Application of a potato UHD genetic linkage map for BAC landing and contig initiation in a region of linkage group V

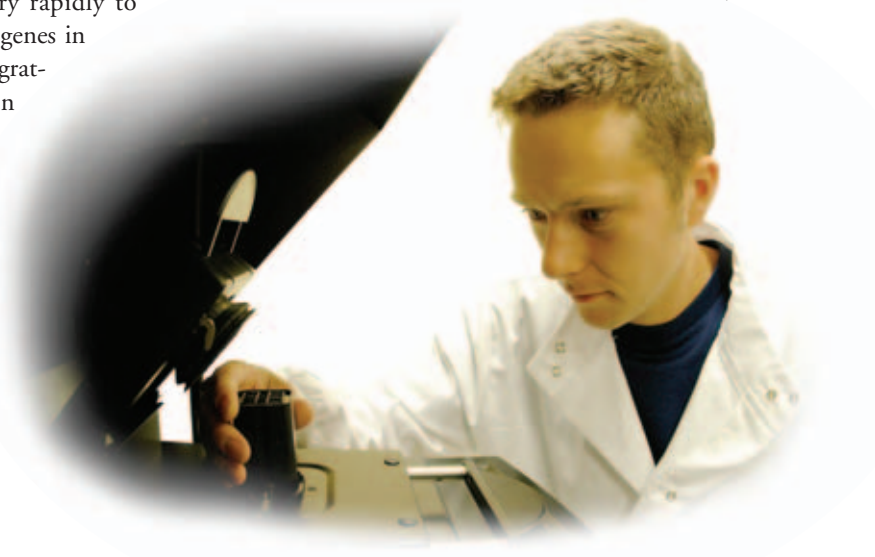
G. Bryan, D. Milbourne, E. Isidore, J. McNicoll, I. Tierney, A. Purvis, S. Williamson, L. Ramsay, K. McLean & R. Waugh

Construction of local or global physical maps is essential for many plant genome research applications, such as map-based gene isolation and genome sequencing. Physical maps are assemblages of overlapping genomic DNA clones, or 'contigs' which can be constructed by a host of different approaches. Existing physical maps have been based (at least in part) on sets of genetically mapped STS (sequence-tagged site) markers, which serve both to assemble contiguous BAC clones, and to anchor these contigs onto the genetic map[1]. Unfortunately, the numbers of characterised STS markers required to make it an effective approach (generally tens of thousands for complex eukaryotes) currently makes it prohibitively expensive for all but model genomes. More typically, physical maps are constructed by random clone fingerprinting, which requires considerable levels of redundancy and sophisticated analytical software. Whole-genome physical mapping studies are costly, and even more challenging in the case of complex heterozygous genomes like potato.

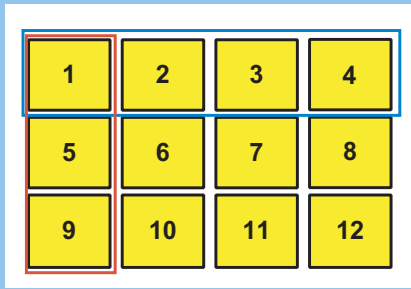
At SCRI, we have set ourselves a less daunting, and more targeted goal – to generate physical maps for key regions of the genome containing genes of prime agronomic and biological importance. For example, the majority of potato disease resistance genes (and several QTL) have been found in 3 or 4 'clusters', on linkage groups IV, V, XI, and XII. Therefore, targeted physical mapping of relatively few genetically defined intervals could yield the resources necessary rapidly to clone the majority of disease resistance genes in potato. We have opted to build an integrated physical and genetic map of a region of potato LG V, defined by RFLP markers GP21 and GP179, that contains a plethora of genes and QTL of agronomic importance. For example, the late blight race specific resistance gene R1, recently isolated, resides in this region[2]. This work is being performed through two EU-funded projects (FAIR-5-PL97-3565 and APOPHYS), the first of which led to the construction of an ultra-high density (UHD) genetic map of the potato genome comprising ~10,000 AFLP markers (http://www.dpw.wau.nl/uhd/). As part of the UHD project, we developed strategies, combining the use of large insert bacterial artificial chromosome (BAC) libraries of the mapping parents, to increase the utility of the UHD map as a versatile resource for potato genetics and genome analysis. The LG V physical mapping work relies heavily on the UHD map, and is aimed at demonstrating its utility in the production of 'seeded' physical maps of defined chromosomal regions of the potato genome.

BAC library construction As a first step towards this goal, we constructed a BAC library of the male parent (RH) of the diploid mapping population used to construct the UHD map. This library consists of 35,712 clones arrayed into ninety-three 384 well microtitre plates. An analysis of approximately 200 random BAC clones showed the RH BAC library has an average insert size of 102kb, with insert sizes range from 30 to 180-kb. This suggests that the library is equivalent to approximately 4-fold coverage of the haploid potato genome, assuming a genome size of 850Mb.

BAC pooling In order to be able to screen the library using PCR, we developed a simple six-dimensional pooling strategy allowing the entire BAC library to be screened in just



Six-dimensional pooling strategy used for screening potato BAC library comprising ~37,000 clones. Library divided into 8 stacks of 12 384-well plates prior to pooling.

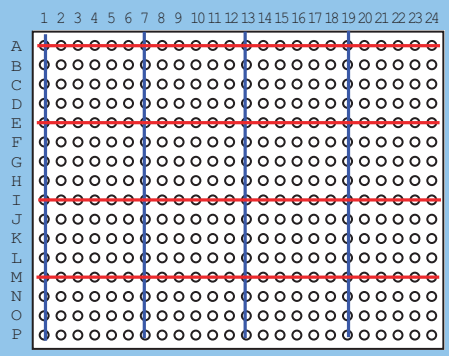


Dimension 1
Pool plates 1,2,3,4 = Pool#1
Pool plates 5,6,7,8 = Pool#2
Pool plates 9,10,11,12 = Pool#3

Dimension 2
Pool plates 1,5,9 = Pool#4
Pool plates 2,6,10 = Pool#5
Pool plates 3,7,11 = Pool#6
Pool plates 4,8,12 = Pool#7

Dimension 3
Pool rows A-D = Pool#8
Pool rows E-H = Pool#9
Pool rows I-L = Pool#10
Pool rows M-P = Pool#11

Dimension 4
Pool columns 1-6 = Pool#12
Pool columns 7-12 = Pool#13
Pool columns 13-18 = Pool#14
Pool columns 19-24 = Pool#15



Dimension 5
Pool rows A,E,I,M = Pool#16
Pool rows B,F,J,N = Pool#17
Pool rows C,G,K,O = Pool#18
Pool rows D,H,L,P = Pool#19

Dimension 6
Pool columns 1,7,13,19 = Pool#20
Pool columns 2,8,14,20 = Pool#21
Pool columns 3,9,15,21 = Pool#22
Pool columns 4,10,16,22 = Pool#23
Pool columns 5,11,17,23 = Pool#24
Pool columns 6,12,18,24 = Pool#25

Figure 1 Pooling strategy used to generate 6 dimensional pools of RH potato BAC library, comprising ~35,000 clones.

200 PCR assays (see Figure 1). The strategy relies on sub-dividing the library into sections representing approximately 0.5x genome equivalents, so that the likelihood of 'hitting' a section more than once with any single copy probe is quite low. Pooling in six-dimensions was then applied to each half-genome stack, allowing the complete address of any single copy sequence in the section to be obtained in 25 PCRs. A single hit in a section yields six co-ordinates, one corresponding to each dimension, which allow the complete address of the relevant clone to be determined. Multiple hits in a section will result in more than six co-ordinates, which requires some degree of 'deconvolution' to identify the individual BACs. The pooling strategy used here attempts to minimise this problem by reducing the probability that multiple hits will be detected in a section. For a section of the RH BAC library containing 12 plates, or 4096 individual clones, this probability is approximately 0.56. Thus, the probability of obtaining two hits in a single section can be approximated to 0.31 (0.56^2), assuming that the two hits are independent events. Furthermore, when considering using segregating AFLPs to screen a library, it is important to take into account that the necessarily heterozygous state of these markers (they must be heterozygous in one or both parents to map in an F_1 population) effectively halves the genome equivalency of the library. Thus a 4x genome equivalent library effectively becomes a 2x genome equivalent library when screening for sequences representing only 1 of the 2 possible alleles at a locus, which gives an even lower probability of obtaining two hits in a section.

Screening the BAC pools The BAC pools were used in conjunction with information from the UHD map for the targeted isolation of BAC clones from the GP21-GP179 interval. This was initiated by mapping GP21 and GP179 as CAPs markers on the UHD map. This allowed the identification of 14 AFLP markers in the interval, of which all but two originated from different primer combinations. A total of 12 AFLP primer combinations, corresponding to 13 AFLPs were deployed in the 6D BAC pools and 8 individual BAC clones were successfully identified. Four of the markers screened on the library yielded no hits, thus 70% (9/13) of the markers screened in the BAC library successfully identified at least one BAC. The RH library was also screened using RFLP markers known to map to this interval (e.g. GP21, GP179, SPUD237, SUT2), a process which identified an additional 10 hits in the library.

Chromosome walking using BAC end sequences The ends of all BACs identified from the library have been sequenced using the T7 and SP6 sites flanking the cloning site in the BAC vector. Resulting sequences were used to search BLAST databases, and for PCR primer design. Primers were tested on genomic DNA, and if a single PCR product of the expected size was amplified, they were tested on the PCR pools. For example, the primers designed to the T7 derived end sequence of RH BAC 56A14 (identified using the primers for SPUD237) were used to screen the pools. Screening the library with this primer identified a total of 4 BACs. One of these was 56A14, the BAC the end sequence was derived from, and another was BAC 13I15, also previously identified by the SPUD237 primers. However, the other 2 BACs identified (48O18 and 64N22) were previously unidentified. This demonstrates that contig extension using the pooled library is feasible, and can be used, in principle to complete the physical map of the region. Problems will occur when BAC end sequences correspond to repetitive elements within the potato genome but these can usually be identified. Interestingly, the end sequences obtained show several homologies to various plant genes, confirming as expected, that this interval is relatively 'gene-dense'.

Towards a physical map of the GP21-GP179 interval of linkage group V Our goal is to illustrate the 'seeding' of an interval on the UHD map of potato LG V with BACs. The results of the first phase of these experiments are summarised graphically in Figure 2, which displays the 'BAC-seeded' genetic map of the GP21-GP179 region. A total of 20 BACs were identified using markers located in the GP21-GP179 interval, and where possible, the relative positions of these BACs are illustrated on the map, by reference to the marker used to identify them. Interestingly, of the three markers that identified BAC 8E5, two co-segregate, but are separated from the third marker by a single recombination event. Examination of the graphical genotype file for the interval supports the validity of this recombination event. This suggests that, in this portion of the interval at least, the genetic to physical distance ratio is relatively high, a favourable situation for physical mapping. Partial sequencing of 8E5 suggests that it contains alleles or homologues of the R1 gene.

Apart from the genetic order of the markers that were used to identify them, very little is currently known about the relationship of the 'seeded' BACs to each other. Groups of BACs identified by the same primers have a very high probability of overlap. Restriction digest fingerprint analysis has confirmed extensive overlap for BACs isolated using the same target sequence (e.g. SPUD237, GP179). An important aspect of the map is that, due to the heterozygous nature of the AFLP markers on the map, BACs identified by these markers are also 'phase' specific. A consequence of this is that primers designed to end-sequences for these BACs will potentially identify not only overlapping clones, but also BACs from both chromatids. Thus, contigs based on these BACs have the potential for rapid growth in terms of both expan-

Use of the UHD map of potato to initiate contig-building across the GP21-GP179 interval of chromosome V. The genetic map of RH chromosome V region in the GP21/GP179 interval is shown, with the two bars representing the two RH alleles. Numbers in the middle are recombination events. Genetic markers on each side (in black) are the markers mapped on the UHD map in the same phase. Individual BAC clones, identified by screening the BAC pools, are displayed in red next to the marker used to identify them.

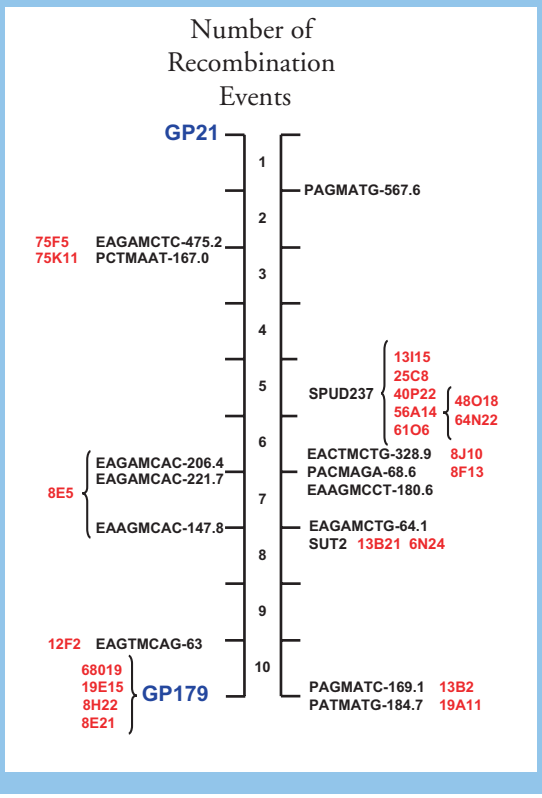


Figure 2 Seeded physical and genetic map of GP21 – GP179 interval of potato LG V, following initial round of screening with AFLPs and STS markers (in E:\POTATO_LGV\map.ppt but needs redrawing for Ann Report)

sion and depth. Moreover, this will facilitate the joining of BACs or contigs based on clones identified by AFLPs alone.

Conclusions This article outlines the construction of a 3-4 fold genome equivalent BAC library of the diploid clone RH, and its successful use to devise a six-dimensional BAC pooling strategy. We have successfully combined the use of AFLP and STS analysis on BAC pools, to identify specific markers present in the interval of interest on the RH genetic map. End sequencing of some of the BACs from the interval has allowed the demonstration of chromosome walking in the region to identify overlapping BAC clones. The identification of 20 BAC clones spanning 11 recombination events from the GP21/GP179 interval is a first step in the construction of a physical map of the region. Its completion relies on use of the end and internal sequences from BACs already identified as 'initiation' points for further BAC isolation from the library. Once a sufficiently large number of BACs has been identified, they can be fingerprinted to identify clone overlaps, which will group the BACs into larger contigs. Yet more chromosome 'walking' can then be performed to fill in the potential remaining gaps in the physical map. This map will be used to identify a minimum 'tiling path' across the region (i.e. the least redundant set of BACs spanning it), which will be used for determining the gene content of this important genetic interval. A further objective of this work is to isolate a Potato Cyst Nematode resistance gene (*Gpa5*) conferring resistance to *G. pallida*, present in the GP21-GP179 interval in the diploid clone JP, for which we also have a BAC library. This will demonstrate further the utility of the markers in the UHD genetic map and will rely on further development of locus-specific markers from the RH BAC clones present in the region.

Acknowledgements

We gratefully acknowledge the financial support of the Scottish Executive Environment & Rural Affairs Department (SEERAD) and the European Union.

Refer ces

[1]Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., Gund, C., McCombie, W.R., McDonald, K., McPherson, J., Mudd, N., Parnell, L., Schein, J., Seim, R., Shelby, P., Waterston R. & Wilson, R. (1999). *Nature Genetics* **22,** 265-270.

[2]Ballvora. A., Ercolano, M.R., Weib, J., Meksem, K., Bormann, C.A., Oberhagemann, P., Salamini, F. & Gebhardt, C. (2002). *The Plant Journal* **30,** 361-371.