)
14	Cải dầu (<i>Brassica rapa</i>)	485	I	2011	(Wang <i>et al.</i> , 2011)
15	Cây gai dầu (<i>Cannabis sativa</i>)	820	4, I	2011	(van Bakel <i>et al.</i> , 2011)
16	Đậu triều (<i>Cajanus cajan</i>)	833	Sa, I	2012	(Varshney <i>et al.</i> , 2012)
17	Cỏ linh lăng (<i>Medicago truncatula</i>)	454	Sa, 4, I	2011	(Young <i>et al.</i> , 2011)
18	Kê (<i>Setaria italica</i>)	490	I	2012	(Zhang <i>et al.</i> , 2012)
19	Cà chua (<i>Solanum lycopersicum</i>)	900	Sa, 4, S, I	2012	(Tomato Genome Consortium 2012)
20	Dưa (<i>Cucumis melo</i>)	450	Sa, 4, I	2012	(Garcia-Mas <i>et al.</i> , 2012)
21	Cây lanh (<i>Linum usitatissimum</i>)	373	I	2012	(Wang <i>et al.</i> , 2012)
22	Chuối (<i>Musa acuminata malaccensis</i>)	523	Sa, 4, I	2012	(D'Hont <i>et al.</i> , 2012)
23	Bông (<i>Gossypium raimondii</i>)	880	I	2012	(Wang <i>et al.</i> , 2012)
24	Cây neem, xoan Ấn Độ (<i>Azadirachta indica</i>)	364	4, I	2012	(Krishnan <i>et al.</i> , 2012)
25	Lê (<i>Pyrus bretschneideri</i>)	527	I	2013	(Wu <i>et al.</i> , 2013)
26	Dưa hấu (<i>Citrullus lanatus</i>)	425	I	2013	(Guo <i>et al.</i> , 2013)
27	Lúa mì (<i>Triticum aestivum</i>)	17000	4	2012	(Brenchley <i>et al.</i> , 2012)
28	Mơ (<i>Prunus mume</i>)	280	I	2012	(Zhang <i>et al.</i> , 2012)
29	Đậu gà (<i>Cicer arietinum</i>)	738	Sa, I	2013	(Varshney <i>et al.</i> , 2013)
30	2150	4, S, I	2013	(Rahman <i>et al.</i> , 2013)	
31	Trúc (<i>Phyllostachys heterocycla</i>)	2075	I	2013	(Peng <i>et al.</i> , 2013)
32	Lúa cỏ (<i>Oryza brachyantha</i>)	300	I	2013	(Chen <i>et al.</i> , 2013)
33	Đào (<i>Prunus persica</i>)	265	Sa	2013	(Verde <i>et al.</i> , 2013)
34	Lúa mì DD (<i>Aegilops tauschii</i>)	4360	4, I	2013	(Jia <i>et al.</i> , 2013)
35	Lúa mì AA (<i>Triticum urartu</i>)	4940	I	2013	(Ling <i>et al.</i> , 2013)
36	Sen (<i>Nelumbo nucifera</i>)	929	I	2013	(Ming <i>et al.</i> , 2013)
37	Rong (<i>Utricularia gibba</i>)	77	4, I	2013	(Ibarra-Laclette <i>et al.</i> , 2013)
38	<i>Capsella rubella</i>	219	Sa	2013	(Slotte <i>et al.</i> , 2013)
39	<i>Amborella trichopoda</i>	870	4, I, Sa	2013	(Chamala <i>et al.</i> , 2013)
40	Tảo đỏ (<i>Chondrus crispus</i>)	105	S	2013	(Collen <i>et al.</i> , 2013)
42	Kiwi (<i>Actinidia chinensis</i>)	758	I	2013	(Huang <i>et al.</i> , 2013)
43	Củ cải đường (<i>Beta vulgaris</i>)	758	I, 4, S	2014	(Dohm <i>et al.</i> , 2014)
44	Ớt (<i>Capsicum annuum</i>)	3480	I, S	2014	(Qin <i>et al.</i> , 2014)
45	Bèo tấm (<i>Spirodela polyrhiza</i>)	158	4, S	2014	(Wang <i>et al.</i> , 2014)
46	Thông (<i>Pinus taeda</i>)	20150	I	2014	(Wegrzyn <i>et al.</i> , 2014)

47	Lúa (<i>Oryza glaberrima</i>)	316	4, S	2014	(Wang <i>et al.</i> , 2014)
48	Cải dầu (<i>Camelina sativa</i>)	785	I, 4	2014	(Kagale <i>et al.</i> , 2014)
49	Cà tím (<i>Solanum melongena</i>)	1100	I, 4	2014	(Hirakawa <i>et al.</i> , 2014)
50	Táo tàu (<i>Ziziphus jujuba</i>)	444	I	2014	(Liu <i>et al.</i> , 2014)
51	Việt quất (<i>Vaccinium macrocarpon</i>)	470	I	2014	(Polashock <i>et al.</i> , 2014)
52	Cỏ tai hùm (<i>Conyza canadensis</i>)	335	4, I, PB	2014	(Peng <i>et al.</i> , 2014)
53	Bông (<i>Gosypium hirsutum</i> TM-1)	2250	I	2014	(Li <i>et al.</i> , 2014)
54	Cà (<i>Solanum commersonii</i>)	840	I	2015	(Aversano <i>et al.</i> , 2015)
55	Hoa bia (<i>Humulus lupulus</i>)	2570	I	2015	(Natsume <i>et al.</i> , 2015)
56	<i>Arachis duranensis</i> & <i>A. ipaensis</i>	1250 & 1560	I	2016	(Bertioli <i>et al.</i> , 2016)

* Sa: Sanger; 4: 454/Roche; I: Illumina/Solexa, S: SOLiD/ABI, PB: Pac Bio System.

Tuy công nghệ giải trình tự gen đã được cải tiến, năng lực của các công cụ lắp ráp đã được nâng cao rất nhiều nhưng trình tự hệ gen thu được thường bị phân tách thành nhiều mảnh nhỏ. Mức độ hoàn chỉnh của hệ gen giải mã thường được biểu thị và có thể so sánh thông qua tham số L50 (số lượng contig/scaffold tối thiểu để tổng chiều dài của chúng chiếm 50% kích thước hệ gen) và N50 (chiều dài của contig/scaffold nhỏ nhất trong bộ L50). Tuy nhiên, việc xây dựng được trình tự hệ gen hoàn chỉnh ở mức độ toàn vẹn từng nhiễm sắc thể vẫn còn là thách thức lớn phía trước. Gần đây, một số chiến thuật giải trình tự và lắp ráp mới (ví dụ NRGene với DenovoMagic hay Dovetail Genomics với kỹ thuật Chicago (Putnam *et al.*, 2016)) có nhiều triển vọng lớn giúp các nhà khoa học giải mã các hệ gen lớn và phức tạp. Để hỗ trợ lắp ráp hoặc đối chiếu, kiểm chứng kết quả giải mã NGS, các thông tin về bản đồ di truyền liên kết (linkage map), bản đồ di truyền tế bào (cytogenetic map), bản đồ vật lý (physical map), bản đồ quang học (optical maps) thường được sử dụng. Trong đó, kỹ thuật FISH được sử dụng trong các dự án giải mã hệ gen như một công cụ hỗ trợ để tăng tính chính xác của hệ gen vì kỹ thuật này không những có thể xác định được vị trí mà còn cho biết được mối liên kết giữa các đoạn gen với nhau để lập nên bản đồ di truyền tế bào trên từng nhiễm sắc thể (Jiang, Gill 2006).

NGUYÊN LÝ VÀ ỨNG DỤNG CỦA KỸ THUẬT FISH

Nguyên lý và một số biến thể của kỹ thuật FISH

FISH là kỹ thuật trung gian giữa di truyền tế bào và di truyền phân tử. Đây là kỹ thuật vô cùng hữu ích được dùng trong các nghiên cứu về cấu trúc và chức năng của nhiễm sắc thể như vùng dị nhiễm sắc, gen mã hóa cho ribosome RNA, tâm động, telomere và subtelomere. Sự phát triển của kỹ thuật FISH đã

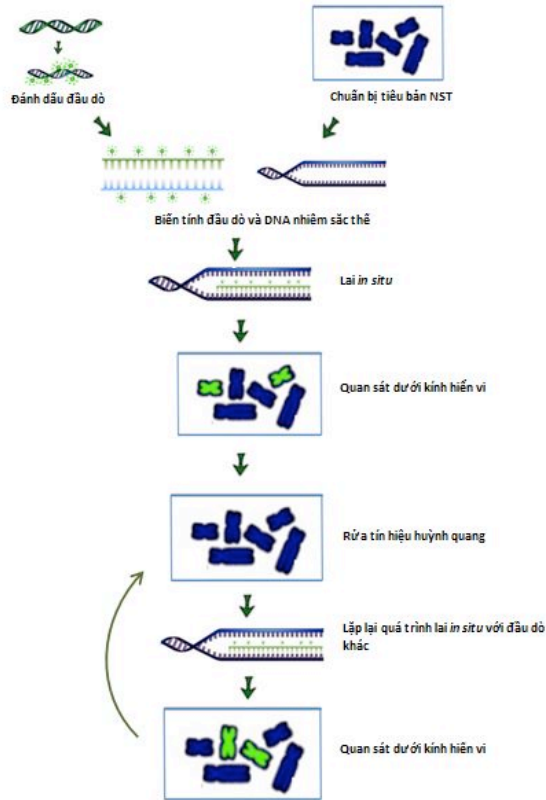
mở ra cơ hội cho các phân tích di truyền học ở nhiều đối tượng khác nhau mà không bị hạn chế bởi hình thái của nhiễm sắc thể (như ở các kỹ thuật nhuộm băng Q, G, R, T và C, Pardue, Gall 1969).

Kỹ thuật FISH cho phép xác định trình tự DNA trên tiêu bản nhiễm sắc thể (thông thường là tiêu bản kỳ giữa nguyên phân) và tiêu bản nhân tế bào kỳ trung gian bằng cách sử dụng các đoạn đầu dò đặc hiệu cho từng nhiễm sắc thể hoặc từng gen. Với việc sử dụng các nucleotide được đánh dấu huỳnh quang trong quá trình chuẩn bị đầu dò đã giúp quan sát và định vị được vị trí của đoạn gen khảo sát trong tế bào, trong nhân và trong từng nhiễm sắc thể riêng biệt (John *et al.*, 1969).

Các bước cơ bản của kỹ thuật FISH được mô tả ở Hình 1 và có thể được tóm tắt như sau:

- i. Đánh dấu đầu dò: đầu dò được đánh dấu trực tiếp hoặc gián tiếp bằng các phương pháp khác nhau (Nick translation, PCR, đánh dấu ngẫu nhiên bằng PCR).
- ii. Chuẩn bị tiêu bản: nhiễm sắc thể kỳ giữa hoặc nhân tế bào kỳ trung gian. Xử lý tiêu bản bằng enzyme để loại bỏ tế bào chất cũng như các protein liên kết với DNA
- iii. Biến tính đầu dò và DNA tiêu bản
- iv. Lai *in situ* trong điều kiện tối ảm (từ 16 – 72h)
- v. Rửa tiêu bản để loại bỏ những liên kết không đặc hiệu
- vi. Dò tìm tín hiệu bằng các chất chỉ thị trung gian (đối với đầu dò được đánh dấu gián tiếp)
- vii. Quan sát tiêu bản dưới kính hiển vi
- viii. Xác định vị trí của đoạn gen hoặc nhiễm sắc thể cần tìm trên tiêu bản
- ix. Tiêu bản có thể được sử dụng nhiều lần với nhiều đầu dò khác nhau bằng cách rửa tín hiệu

huỳnh quang (chỉ hoạt động tốt với đầu dò trực tiếp) và tiếp tục lai với các đầu dò khác để có được thông tin hoàn chỉnh trên cùng 1 tiêu bản.



Hình 1. Các bước cơ bản của kỹ thuật FISH.

Một số cải tiến đã được áp dụng dựa trên nguyên lý cơ bản của kỹ thuật FISH đã được nghiên cứu và ứng dụng trong các phòng thí nghiệm thực vật như Tyr – FISH, 3-D FISH, hay FISH trên các nhiễm sắc thể siêu kéo dài. Tyr - FISH sử dụng kháng thể cộng hợp với peroxidase (peroxidase-conjugated antibody) sau đó sử dụng tyramides (đã được đánh dấu huỳnh quang) là cơ chất để khuếch đại và phát hiện tín hiệu FISH. Bằng phương pháp này, tín hiệu có thể được khuếch đại 10 – 100 lần và phù hợp cho việc xác định các đoạn gen có kích thước nhỏ hơn 1 kb (Raap *et al.*, 1995; van Gijlswijk *et al.*, 1997). Tuy nhiên, việc khuếch đại tín hiệu lại dẫn đến các tín hiệu không đặc hiệu (background) xuất hiện nhiều hơn vì thế phương pháp này cần phải được tối ưu hóa để giảm các tín hiệu nhiễu. Trong khi đó, 3-D FISH sử dụng kính hiển vi quang học cắt lớp (optical-sectioning microscopy): bao phần được cố định trong dung dịch đệm nhằm bảo vệ cấu trúc của

nhiễm sắc thể. Sau đó các tế bào mẹ hạt phấn được tách một cách nhẹ nhàng khỏi dung dịch cố định và được đặt lên lam kính polyacrylamide sạch để nhuộm và quan sát. Các hình kết quả FISH được chụp dưới dạng từng hình 3-D riêng lẻ. Các nhiễm sắc thể có tín hiệu FISH sẽ được xử lý và duỗi thẳng bằng phần mềm máy tính. Bằng kỹ thuật này, cấu trúc của nhiễm sắc thể được cố định đã giúp cho việc dự đoán vị trí của đoạn gen trên nhiễm sắc thể ở trong nhân (Harper, Cande 2000; Koumaris, Bass 2003). Tuy nhiên, phương pháp này phức tạp và chi phí cao hơn nhiều so với việc sử dụng kính hiển vi thông dụng. Năm 2006, Wang và đồng tác giả (2006) đã công bố rằng các nhiễm sắc thể giảm phân (pachytene chromosome) được chuẩn bị bằng phương pháp squash (tiêu bản giọt ép) cũng cho cấu trúc tương tự với phương pháp 3-D và thuận lợi hơn trong việc bản đồ hóa vị trí các single copy (các đoạn gen chỉ có 1 phiên bản duy nhất trong bộ gen). Tuy phương pháp 3-D FISH không phát huy hiệu quả trong việc lập bản đồ phân tử nhưng lại rất hữu ích trong việc khảo sát sự sắp xếp của từng phần DNA trong nhân và trong lai miễn dịch protein. Đối với các thí nghiệm FISH trên các nhiễm sắc thể siêu kéo dài, nhiễm sắc thể có thể có kích thước dài gấp 100 lần so với nhiễm sắc thể ở kỳ giữa nguyên phân. Vì thế, FISH trên tiêu bản này giúp lập được bản đồ phân tử với độ phân giải lên tới 70 kb, tương tự với các nhiễm sắc thể pachytene của giảm phân (Valarik *et al.*, 2004). Vì thế, đây là phương pháp thay thế hiệu quả trong nghiên cứu di truyền trên các đối tượng thực vật không hoặc khó chuẩn bị được tiêu bản nhiễm sắc thể giảm phân.

Ứng dụng của kỹ thuật FISH trong các dự án giải mã hệ gen

Các thuật toán lắp ráp trình tự từ những hệ thống giải mã đoạn ngắn (short read sequencing technology) gặp nhiều khó khăn để có thể xây dựng thông tin hoàn chỉnh của 1 hệ gen hoặc 1 nhiễm sắc thể. Đồng thời kết quả dựa hoàn toàn vào thuật toán cũng có tiềm ẩn nguy cơ sai sót. Vì thế, kỹ thuật FISH được sử dụng như một công cụ hỗ trợ trong các dự án giải mã hệ gen để xác nhận độ chính xác cũng như phát hiện ra những sai sót của quá trình lắp ráp hệ gen. Bên cạnh đó, kỹ thuật này còn đóng vai trò quan trọng trong lập bản đồ di truyền tế bào và nghiên cứu quá trình tiến hóa của nhiễm sắc thể.

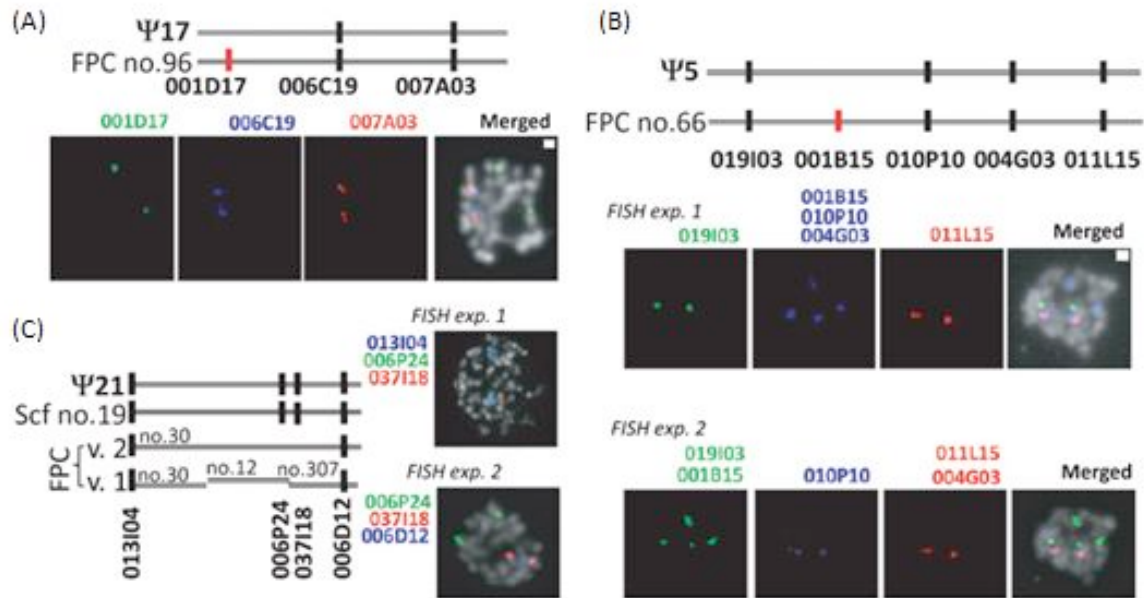
Ứng dụng trong kiểm định kết quả lắp ráp trình tự

Kỹ thuật FISH có thể được dùng để xác định mối quan hệ liên kết giữa các đoạn gen mà nằm rất

xa nhau trên nhiễm sắc thể và/ hoặc bị xen giữa bởi những đoạn dài các trình tự lặp lại. Vì thế, đây được xem như là một kỹ thuật độc lập để đánh giá, chỉnh sửa kết quả tích hợp từ NGS (Cao *et al.*, 2016). Từ 1990, kỹ thuật FISH đã được sử dụng để xác định chiều dài cũng như vị trí sắp xếp của các trình tự quan tâm trên nhiễm sắc thể 11 của người, và để lập bản đồ vật lý với độ phân giải 1 -3 cM (Lichter *et al.*, 1990). Các đoạn DNA được tách dòng trong nhiễm sắc thể nhân tạo của vi khuẩn (Bacterial Artificial Chromosome, BAC) là nguồn vật liệu ưa thích để sử dụng làm đầu dò trong các thí nghiệm FISH (BAC-FISH). BAC FISH đã được sử dụng để kiểm định kết quả lắp ráp trình tự để giải mã hệ gen tiêu chuẩn ở bèo tấm *S. polyrhiza* (Hình 2, Cao *et al.*, 2016) hay ở cây *Amborella trichopoda* (hệ gen cơ sở cho các loài thực vật hạt kín, Chamala *et al.*,

2013). Ngoài ra, các trình tự tách dòng từ cDNA (cDNA clone trên nhiễm sắc thể số 7 của lúa mạch, Karafiatova *et al.*, 2013) hoặc các sản phẩm PCR (sản phẩm PCR kích thước từ 7-15 kbp trên cây ăn thịt *Genlisea nigrocaulis*, Vu *et al.*, 2015) cũng có thể sử dụng như là các đầu dò đánh dấu. Đặc biệt, kết quả nghiên cứu trên NST số 10 ở lúa cho thấy FISH trên sợi DNA là một kỹ thuật hiệu quả để xác nhận bản đồ BAC-contig và đánh giá kích thước của các khoảng trống trình tự lớn giữa các marker (Cheng *et al.*, 2002; Zhou *et al.*, 2007).

Qua một vài ví dụ trên, có thể thấy rằng việc sử dụng kỹ thuật FISH để xác định vị trí, mối liên kết của các đoạn gen trên nhiễm sắc thể cũng như phát hiện ra các sai sót của bản đồ di truyền là rất quan trọng để thiết lập thông tin hoàn chỉnh về hệ gen của đối tượng nghiên cứu.



Hình 2. Kiểm định kết quả lắp ráp trình tự của *S. polyrhiza* bằng kỹ thuật FISH (nguồn từ Cao *et al.*, 2016). Các DNA BAC trên pseudomolecule (Ψ), scaffold (Scf) (của bản đồ BAC end sequences) hoặc fingerprint contigs (FPC) (của bản đồ di truyền) được lai trên tiêu bản nhiễm sắc thể ký giữa của *S. polyrhiza*. (A) Kết quả FISH cho thấy quá trình lắp ráp trình tự đã đúng khi cho rằng vị trí của BAC 001D17 (màu xanh lục) không nằm trên Ψ 17 như giả định của bản đồ fingerprinting (FPC96); (B) Tương tự, BAC 001B15 được xác định là không nằm trên Ψ 5 như kết quả FPC66. Trong thí nghiệm FISH lần thứ nhất, một tập hợp DNA của 3 BAC được trộn lẫn, đánh dấu cùng 1 màu (xanh dương), và lai trên một tiêu bản nhiễm sắc thể của *S. polyrhiza*. Việc có thể rửa bỏ kết quả FISH (lần 1) và tiến hành với các tổ hợp mẫu DNA đánh dấu khác trên cùng 1 tiêu bản đã giúp nhóm tác giả xác định được BAC 001B15 nằm bên ngoài nhóm liên kết; (C) FISH đã giúp xác định một sai sót trong quá trình lắp ráp Ψ 21 (Scf19). Nhóm BAC 013104 và 006P24 nằm trên cặp nhiễm sắc thể khác so với nhóm BAC 037118 và 006D12. Thông tin về phiên bản khác nhau (v. 1 và v. 2) của FPC có thể giúp dự đoán vị trí điểm sai sót giữa BAC 006P24 và 037118 (FPC12 và FPC307 của v. 1).

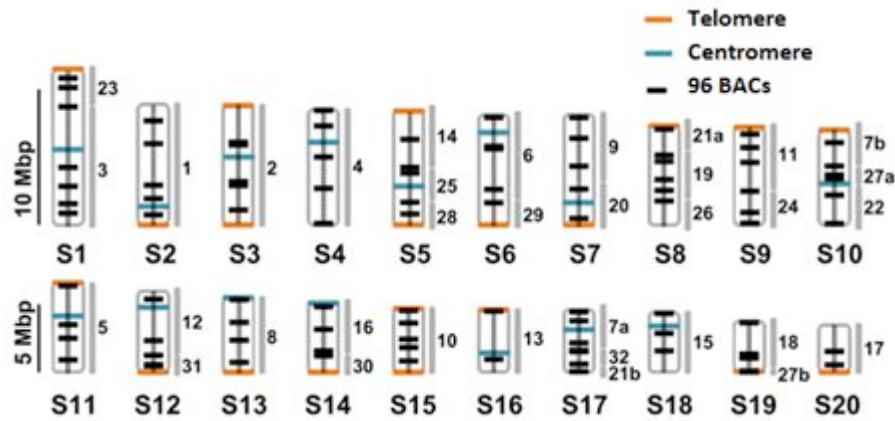
Ứng dụng trong lập bản đồ vật lý / bản đồ di truyền tế bào

Bản đồ hệ gen thể hiện vị trí của các locus trên

nhiễm sắc thể cũng như khoảng cách giữa các gen với nhau. Hiện nay có hai loại bản đồ gen là bản đồ di truyền và bản đồ vật lý. Bản đồ di truyền (hay còn

gọi là bản đồ di truyền liên kết): sự sắp xếp của các gen và gen chỉ thị trên nhiễm sắc thể được tính toán dựa trên tần số hoán vị gen (đơn vị tính là cM). Bản đồ vật lý thể hiện khoảng cách vật lý giữa các gen trên nhiễm sắc thể (đơn vị tính là base pair). Kỹ thuật FISH được sử dụng phổ biến trong các nghiên cứu lập bản đồ vật lý của các đoạn gen quan tâm trên nhiều đối tượng khác nhau như: (i) rDNA (5S và 45S) ở ngô (Li, Arumuganathan 2001), đậu phộng (Seijo *et al.*, 2004), anh đào (Maghuly *et al.*, 2010); (ii) gen kháng bệnh phần trắng ở lúa mì (Wang *et al.*, 2015); (iii) gen Rfm1 ở lúa mạch (Ui *et al.*, 2015); (iv) gen aquaporin ở lúa mì (Forrest, Bhavé 2010). Bên cạnh đó, kỹ thuật FISH còn được sử dụng để lập bản đồ vật lý / bản đồ di truyền tế bào của 50 BACs trên NST số 11 của người (Lichter *et al.*, 1990); 28

BACs trên NST số 2 của cà chua (Koo *et al.*, 2008); 10 BACs trên NST số 1 của *Gossypium herbaceum* (Cui *et al.*, 2015); ngô (Wang, Chen 2005); 23 đầu dò đặc hiệu trên NST số 3, 4 và 7 ở đậu (*Phaseolus vulgaris* L., Pedrosa-Harand *et al.*, 2009); 96 BACs trên 20 cặp nhiễm sắc thể bèo tấm *S. polyrhiza* (Hình 3, Cao *et al.*, 2016). Đặc biệt, vùng gen gần tâm động của các nhiễm sắc thể thường là những nơi đặc biệt khó khăn để có được trình tự giải mã hoàn chỉnh. Bởi vì vùng này thường chứa rất nhiều trình tự lặp lại và không có thông tin trên bản đồ di truyền (do tần số trao đổi chéo rất thấp hoặc gần như không có). Các kết quả từ vùng tâm động của nhiễm sắc thể 3H ở lúa mạch (Aliyeva-Schnorr *et al.*, 2015) cho thấy FISH có thể sử dụng để giải quyết các trường hợp tương tự.



Hình 3. Bản đồ phân bố của 96 BACs trên 20 nhiễm sắc thể của bèo tấm *S. polyrhiza* (nguồn từ Cao *et al.*, 2016). Các nhiễm sắc thể được sắp xếp (đánh số thứ tự) dựa trên chiều dài đoạn trình tự liên tục do lắp ráp. Vị trí của các pseudomolecules trên mỗi NST được thể hiện qua đường kẻ dọc màu xám. Các thanh ngang là vị trí của telomere (TTTAGGG)_n (màu cam); centromere (màu xanh) và BACs (màu đen).

Ứng dụng trong nghiên cứu mô tả tổng quan hệ gen và phân bố trình tự lặp lại

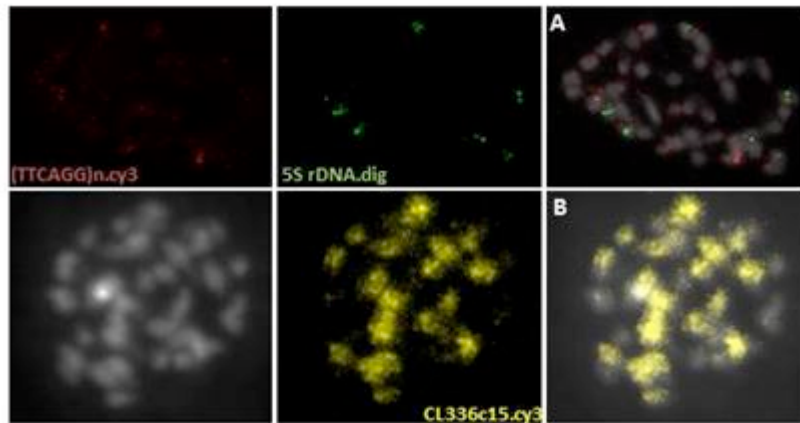
Trình tự DNA lặp lại đóng vai trò quan trọng trong điều hòa hoạt động gen, hoạt động của NST giới tính và quá trình tiến hóa nhiễm sắc thể. Các trình tự lặp lại này chiếm một phần lớn trong toàn bộ hệ gen sinh vật nhân chuẩn và có độ đa dạng cao vì thế, sự có mặt của các trình tự lặp lại này là nguyên nhân gây trở ngại cho các nghiên cứu, phân tích chi tiết hệ gen. Với kết quả thu được từ NGS có thể phân tích và dự đoán tỷ phần của các trình tự lặp lại trong hệ gen nghiên cứu.

Để phân loại các đoạn lặp trong một cơ sở dữ liệu gồm hàng triệu đoạn đọc được tạo ra từ kết quả giải trình tự thế hệ mới thì một công cụ có thể sử

dụng là RepeatExplorer. Hàng triệu đoạn đọc ngắn được phân tích trong một lần chạy, thuật toán nhóm được sử dụng bằng cách tìm kiếm và đánh giá sự giống nhau về trình tự của các đoạn đọc để đưa về nhóm giống nhau và hiển thị sự phân tích cơ sở dữ liệu để xác định các nhóm thuộc loại trình tự lặp nào. Số lượng đoạn đọc trong một lần tính toán của RepeatExplorer thường vào khoảng 5% của toàn bộ hệ gen của một thực vật bậc cao. Các loại đoạn lặp có thể gồm các DNA satellite, LTR retrotransposon hay rDNA. Đồng thời, RepeatExplorer sẽ cho biết thông tin về số nhóm lặp, số đoạn đọc thuộc 1 nhóm và tỉ lệ mỗi loại đoạn lặp. Số liệu về tỉ lệ của các loại đoạn lặp trong phân tích của RepeatExplorer cũng phản ánh một tỉ lệ tương tự trong toàn hệ gen (Novak *et al.*, 2010; Novak *et al.*,

2013). Các DNA lặp với motif lặp từ 2bp cho đến trên 10kb là những thành phần chính của genome thực vật, chúng có thể chiếm đến 85% DNA nhân (Schnable *et al.*, 2009). Trong nghiên cứu của Heitkam và đồng tác giả (2015), khi phân tích số liệu trình tự genome của cây trà Nhật Bản *Camellia japonica* bằng Repeat Explorer kết quả cho thấy hàm lượng đoạn lặp chiếm 73%, trong đó tỉ lệ mỗi loại lặp như sau LTR/gypsy 47.3%; Satellite 17%; đoạn lặp chưa định danh 29.4%; LTR/copia 4.9%; rDNA 1.1%; LINE 0.2% và DNA transposon 0.1%. Từ kết quả phân tích *in silico* như vậy, có thể tiến hành làm các nghiên cứu tiếp theo để xác định vị trí và sự phân bố của các loại đoạn lặp trên NST thông qua kỹ thuật FISH.

Do công nghệ NGS có thể giải mã toàn bộ hệ gen cung cấp thông tin về các loại đoạn lặp và tỉ lệ các đoạn lặp nên khi kết hợp với FISH thì sẽ cho biết được tổ chức và kiểu phân bố của các đoạn lặp trên NST của thực vật. Bằng cách đó, FISH có thể giúp xác định trình tự đoạn lặp đóng vai trò sinh học cơ bản trong hệ gen/ tế bào như DNA tâm động (centromere DNA repeat), telomere của các nhiễm sắc thể (Hình 4). Ví dụ, các trình tự centromere và telomere khác nhau đã phát sinh trong quá trình tiến hóa của chi thực vật *Genlisea* (Tran *et al.*, 2015) được phát hiện nhờ kỹ thuật FISH kết hợp nhuộm miễn dịch (immunostaining). Ngoài ra, FISH cũng có thể được sử dụng để so sánh bản đồ vật lý của các đoạn lặp trong những loài có liên quan (Begum *et al.*, 2013).



Hình 4. Đặc điểm di truyền tế bào của *Genlisea subglabra* ($2n = 40$) (nguồn từ Tran *et al.*, 2015). (A) Các cấu trúc lặp lại có chức năng quan trọng trong tế bào như DNA telomere repeat (TTCAGG) n hay rDNA loci có thể quan sát thông qua kỹ thuật FISH. (B) Phân bố của 1 loại đoạn lặp trên nhiễm sắc thể có thể cung cấp các thông tin về nguồn gốc tiến hóa của loài. Ví dụ loại đoạn lặp CL336c15 ở loài *G. subglabra* được tìm thấy ở 10 trong tổng số 20 cặp nhiễm sắc thể cùng có cho giả thuyết về nguồn gốc hệ gen dị tứ bội (allotetraploid) của loài.

Ứng dụng trong nghiên cứu quá trình tiến hóa nhiễm sắc thể

Bên cạnh việc tìm ra vị trí cũng như mối liên kết giữa các đoạn gen để làm cơ sở cho việc lắp ráp hệ gen hoàn chỉnh, kỹ thuật FISH còn là một kỹ thuật rất quan trọng để nghiên cứu quá trình tiến hóa nhiễm sắc thể giữa các loài thực vật có quan hệ họ hàng với nhau mà không cần phải giải mã hệ gen. Trong các nghiên cứu này, những đột biến trong cấu trúc (chuyển đoạn, đảo đoạn) và số lượng (đa bội, dị bội, dung hợp NST) NST giữa các loài khảo sát sẽ được thể hiện qua vị trí và số lượng tín hiệu FISH của đầu dò. Dựa trên số lượng nhiễm sắc thể cùng với thông tin thu được từ kỹ thuật FISH, các nhà

nghiên cứu có thể mô hình hóa con đường tiến hóa nhiễm sắc thể từ loài nguyên thủy đến loài mới phát sinh và sự xuất hiện của hiện tượng nhân đôi toàn bộ hệ gen (whole genome duplication, WGD).

Quá trình tiến hóa NST của nhiều loài đã được nghiên cứu dựa trên kỹ thuật FISH và kết quả cho thấy hiện tượng chuyển đoạn (translocation), đảo đoạn (inversion), dung hợp (fusion) giữa các NST và bất hoạt chức năng tâm động chịu trách nhiệm cho quá trình tiến hóa giảm số lượng NST từ $n = 8$ (*Arabidopsis lyrata*) thành $n = 5$ (*Arabidopsis thaliana*) (Lysak *et al.*, 2006); từ 8 NST nguyên thủy ACK (Ancestral Crucifer Karyotype) thành 7 NST ở họ Cải Brassicaceae (Mandakova, Lysak

2008); hiện tượng phân chia tại tâm động (centric fission) và đảo đoạn (pericentric inversion) đã giải thích cho cơ chế hình thành nhiễm sắc thể Het' và Del ở các dòng *B. polyantha* trinh sản ($2n = 15$) (Mandakova *et al.*, 2015).

Kỹ thuật nhuộm so sánh NST (comparative chromosome painting, CCP) giữa các loài nghiên cứu cho thấy các sự kiện WGD kết hợp với đột biến cấu trúc NST là nguyên nhân làm thay đổi kích thước hệ gen và số lượng NST như giữa *G. nigrocaulis* ($2n = 40$; 86 Mbp) và *G. pygmaea* ($2n = 80$; 179 Mbp) (Vu *et al.*, 2015), từ 8 NST nguyên thủy ACK đến $2n = 2x = 18$ ở *B. laevigata* subsp. *varia*. (Geiser *et al.*, 2016). Sự tương đồng của một số gen quan tâm giữa các loài họ hàng cũng được thể hiện qua kết quả FISH như giữa *Brachypodium distachyon* và *Hordeum vulgare* (Ma *et al.*, 2010); giữa nhiễm sắc thể 2BL ở lúa mì với nhiễm sắc thể số 4 của lúa (Lee *et al.*, 2010).

ƯU, NHƯỢC ĐIỂM VÀ TRIỂN VỌNG CỦA KỸ THUẬT FISH

Ưu và nhược điểm của kỹ thuật FISH

Ưu điểm của kỹ thuật FISH bao gồm: (i) bản đồ BAC-FISH là cơ sở để giải thích cho sự thay đổi (đột biến) vị trí của các BAC cũng như khoảng cách giữa các contig; (ii) kỹ thuật FISH giúp xác định sự phân bố và ước tính được độ lớn các trình tự lặp lại trên nhiễm sắc thể mà không cần phải giải trình tự gen; (iii) BAC-FISH giúp định vị được vị trí các đoạn gen của vùng dị nhiễm sắc gần tâm động- khu vực ít xảy ra hiện tượng trao đổi chéo; (iv) kỹ thuật này còn được dùng để làm sáng tỏ các đột biến xảy ra do hiện tượng trao đổi chéo trong giảm phân và sự khác biệt của vùng gen quan tâm giữa các loài khác nhau mà không cần phải có thông tin cơ sở từ di truyền phân tử và (v) trạng thái hoạt động của các vùng nhiễm sắc được thể hiện từ thông tin của BAC-FISH cùng với phân tích miễn dịch nhiễm sắc thể.

Bên cạnh những ưu điểm trên thì kỹ thuật FISH vẫn còn một vài nhược điểm như: (i) đòi hỏi nhiều thời gian và kỹ thuật khác nhau (di truyền học, kính hiển vi, kỹ thuật phân tử, phân tích và giải thích hình ảnh); (ii) số lượng phòng thí nghiệm sử dụng kỹ thuật này rất hạn chế; (iii) không có quy trình chuẩn cho việc thiết kế, đánh dấu đầu dò và lai mà cần phải được tối ưu tùy vào đối tượng nghiên cứu và (iv) chi phí cho kính hiển vi và chất huỳnh quang đánh dấu cao.

Triển vọng của kỹ thuật FISH

Nhờ vào những đóng góp của mình trong nghiên cứu di truyền nhiễm sắc thể mà kỹ thuật FISH ngày càng được sử dụng rộng rãi trong các nghiên cứu về phân loại thực vật và chọn tạo giống cây trồng. Các công ty chọn tạo giống ngày càng quan tâm nhiều hơn đến kỹ thuật này vì bản đồ vị trí các BAC sẽ cung cấp thông tin về tiềm năng xảy ra hiện tượng trao đổi chéo giữa các đoạn gen quan tâm. Từ đó, các chương trình lai tạo giống cây trồng có thể được nghiên cứu và thực hiện dựa trên sự phân bố của các đoạn gen mang tính trạng mong muốn như chịu hạn, chịu mặn, kháng bệnh (Szinay *et al.*, 2010). Một vài chương trình lai tạo giống thành công giữa các đối tượng thực vật trong cùng 1 loài hoặc trong cùng 1 chi có thể được kể đến như: khoai tây và cà chua (Jacobsen *et al.*, 1992), lúa mì x lúa mạch đen (Sanchez-Moran *et al.*, 1999), củ cải đường x *Beta corolliflora* (Desel *et al.*, 2002).

Không chỉ giới hạn trong các nghiên cứu trên thực vật, kỹ thuật FISH ngày càng được sử dụng rộng rãi trong các nghiên cứu trên động vật, đặc biệt là trong lĩnh vực y học. Vì thế, hiện nay FISH được xem như là một kỹ thuật tất yếu trong các chẩn đoán y sinh như chẩn đoán viêm phổi do *Pneumocystis carinii* ở ngựa con và lợn (Jensen *et al.*, 2001); đột biến số lượng và cấu trúc của NST ở tế bào trứng hay noãn bào (Weise *et al.*, 2009); các hiện tượng chuyển đoạn ở bò (*Bos taurus*), lợn (*Sus scrofa domestica*), ngựa (*Equus caballus*) (Rubes *et al.*, 2009); phát hiện hiện tượng chuyển vị BCR/ABL1, khuếch đại HER2 và tái sắp xếp của ALK là yếu tố rất quan trọng trong xây dựng liệu pháp chữa trị cho các bệnh nhân ung thư máu, vú, phổi, tiền liệt tuyến (Hu *et al.*, 2014).

Như vậy, kỹ thuật FISH sẽ ngày càng được ứng dụng rộng rãi trong nhiều lĩnh vực nghiên cứu khác nhau. Sự kết hợp giữa kỹ thuật FISH với sự phát triển nhanh chóng, liên tục của công nghệ giải trình tự gen và công cụ lắp ráp hệ gen *de novo* sẽ giúp giảm chi phí và tăng tốc độ giải mã hệ gen cũng như độ chính xác của hệ gen sau khi được hoàn thiện. Nguồn thông tin này là vô cùng hữu ích và có thể được sử dụng cho rất nhiều mục đích nghiên cứu khác nhau về sau.

TÀI LIỆU THAM KHẢO

Aliyeva-Schnorr L, Beier S, Karafiatova M, Schmutz T, Scholz U, Dolezel J, Stein N, Houben A (2015) Cytogenetic mapping with centromeric bacterial artificial

- chromosomes contigs shows that this recombination-poor region comprises more than half of barley chromosome 3H. *Plant J* 84(2): 385-394.
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814): 796-815.
- Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Bérard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahiri M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C. (2011) The genome of *Theobroma cacao*. *Nat Genet* 43(2): 101-108.
- Aversano R, Contaldi F, Ercolano MR, Grosso V, Iorizzo M, Tatino F, Xumerle L, Dal Molin A, Avanzato C, Ferrarini A, Delledonne M, Sanseverino W, Cigliano RA, Capella-Gutierrez S, Gabaldón T, Frusciantè L, Bradeen JM, Carpato D. (2015) The *Solanum commersonii* Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. *Plant Cell* 27(4): 954-968.
- Begum R, Zakrzewski F, Menzel G, Weber B, Alam SS, Schmidt T (2013) Comparative molecular cytogenetic analyses of a major tandemly repeated DNA family and retrotransposon sequences in cultivated jute *Corchorus species* (Malvaceae). *Ann Bot* 112(1): 123-134.
- Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araujo ACG, Kozik A, Kim KD, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimaraes PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SCM, Xun X, Jackson SA, Michelmore R, Ozias-Akins P (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet* 48(4): 438-446.
- Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426): 705-710.
- Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res* 18(5): 810-820.
- Cao HX, Vu GT, Wang W, Appenroth KJ, Messing J, Schubert I (2016) The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. *New Phytol* 209(1): 354-363.
- Chamala S, Chanderbali AS, Der JP, Lan T, Walts B, Albert VA, dePamphilis CW, Leebens-Mack J, Rounsley S, Schuster SC, Wing RA, Xiao N, Moore R, Soltis PS, Soltis DE, Barbazuk WB. (2013) Assembly and validation of the genome of the nonmodel basal angiosperm *Amborella*. *Science* 342(6165): 1516-1517.
- Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Luis Goicoechea J, Liang C, Chen C, Zhang W, Sun S, Liao Y, Zhang X, Yang L, Song C, Wang M, Shi J, Liu G, Liu J, Zhou H, Zhou W, Yu Q, An N, Chen Y, Cai Q, Wang B, Liu B, Min J, Huang Y, Wu H, Li Z, Zhang Y, Yin Y, Song W, Jiang J, Jackson SA, Wing RA, Wang J, Chen M. (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun* 4: 1595.
- Cheng Z, Buell CR, Wing RA, Jiang J (2002) Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Res* 10(5): 379-387.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10(6): 563-569.
- Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, Barbeyron T, Michel G, Noel B, Valentin K, Elias M, Artiguenave F, Arun A, Aury JM, Barbosa-Neto JF, Bothwell JH, Bouget FY, Brillet L, Cabello-Hurtado F, Capella-Gutiérrez S, Charrier B, Cladière L, Cock JM, Coelho SM, Colleoni C, Czjzek M, Da Silva C, Delage L, Denoeud F, Deschamps P, Dittami SM, Gabaldón T, Gachon CM, Groisillier A, Hervé C, Jabbari K, Katinka M, Kloareg B, Kowalczyk N, Labadie K, Leblanc C, Lopez PJ, McLachlan DH, Meslet-Cladière L, Moustafa A, Nehr Z, Nyvall Collén P, Panaud O, Partensky F, Poulain J, Rensing SA, Rousvoal S, Samson G, Symeonidi A, Weissenbach J, Zambounis A, Wincker P, Boyen C. (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *PNAS* 110(13): 5247-5252.
- Cui X, Liu F, Liu Y, Zhou Z, Zhao Y, Wang C, Wang X, Cai X, Wang Y, Meng F, Peng R, Wang K (2015) Construction of cytogenetic map of *Gossypium herbaceum* chromosome 1 and its integration with genetic maps.

Molecular cytogenetics 8(1): 2.

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, Da Silva C, Jabbari K, Cardi C, Poulain J, Souquet M, Labadie K, Jourda C, Lengellé J, Rodier-Goud M, Alberti A, Bernard M, Correa M, Ayyampalayam S, Mckain MR, Leebens-Mack J, Burgess D, Freeling M, Mbéguié-A-Mbéguié D, Chabannes M, Wicker T, Panaud O, Barbosa J, Hribova E, Heslop-Harrison P, Habas R, Rivallan R, Francois P, Poirion C, Kilian A, Burthia D, Jenny C, Bakry F, Brown S, Guignon V, Kema G, Dita M, Waalwijk C, Joseph S, Dievert A, Jaillon O, Leclercq J, Argout X, Lyons E, Almeida A, Jeridi M, Dolezel J, Roux N, Risterucci AM, Weissenbach J, Ruiz M, Glaszmann JC, Quétier F, Yahiaoui N, Wincker P (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488(7410): 213-217.

Deamer D, Akeson M, Branton D (2016) Three decades of nanopore sequencing. *Nat. Biotechnol.* 34(5): 518-524.

Desel C, Jansen R, Dedong G, Schmidt T (2002) Painting of parental chromatin in Beta hybrids by multi-colour fluorescent in situ hybridization. *Ann Bot* 89(2): 171-181.

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sørensen TR, Stracke R, Reinhardt R, Goemann A, Kraft T, Schulz B, Stadler PF, Schmidt T, Gabaldón T, Lehrach H, Weisshaar B, Himmelbauer H. (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505(7484): 546-549.

Escalona M, Rocha S, Posada D (2016) A comparison of tools for the simulation of genomic next-generation sequencing data. *Nat Rev Genet* 17(8): 459-469.

Forrest KL, Bhave M (2010) Physical mapping of wheat aquaporin genes. *Theor Appl Genet* 120(4): 863-873.

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Picó B, Gabaldón T, Roma G, Guigó R, Casacuberta JM, Arús P, Puigdomènech P. (2012) The genome of melon (*Cucumis melo* L.). *PNAS* 109(29): 11872-11877.

Geiser C, Mandakova T, Arrigo N, Lysak MA, Parisod C (2016) Repeated Whole-Genome Duplication, Karyotype Reshuffling, and Biased Retention of Stress-Responding Genes in Buckler Mustard. *Plant Cell* 28(1): 17-27.

Gilchrist CA, Turner SD, Riley MF, Petri WA, Jr., Hewlett EL (2015) Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev* 28(3): 541-563.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17(6): 333-351.

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R, Wu M, Levi A, Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z, Xu Y. (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet* 45(1): 51-58.

Harper LC, Cande WZ (2000) Mapping a new frontier; development of integrated cytogenetic maps in plants. *Funct. Integr. Genomics* 1(2): 89-98.

Heitkam T, Petrasch S, Zakrzewski F, Kogler A, Wenke T, Wanke S, Schmidt T (2015) Next-generation sequencing reveals differentially amplified tandem repeats as a major genome component of Northern Europe's oldest *Camellia japonica*. *Chromosome Res* 23(4): 791-806.

Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S, Ohyama A, Yamaguchi H, Sato S, Isobe S, Tabata S, Fukuoka H. (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. *DNA Res* 21(6): 649-660.

Hu L, Ru K, Zhang L, Huang Y, Zhu X, Liu H, Zetterberg A, Cheng T, Miao W, (2014) Fluorescence in situ hybridization (FISH): an increasingly demanded tool for biomarker research and personalized medicine. *Biomarker Research* 2(1): 3.

Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, Niu X, Zhang X, Meng M, Yu J, Liu J, Han Y, Shi W, Zhang D, Cao S, Wei Z, Cui Y, Xia Y, Zeng H, Bao K, Lin L, Min Y, Zhang H, Miao M, Tang X, Zhu Y, Sui Y, Li G, Sun H, Yue J, Sun J, Liu F, Zhou L, Lei L, Zheng X, Liu M, Huang L, Song J, Xu C, Li J, Ye K, Zhong S, Lu BR, He G, Xiao F, Wang HL, Zheng H, Fei Z, Liu Y. (2013) Draft genome of the kiwifruit *Actinidia chinensis*. *Nat Commun* 4: 2640.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma

- L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41(12): 1275-1281.
- Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juárez MJ, Simpson J, Fernández-Cortés A, Arteaga-Vázquez M, Góngora-Castillo E, Acevedo-Hernández G, Schuster SC, Himmelbauer H, Minoche AE, Xu S, Lynch M, Oropeza-Aburto A, Cervantes-Pérez SA, de Jesús Ortega-Estrada M, Cervantes-Luevano JI, Michael TP, Mockler T, Bryant D, Herrera-Estrella A, Albert VA, Herrera-Estrella L. (2013) Architecture and evolution of a minute plant genome. *Nature* 498(7452): 94-98.
- Illumina. (2016) An introduction to Next-generation sequencing technology. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
- International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431(7011): 931-945.
- Istrail S, Sutton GG, Florea L, Halpern AL, Mobarry CM, Lippert R, Walenz B, Shatkay H, Dew I, Miller JR, Flanigan MJ, Edwards NJ, Bolanos R, Fasulo D, Halldórsson BV, Hannenhalli S, Turner R, Yooseph S, Lu F, Nusskern DR, Shue BC, Zheng XH, Zhong F, Delcher AL, Huson DH, Kravitz SA, Mouchard L, Reinert K, Remington KA, Clark AG, Waterman MS, Eichler EE, Adams MD, Hunkapiller MW, Myers EW, Venter JC. (2004) Whole-genome shotgun assembly and comparison of human genome assemblies. *PNAS* 101(7): 1916-1921.
- Jacobsen E, Reinhout P, Bergervoet JE, de Looft J, Abidin PE, Huigen DJ, Ramanna MS (1992) Isolation and characterization of potato-tomato somatic hybrids using an amylose-free potato mutant as parental genotype. *Theor. Appl Genet* 85(2-3): 159-164.
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463-467.
- Jensen TK, Boye M, Bille-Hansen V (2001) Application of fluorescent in situ hybridization for specific diagnosis of *Pneumocystis carinii* pneumonia in foals and pigs. *Veterinary pathology* 38(3): 269-274.
- Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KF, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, Yang H, Liu X, He Z, Mao L, Wang J. (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496(7443): 91-95.
- Jiang J, Gill BS (2006) Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. *Genome* 49(9): 1057-1068.
- John HA, Birnstiel ML, Jones KW (1969) RNA-DNA hybrids at the cytological level. *Nature* 223(5206): 582-587.
- Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, Higgins EE, Huebert T, Sharpe AG, Parkin IA. (2014) The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat Commun* 5: 3706.
- Karafiatova M, Bartos J, Kopecky D, Ma L, Sato K, Houben A, Stein N, Dolezel J (2013) Mapping nonrecombining regions in barley using multicolor FISH. *Chromosome Res* 21(8): 739-751.
- Knoppers BM, Zawati MH, Senecal K (2015) Return of genetic testing results in the era of whole-genome sequencing. *Nat Rev Genet* 16(9): 553-559.
- Koo DH, Jo SH, Bang JW, Park HM, Lee S, Choi D (2008) Integration of cytogenetic and genetic linkage maps unveils the physical architecture of tomato chromosome 2. *Genetics* 179(3): 1211-1220.
- Koumbaris GL, Bass HW (2003) A new single-locus cytogenetic mapping system for maize (*Zea mays* L.): overcoming FISH detection limits with marker-selected sorghum (*S. propinquum* L.) BAC clones. *Plant J* 35(5): 647-659.
- Krishnan NM, Pattnaik S, Jain P, Gaur P, Choudhary R, Vaidyanathan S, Deepak S, Hariharan AK, Krishna PB, Nair J, Varghese L, Valivarthi NK, Dhas K, Ramaswamy K, Panda B (2012) A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics* 13: 464.
- Lee TG, Lee YJ, Kim DY, Seo YW (2010) Comparative physical mapping between wheat chromosome arm 2BL and rice chromosome 4. *Genetica* 138(11-12): 1277-1296.

- Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, Chen W, Liang X, Shang H, Liu W, Shi C, Xiao G, Gou C, Ye W, Xu X, Zhang X, Wei H, Li Z, Zhang G, Wang J, Liu K, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S. (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet* 46(6): 567-572.
- Li L, Arumuganathan K (2001) Physical mapping of 45S and 5S rDNA on maize metaphase and sorted chromosomes by FISH. *Hereditas* 134(2): 141-145.
- Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC (1990) High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 247(4938): 64-69.
- Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J. (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496(7443): 87-90.
- Liu MJ, Zhao J, Cai QL, Liu GC, Wang JR, Zhao ZH, Liu P, Dai L, Yan G, Wang WJ, Li XS, Chen Y, Sun YD, Liu ZG, Lin MJ, Xiao J, Chen YY, Li XF, Wu B, Ma Y, Jian JB, Yang W, Yuan Z, Sun XC, Wei YL, Yu LL, Zhang C, Liao SG, He RJ, Guang XM, Wang Z, Zhang YY, Luo LH. (2014) The complex jujube genome provides insights into fruit tree biology. *Nat Commun* 5: 5315.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1(1): 18.
- Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *PNAS* 103(13): 5224-5229.
- Ma L, Vu GT, Schubert V, Watanabe K, Stein N, Houben A, Schubert I (2010) Synteny between *Brachypodium distachyon* and *Hordeum vulgare* as revealed by FISH. *Chromosome Res* 18(7): 841-850.
- Maghuly F, Schmoellerl B, Tensch EM, Laimer M (2010) Genome size, karyotyping and FISH physical mapping of 45S and 5S genes in two cherry rootstocks: *Prunus subhirtella* and *Prunus incisa xserrula*. *J Biotechnol* 149(1-2): 88-94.
- Mandakova T, Lysak MA (2008) Chromosomal phylogeny and karyotype evolution in $x=7$ crucifer species (Brassicaceae). *Plant Cell* 20(10): 2559-2570.
- Mandakova T, Schranz ME, Sharbel TF, de Jong H, Lysak MA (2015) Karyotype evolution in apomictic *Boechera* and the origin of the aberrant chromosomes. *Plant J* 82(5): 785-793.
- Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. *Genomics* 95(6): 315-327.
- Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452(7190): 991-996.
- Ming R, VanBuren R, Liu Y, Yang M, Han Y, Li LT, Zhang Q, Kim MJ, Schatz MC, Campbell M, Li J, Bowers JE, Tang H, Lyons E, Ferguson AA, Narzisi G, Nelson DR, Blaby-Haas CE, Gschwend AR, Jiao Y, Der JP, Zeng F, Han J, Min XJ, Hudson KA, Singh R, Grennan AK, Karpowicz SJ, Watling JR, Ito K, Robinson SA, Hudson ME, Yu Q, Mockler TC, Carroll A, Zheng Y, Sunkar R, Jia R, Chen N, Arro J, Wai CM, Wafula E, Spence A, Han Y, Xu L, Zhang J, Peery R, Haus MJ, Xiong W, Walsh JA, Wu J, Wang ML, Zhu YJ, Paull RE, Britt AB, Du C, Downie SR, Schuler MA, Michael TP, Long SP, Ort DR, Schopf JW, Gang DR, Jiang N, Yandell M, dePamphilis CW, Merchant SS, Paterson AH, Buchanan BB, Li S, Shen-Miller J. (2013) Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol* 14(5): R41.
- Natsume S, Takagi H, Shiraiishi A, Murata J, Toyonaga H, Patzak J, Takagi M, Yaegashi H, Uemura A, Mitsuoka C, Yoshida K, Krofta K, Satake H, Terauchi R, Ono E. (2015) The Draft Genome of Hop (*Humulus lupulus*), an Essence for Brewing. *Plant Cell Physiol* 56(3): 428-441.
- Novak P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11: 378.
- Novak P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-

- wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29(6): 792-793.
- Pardue ML, Gall JG (1969) Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *PNAS* 64(2): 600-604.
- Pedrosa-Harand A, Kami J, Gepts P, Geffroy V, Schweizer D (2009) Cytogenetic mapping of common bean chromosomes reveals a less compartmentalized small-genome plant species. *Chromosome Res* 17(3): 405-417.
- Peng Y, Lai Z, Lane T, Nageswara-Rao M, Okada M, Jasieniuk M, O'Geen H, Kim RW, Sammons RD, Rieseberg LH, Stewart CN Jr. (2014) De novo genome assembly of the economically important weed horseweed using integrated data from multiple sequencing platforms. *Plant Physiol* 166(3): 1241-1254.
- Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, Lu H, Hu T, Yao N, Liu K, Li Y, Fan D, Guo L, Li W, Lu Y, Weng Q, Zhou C, Zhang L, Huang T, Zhao Y, Zhu C, Liu X, Yang X, Wang T, Miao K, Zhuang C, Cao X, Tang W, Liu G, Liu Y, Chen J, Liu Z, Yuan L, Liu Z, Huang X, Lu T, Fei B, Ning B, Han B, Jiang Z (2013) The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat. Genet.* 45(4): 456-461, 461e451-452.
- Polashock J, Zelzion E, Fajardo D, Zalapa J, Georgi L, Bhattacharya D, Vorsa N (2014) The American cranberry: first insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biol* 14: 165.
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* 26(3): 342-350.
- Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y, Wu Z, Mao L, Wu H, Ling-Hu C, Zhou H, Lin H, González-Morales S, Trejo-Saavedra DL, Tian H, Tang X, Zhao M, Huang Z, Zhou A, Yao X, Cui J, Li W, Chen Z, Feng Y, Niu Y, Bi S, Yang X, Li W, Cai H, Luo X, Montes-Hernández S, Leyva-González MA, Xiong Z, He X, Bai L, Tan S, Tang X, Liu D, Liu J, Zhang S, Chen M, Zhang L, Zhang L, Zhang Y, Liao W, Zhang Y, Wang M, Lv X, Wen B, Liu H, Luan H, Zhang Y, Yang S, Wang X, Xu J, Li X, Li S, Wang J, Palloix A, Bosland PW, Li Y, Krogh A, Rivera-Bustamante RF, Herrera-Estrella L, Yin Y, Yu J, Hu K, Zhang Z (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *PNAS* 111(14): 5135-5140.
- Raap AK, van de Corput MPC, Vervenne RAW, van Gijlswijk RP, Tanke HJ, Wiegant J (1995) Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum Mol Genet* 4(4): 529-534.
- Rahman AY, Usharraj AO, Misra BB, Thottathil GP, Jayasekaran K, Feng Y, Hou S, Ong SY, Ng FL, Lee LS, Tan HS, Sakaff MK, Teh BS, Khoo BF, Badai SS, Aziz NA, Yuryev A, Knudsen B, Dionne-Laporte A, Mchunu NP, Yu Q, Langston BJ, Freitas TA, Young AG, Chen R, Wang L, Najimudin N, Saito JA, Alam M. (2013) Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics* 14: 75.
- Rensing SA1, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin-I T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319(5859): 64-69.
- Rubes J, Pinton A, Bonnet-Garnier A, Fillon V, Musilova P, Michalova K, Kubickova S, Ducos A, Yerle M. (2009) Fluorescence in situ hybridization applied to domestic animal cytogenetics. *Cytogenet. Genome Res* 126(1-2): 34-48.
- Sanchez-Moran E, Benavente E, Orellana J (1999) Simultaneous identification of A, B, D and R genomes by genomic in situ hybridization in wheat-rye derivatives. *Heredity (Edinb)* 83 (Pt 3): 249-252.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278): 178-183.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson

- SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddelloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956): 1112-1115.
- Seijo JG, Lavia GI, Fernandez A, Krapovickas A, Ducasse D, Moscone EA (2004) Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). *Am J Bot* 91(9): 1294-1303.
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJ, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troglio M, Viola R, Ashman TL, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW Jr, Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffels A, Salama DY, Carter J, Lopez Girona E, Zdepksi A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE, Folta KM. (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43(2): 109-116.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res* 19(6): 1117-1123.
- Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo YL, Steige K, Platts AE, Escobar JS, Newman LK, Wang W, Mandáková T, Vello E, Smith LM, Henz SR, Steffen J, Takuno S, Brandvain Y, Coop G, Andolfatto P, Hu TT, Blanchette M, Clark RM, Quesneville H, Nordborg M, Gaut BS, Lysak MA, Jenkins J, Grimwood J, Chapman J, Prochnik S, Shu S, Rokhsar D, Schmutz J, Weigel D, Wright SI. (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet* 45(7): 831-835.
- Szinay D, Bai Y, Visser R, de Jong H (2010) FISH applications for genomics and plant breeding strategies in tomato and other solanaceous crops. *Cytogenet. Genome Res* 129(1-3): 199-210.
- Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400): 635-641.
- Tran TD, Cao HX, Jovtchev G, Neumann P, Novák P, Fojtová M, Vu GT, Macas J, Fajkus J, Schubert L, Fuchs J. (2015) Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*. *The Plant J* 84(6): 1087-1099.
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroove S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehrling J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793): 1596-1604.
- Ui H, Sameri M, Pourkheirandish M, Chang MC, Shimada H, Stein N, Komatsuda T, Handa H (2015) High-resolution genetic mapping and physical map construction for the fertility restorer Rfm1 locus in barley. *Theor. Appl. Genet* 128(2): 283-290.

- Valarik M, Bartos J, Kovarova P, Kubalaková M, de Jong JH, Dolezel J (2004) High-resolution FISH on super-stretched flow-sorted plant chromosomes. *Plant J* 37(6): 940-950.
- van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* 12(10): R102.
- van Gijlswijk RP, Zijlmans HJ, Wiegant J, Bobrow MN, Erickson TJ, Adler KE, Tanke HJ, Raap AK (1997) Fluorochrome-labeled tyramides: use in immunocytochemistry and fluorescence in situ hybridization. *J Histochem Cytochem* 45(3): 375-382.
- Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang S-P, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol* 30(1): 83-89.
- Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, Thudi M, Gowda CL, Singh NP, Lichtenzweig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR. (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol* 31(3): 240-246.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troglio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niaz F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchiotti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R. (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet* 42(10): 833-839.
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, Zuccolo A, Rossini L, Jenkins J, Vendramin E, Meisel LA, Decroocq V, Sosinski B, Prochnik S, Mitros T, Policriti A, Cipriani G, Dondini L, Ficklin S, Goodstein DM, Xuan P, Del Fabbro C, Aramini V, Copetti D, Gonzalez S, Horner DS, Falchi R, Lucas S, Mica E, Maldonado J, Lazzari B, Bielenberg D, Pirona R, Miculan M, Barakat A, Testolin R, Stella A, Tartarini S, Tonutti P, Arús P, Orellana A, Wells C, Main D, Vizzotto G, Silva H, Salamini F, Schmutz J, Morgante M, Rokhsar DS. (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 45(5): 487-494.
- Vu GTH, Schmutzer T, Bull F, Cao HX, Fuchs J, Tran DT, Jovtchev G, Pistrick K, Stein N, Pecinka A, Neumann P, Novak P, Macas J, Dear PH, Blattner FR, Scholz U, Schubert I (2015) Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus. *The Plant Genome* 8 DOI: 10.3835/plantgenome2015.04.0021.
- Wang CJ, Chen CC (2005) Cytogenetic mapping in maize. *Cytogenet. Genome Res.* 109(1-3): 63-69.
- Wang CJ, Harper L, Cande WZ (2006) High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *Plant Cell* 18(3): 529-544.
- Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S. (2012) The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* 44(10): 1098-1103.
- Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, Zuccolo A, Song X, Kudrna D, Ammiraju JS, Cossu RM, Maldonado C, Chen J, Lee S, Sisneros N, de Baynast K, Golser W, Wissotski M, Kim W, Sanchez P, Ndjiondjop MN, Sanni K, Long M, Carney J, Panaud O, Wicker T, Machado CA, Chen M, Mayer KF, Rounsley S, Wing RA. (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet* 46(9): 982-988.
- Wang W, Haberer G, Gundlach H, Gläber C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, Byrant DW, Mockler TC, Appenroth KJ, Grimwood J, Jenkins J, Chow J, Choi C, Adam C, Cao XH, Fuchs J, Schubert I, Rokhsar D, Schmutz J, Michael TP, Mayer KF, Messing J. (2014) The *Spirodela polyrhiza* genome reveals insights into its neotenus reduction fast growth and aquatic lifestyle. *Nat Commun* 5: 3311.
- Wang X1, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua

- W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z; (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43(10): 1035-1039.
- Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldine A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK, Wang J, Deyholos MK. (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J* 72(3): 461-473
- Wang Z, Li H, Zhang D, Guo L, Chen J, Chen Y, Wu Q, Xie J, Zhang Y, Sun Q, Dvorak J, Luo M, Liu Z(2015) Genetic and physical mapping of powdery mildew resistance gene MIHLT in Chinese wheat landrace Hulutou. *Theor Appl Genet* 128(2): 365-373.
- Wegrzyn JL, Liechty JD, Stevens KA, Wu LS, Loopstra CA, Vasquez-Gross HA, Dougherty WM, Lin BY, Zieve JJ, Martínez-García PJ, Holt C, Yandell M, Zimin AV, Yorke JA, Crepeau MW, Puiu D, Salzberg SL, Dejong PJ, Mockaitis K, Main D, Langley CH, Neale DB. (2014) Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196(3): 891-909.
- Weise A, Mrasek K, Ewers E, Mkrtychyan H, Kosyakova N, Liehr T (2009) Diagnostic applications of fluorescence in situ hybridization. *Expert opinion on medical diagnostics* 3(4): 453-460.
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, Chen NJ, Nishio T, Xu X, Cong L, Qi K, Huang X, Wang Y, Zhao X, Wu J, Deng C, Gou C, Zhou W, Yin H, Qin G, Sha Y, Tao Y, Chen H, Yang Y, Song Y, Zhan D, Wang J, Li L, Dai M, Gu C, Wang Y, Shi D, Wang X, Zhang H, Zeng L, Zheng D, Wang C, Chen M, Wang G, Xie L, Sovero V, Sha S, Huang W, Zhang S, Zhang M, Sun J, Xu L, Li Y, Liu X, Li Q, Shen J, Wang J, Paull RE, Bennetzen JL, Wang J, Zhang S. (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res* 23(2): 396-408.
- Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, Orjeda G, Guzman F, Torres M, Lozano R, Ponce O, Martinez D, De la Cruz G, Chakrabarti SK, Patil VU, Skryabin KG, Kuznetsov BB, Ravin NV, Kolganova TV, Beletsky AV, Mardanov AV, Di Genova A, Bolser DM, Martin DM, Li G, Yang Y, Kuang H, Hu Q, Xiong X, Bishop GJ, Sagredo B, Mejía N, Zagorski W, Gromadka R, Gawor J, Szczesny P, Huang S, Zhang Z, Liang C, He J, Li Y, He Y, Xu J, Zhang Y, Xie B, Du Y, Qu D, Bonierballe M, Ghislain M, Herrera Mdel R, Giuliano G, Pietrella M, Perrotta G, Facella P, O'Brien K, Feingold SE, Barreiro LE, Massa GA, Diambra L, Whitty BR, Vaillancourt B, Lin H, Massa AN, Geoffroy M, Lundback S, DellaPenna D, Buell CR, Sharma SK, Marshall DF, Waugh R, Bryan GJ, Destefanis M, Nagy I, Milbourne D, Thomson SJ, Fiers M, Jacobs JM, Nielsen KL, Sønderkær M, Iovene M, Torres GA, Jiang J, Veilleux RE, Bachem CW, de Boer J, Borm T, Kloosterman B, van Eck H, Datema E, Hekkert Bt, Goverse A, van Ham RC, Visser RG. (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355): 189-195.
- Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou S, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KA, Tang H, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Bergès H, Bidwell S, Bisseling T, Choise N, Couloux A, Denny R, Deshpande S, Dai X, Doyle JJ, Duzé AM, Farmer AD, Foutreau S, Franken C, Gibelin C, Gish J, Goldstein S, González AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong DH, Jing Y, Jöcker A, Kenton SM, Kim DJ, Klee K, Lai H, Lang C, Lin S, Macmill SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun JH, Najjar FZ, Nicholson C, Noirot C, O'Bleness M, Paule CR, Poulain J, Prion F, Qin B, Qu C, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi R, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang BB, Wang K, Wang M, Wang X, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing Y, Yang L, Yao Z, Ying F, Zhai J, Zhou L, Zuber A, Dénarié J, Dixon RA, May GD, Schwartz DC, Rogers J, Quétier F, Town CD, Roe BA. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480(7378): 520-524.
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He

- S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296(5565): 79-92.
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5): 821-829.
- Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, Tao Y, Bian C, Han C, Xia Q, Peng X, Cao R, Yang X, Zhan D, Hu J, Zhang Y, Li H, Li H, Li N, Wang J, Wang C, Wang R, Guo T, Cai Y, Liu C, Xiang H, Shi Q, Huang P, Chen Q, Li Y, Wang J, Zhao Z, Wang J (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 30(6): 549-554.
- Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G, Xing Z, Han C, Pan H, Zhong X, Shi W, Liang X, Du D, Sun F, Xu Z, Hao R, Lv T, Lv Y, Zheng Z, Sun M, Luo L, Cai M, Gao Y, Wang J, Yin Y, Xu X, Cheng T, Wang J (2012) The genome of *Prunus mume* *Nat Commun* 3: 1318.
- Zhou S, Bechner MC, Place M, Churas CP, Pape L, Leong SA, Runnheim R, Forrest DK, Goldstein S, Livny M, Schwartz DC (2007) Validation of rice genome sequence by optical mapping. *BMC Genomics* 8: 278.
- Zimin AV, Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. *Bioinformatics* 29(21): 2669-2677.

APPLICATIONS OF FLUORESCENCE *IN SITU* HYBRIDIZATION (FISH) IN GENOME RESEARCH

Hoang Thi Nhu Phuong^{1,3}, Huynh Thi Thu Hue², Cao Xuan Hieu^{1,4}

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

²Institute of Genome Research (IGR), Vietnam Academy of Science and Technology (VAST), Vietnam

³Dalat University, Vietnam

⁴Institute of Biology, Martin-Luther University of Halle-Wittenberg, Germany

SUMMARY

Fluorescence *in situ* hybridization (FISH) technique enables the direct detection of DNA sequences in intact cellular materials (e.g. individual chromosomes in metaphase spreads). This review article focuses on the applications of FISH in genome research, including validation and correction of the genome assembly from the next-generation sequencing (NGS) projects. DNA probes for specific DNA fragments of the assembly can be obtained from PCR amplicon or cloned products using different vector systems. Localization of these probes on their respective chromosomal regions can be visualized by FISH, providing useful information to cross-check the assembly data. Furthermore, the recent refinements in the FISH technology including using smart pooling scheme of differently colored DNA probes, together with consecutive FISH experiments (stripping and reprobing of the same slide) are described. These advances in multicolor FISH can provide crucial linkage information on association of linkage groups and assembly scaffolds, resulting in so-called cytogenetic maps. Integration of the cytogenetic maps and assembly sequences assists to resolve the chromosome-level genome assembly and to reveal new insights in genome architecture and genome evolution. Especially, comparative chromosome painting with pooled DNA probes from one reference species can be used to investigate ancestral relationships (chromosome homeology and rearrangements) among other not-yet-sequenced species. In addition, FISH using DNA probes for certain specific classes of repetitive DNA elements as well as for basic chromosome structures (e.g. centromere or telomere DNA repeats, ribosomal DNA loci) can be used to study the genome organization and karyotype differentiation. We also discussed about limitations and future perspectives of the FISH technology.

Keywords: cytogenetics, fluorescence *in situ* hybridization, genomics, next-generation sequencing, whole genome sequencing