Theor Appl Genet (2011) 122:1375–1383
DOI 10.1007/s00122-011-1538-3

ORIGINAL PAPER

Combining genetical genomics and bulked segregant analysis-based differential expression: an approach to gene localization

Xinwei Chen · Peter E. Hedley · Jenny Morris · Hui Liu · Rients E. Niks · Robbie Waugh

Received: 9 November 2010 / Accepted: 6 January 2011 / Published online: 26 January 2011

Abstract Positional gene isolation in unsequenced species generally requires either a reference genome sequence or an inference of gene content based on conservation of synteny with a genomic model. In the large unsequenced genomes of the Triticeae cereals the latter, i.e. conservation of synteny with the rice and *Brachypodium* genomes, provides a powerful proxy for establishing local gene content and order. However, efficient exploitation of conservation of synteny requires 'homology bridges' between the model genome and the target region that contains a gene of interest. As effective homology bridges are generally the sequences of genetically mapped genes, increasing the density of these genes around a target locus is an important step in the process. We used bulked segregant analysis (BSA) of transcript abundance data to identify genes located in a specific region of the barley genome. The approach is valuable because only a relatively small proportion of barley genes are currently placed on a genetic map. We analyzed eQTL datasets from the reference Steptoe × Morex doubled haploid population and showed a strong association between differential gene expression and *cis*-regulation, with 83% of differentially expressed genes co-locating with their eQTL. We then performed BSA by assembling allele-specific pools based on the genotypes of individuals at the partial resistance QTL *Rphq11*. BSA identified a total of 411 genes as differentially expressed, including *HvPHGPx*, a gene previously identified as a promising candidate for *Rphq11*. The genetic location of 276 of these genes could be determined from both eQTL datasets and conservation of synteny, and 254 (92%) of these were located on the target chromosome. We conclude that the identification of differential expression by BSA constitutes a novel method to identify genes located in specific regions of interest. The datasets obtained from such studies provide a robust set of candidate genes for the analysis and serve as valuable resources for targeted marker development and comparative mapping with other grass species.

Communicated by T. Luebberstedt.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1538-3) contains supplementary material, which is available to authorized users.

X. Chen (✉) · P. E. Hedley · J. Morris · H. Liu · R. Waugh (✉)
Genetics Programme,
Scottish Crop Research Institute, Invergowrie,
Dundee DD2 5DA, Scotland, UK
e-mail: xinwei.chen@scri.ac.uk

P. E. Hedley
e-mail: pete.hedley@scri.ac.uk

J. Morris
e-mail: jenny.morris@scri.ac.uk

R. Waugh
e-mail: robbie.waugh@scri.ac.uk

R. E. Niks
Laboratory of Plant Breeding,
Graduate School for Experimental Plant Sciences,
Wageningen University, Wageningen, The Netherlands
e-mail: rients.niks@wur.nl

Introduction

Barley is an important diploid crop species and a good genomic model for staple cereals. A great deal of progress has been achieved over the past two decades within the barley community in establishing well-characterized genetic stocks and mutant collections (http://www.untamo.net/cgi-bin/ace/searches/basic) (Caldwell et al. 2004; Lundqvist et al. 1996), bacterial artificial chromosome

(BAC) libraries (Isidore et al. 2005; Yu et al. 2000) and numerous genetic linkage maps (Varshney et al. 2004). Recently, thanks to second generation sequencing (2GS) technologies, the volume of expressed sequence tag (EST) information has dramatically increased, supplementing over 500,000 Sanger-based ESTs released in Genbank (Close et al. 2009). The 2GS supports and extends the partial sequence information available for ~50,000 tentative unigenes (Sreenivasulu et al. 2008). Based on these coding sequences, high throughput gene-based genotyping technologies have been developed and many genes have been mapped on the barley genome (Varshney et al. 2006; Rostoks et al. 2005; Stein et al. 2007; Wenzl et al. 2004) supplementing historical DNA markers based on RFLP (Kleinhofs and Graner 2001) and AFLP (Hori et al. 2003). Currently over 5,000 non-redundant barley genes have been genetically mapped.

However, due to the lack of a complete reference genome sequence and low gene density in barley, fine mapping and map-based gene isolation often relies on exploiting conserved synteny with model grass species (e.g. rice and *Brachypodium*). However, only 50% of barley genes that are highly homologous to their rice orthologues appear to be collinearly arranged in both genomes (Stein et al. 2007), a limitation that may not be significantly improved with the more closely related grass model *Brachypodium* (Draper et al. 2001). While an International Barley Genome Sequencing Consortium (IBSC, http://barleygenome.org) has been formed, the complete genome of barley may not be sequenced soon, due to its large genome size >5 Gbp (Bennett and Smith 1976) and highly repetitive nature (Flavell et al. 1974). Under such circumstances, a systematic and versatile methodology for gene localization at a specific region of interest would be exceptionally valuable.

Microarray technology has been used primarily to address biological questions related to gene expression. However, it has also been used to locate genes represented on the arrays to meiotic maps using single feature polymorphisms (SFPs) and transcript derived markers (TDMs) (Rostoks et al. 2005; Luo et al. 2007; Potokina et al. 2008a, b). This information is collected when investigating variation in transcript abundance as a proxy for phenotypic traits and has been termed expression quantitative trait loci (eQTL) mapping or genetical genomics (GG) (Jansen and Nap 2001). GG is a systems biology approach used to identify genetic determinants of gene expression variation represented as probes on a species-specific microarray. In GG experiments, mRNA from a selected tissue at the same physiological stage or under same treatment for each of the individuals of a segregating population are analysed and the observed transcript abundance data interrogated genetically as eQTL. These are broadly categorized depending on their location relative to their corresponding gene, where *cis*-eQTL represent a polymorphism physically located near the gene itself and *trans*-eQTL reside at locations distant from the genes, frequently on independent chromosomes.

Several eQTL studies have now been reported in plant species that include *Arabidopsis*, maize and eucalyptus (West et al. 2007; Keurentjes et al. 2007; Schadt et al. 2003; Shi et al. 2007; Kirst et al. 2005), along with barley: two with germinating embryos (Potokina et al. 2008a, b) and *Puccinia hordei*-infected seedling leaves (Chen et al. 2010a, b) and one with *P. tritici*-infected leaves (Wise, unpublished data). These have revealed common characteristics of *cis*- and *trans*-eQTL: *cis*-eQTL generally have larger, and *trans*-eQTL smaller effects (Keurentjes et al. 2007; West et al. 2007; Potokina et al. 2008a, b; Chen et al. 2010a, b). In addition, *cis*-eQTL are strongly context dependent often showing strong tissue specificity (Potokina et al. 2008a, b), whereas *trans*-eQTL show high QTL-by-environment interaction (Li et al. 2006). eQTL for a substantial number of genes that are not significantly differentially expressed between two parents can also frequently be detected in populations due to the multi-genic and transgressive nature of the expression traits. Perhaps not surprisingly, alleles that exhibit highly significant differential expression have a higher chance of being detected as eQTL in populations. For example, Potokina et al. (2008a, b) showed that over 90% of the most highly differentially expressed (DE) genes ($P \leq 0.05$) had detectable eQTL, and showed that DE genes represented 32% of all robust eQTL. In none of these eQTL studies has the locational correspondence between DE genes and their eQTL been fully exploited in the context of saturation genetic mapping.

BSA (Michelmore et al. 1991) has been widely adopted as a method to rapidly identify molecular makers in specific regions of a genome. The underlying principle is the bulking of individuals from a segregating population into pools each having an alternative phenotypes or genotypes at a particular locus, the outcome of which is assumed to generate a random genetic background at all other unlinked loci. The contrasting pools are then commonly screened to identify molecular markers or marker alleles that are exclusively found in only one of the two pools. The presence/absence of a marker or marker allele in a pool indicates close linkage to one of the two alleles of the phenotype or genotype used initially for pool construction. BSA has recently been used to identify differentially abundant anonymous transcripts using the cDNA-AFLP approach (Guo et al. 2006; Fernandez-del-Carmen et al. 2007) and for identification of genes with expression variation associated with both qualitative and quantitative traits in potatoes using microarrays (Kloosterman et al. 2010).

Here we investigated the potential of using BSA of differential transcript abundance to identify genes located near *Rphq11*, a major QTL for resistance to barley leaf rust (Marcel et al. 2007). We first re-analysed eQTL datasets generated from the Steptoe × Morex doubled haploid population using a 15K Agilent custom microarray (Chen et al. 2010a, b). Using the genotypic data from individuals in the population, we assembled *Rphq11* allele-specific bulked mRNA pools and analysed them using a custom 44K Agilent array. We found that DE genes were predominantly located in, or linked to, the target regions and conclude that BSA constitutes a novel method to identify genes located in specific genomic regions of interest, providing a platform for high-resolution genetic analysis either directly or by exploiting conservation of synteny with genomic models.

Methods

Plant materials, treatments and RNA samples

Barley cultivars Steptoe (*St*) and Morex (*Mx*), and the doubled haploid (DH) lines from the F1 cross between the two cultivars were used throughout. Total RNA samples used here were the same as reported by Chen et al. (2010a, b) and were prepared from replicated (four biological replicates for *St* and *Mx*) *Puccinia hordei*-infected seedlings (all lines) sampled 18 h post infection (hpi).

Construction of RNA bulks

Using the genotypic SNP dataset of 150 *St/Mx* DH lines from Close et al. (2009), a subset of 112 DH lines without missing values at the *Rphq11* locus, a QTL for resistance to barley leaf rust on chromosome 2H (Marcel et al. 2007) was selected for making the two bulks differing in genotype at this locus. In total, 64 and 48 lines had *St*- and *Mx*-genotypes at the locus, respectively. These two allele-specific groups were randomly split up into four sub-groups to represent four biological replicates. As a result, each *St*- and *Mx*-specific bulk comprised 16 and 12 DH lines, respectively. The RNA bulks were made by mixing an equal amount of RNA from respective *P. hordei*-infected individuals of each replicate.

Barley custom agilent microarray

BSA of differential expression was performed on Barley Agilent 44K arrays designed as previously described for Barley Agilent 15K arrays (Chen et al. 2010a, b) using eArray (Agilent http://www.chem.agilent.com; design number 015862 for the 15K array and 020599 for the 44K array). The microarray contains 42,302 60-mer oligonucleotide probes and was fabricated using Agilent's proprietary technology (http://www.chem.agilent.com). The probe identifiers and their corresponding DNA sequences of both arrays can be found at ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/; accession # A-MEXP-1533 (15 k) and A-MEXP-1728 (44 k)).

Deposition of microarray data

The raw microarray data and relevant experimental metadata, which are MIAME (minimum information about a microarray experiment) compliant, are deposited at the ArrayExpress microarray data archive (http://www.ebi.ac.uk/microarray-as/ae/) at the European Bioinformatics Institute (accession numbers: E-TABM-1069).

Microarray processing and data analysis for differential expression

mRNA from the two allele-specific bulks with four replicates were co-hybridised on the same two-channel (Cy3 vs. Cy5) arrays to measure the contrast in hybridization intensity between the two bulks. Microarray processing, including dye-swaps, and data extraction from microarray images were carried out as described previously by Chen et al. (2010a, b). Analysis of differential expression was performed using GeneSpring (v.7.3) after Lowess (LOcally WEighted polynomial regreSSion) normalisation. Differential expression was identified by *t* tests on log-transformed normalised ratio data with the significance threshold set at $P < 0.05$.

Conservation of synteny

In total, 3,000 gene-based homology bridges provided by the SNP-based barley gene map of Close et al. (2009) were used for comparative alignment with the rice genome. The best BlastN hits were used for the alignment, which revealed established blocks of conserved synteny.

Results

Positional relationship between differentially expressed genes and their eQTL

In a previous eQTL study with the Steptoe × Morex (*St/Mx*) reference barley population, we identified a total of 4,306 differentially expressed (DE) genes ($p < 0.05$) between the *P. hordei*-infected parental lines and a total of 15,685 eQTL observed from the genetic analysis of transcript abundance of 9,557 genes in the segregating progeny (Chen et al. 2010a, b). We further used these data here, assembling eQTL into two categories based on their LOD scores, then further dividing each category into DE and non-DE genes. We found that the majority of the DE genes were

associated with high LOD scores, and that low LOD scores were associated with the non-DE genes (Fig. 1). We then examined the map positions of eQTL for DE genes in relation to the location of their structural genes. We used both stringent ($P < 0.01$) and relaxed ($P < 0.05$) thresholds for the identification of DE genes and compared the results to test the effect of statistical stringency on DE gene identification and the robustness of eQTL prediction. Of the 9,557 genes with eQTL, 3,732 (39%) and 1,998 (20.9%) showed significant expression differences at $P < 0.05$ and $P < 0.01$, respectively between the infected parents. Out of the 3,732 and 1,998 DE genes, there were 694 and 458 that had previously defined map positions (Close et al. 2009; Potokina et al. 2008a, b), and 578 (83.3%) and 419 (91.5%) of these genes, respectively, detected eQTL that were located within 15 cM of their previously assigned map location (Table 1) which we considered most likely to be *cis*-eQTL. DE genes that had a single eQTL most frequently co-located with their structural genes (89.1 and 94.7% at $P < 0.05$ and $P < 0.01$). Overall, the results indicate that high LOD eQTL were more frequently observed from DE genes and that DE genes primarily detected eQTL close to their corresponding genes. Relaxing the stringency ($P < 0.01$ to $P < 0.05$) almost doubled the number of DE genes identified (1,998 vs. 3,732), while the fraction of those co-locating with their structural genes was only slightly compromised (91.5 vs. 83.3%). We therefore adopted the relaxed threshold in the following BSA experiment.

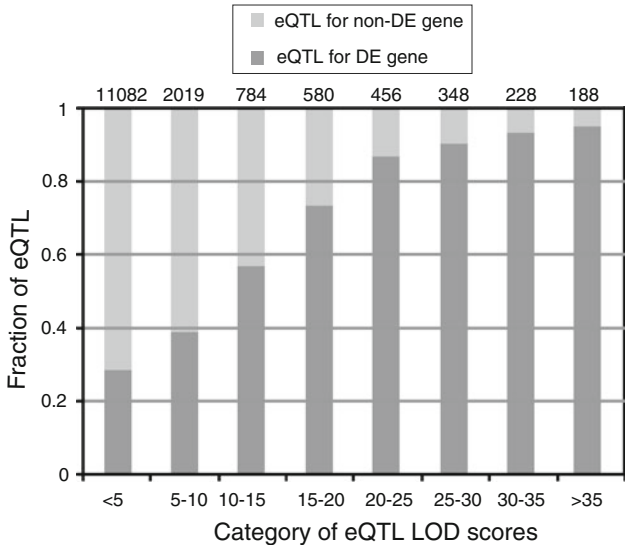

Fig. 1 Distribution of eQTL and its relationship with differentially expressed genes. Numbers over the *grey bars* represent the total number of eQTL in respective categories. DE genes and non-DE genes: expression differences between parents at $P < 0.05$ and $P > 0.05$, respectively

Genotypic characterization of the pools assembled for BSA

We used genome-wide genotypic data of the *St/Mx* DH population to characterize allele frequencies across the whole genome in the replicate bulks containing contrasting alleles at *Rphq11*. In a few cases, missing data were inferred from the genotype of neighboring markers. The proportions of *St* and *Mx* alleles were calculated on each of the four pairs of bulks individually, with the average allele frequency across all four bulks is shown in Fig. 2 (a1 and a2). In each individual case, 100% homogeneity of the *St* or *Mx* allele was obtained in an 8 cM interval defined by OPA-SNP flanking markers 1_0475 and 1_0649 (pilot OPA names) at 85–93 cM on chromosome 2H. The allelic homogeneity decayed on both sides (towards distal ends of chromsome 2H) as the distance between the markers and the target locus increased. The relative allele frequency at distal ends of chromosome 2H were 0.48 and 0.61 for the *St*-specific bulk (Fig. 2, a1) and 0.56 and 0.47 for the *Mx*-specific bulk (Fig. 2, a2). The overall allele frequency from the six other chromosomes were 0.50 and 0.48 for *St*-specific and *Mx*-specific bulk, respectively (SD 0.049

Table 1 Number of differentially expressed genes and their map locations relative to their eQTL

No. eQTL per gene	Overall genes	No. DE genes and percentages		Map position of gene relative to eQTL on SNP or TDM							
		$P < 0.05$	$P < 0.01$	$P < 0.05$				$P < 0.01$			
				All	Local	Distant	Local (%)	All	Local	Distant	Local (%)
1	5,103	2,081 (40.8%)	1,204 (23.6%)	460	410	50	89.1	321	304	17	94.7
2	3,074	1,162 (37.8%)	577 (18.8%)	165	126	39	76.4	101	86	15	85.1
3	1,122	404 (36.0%)	181 (16.1%)	54	35	19	64.8	30	25	5	83.3
4	227	76 (33.5%)	32 (14.1%)	13	6	7	46.2	6	4	2	66.7
5	26	9 (34.6%)	4 (15.4%)	2	1	1	50.0	0	0	0	–
6	5	0 (0.0%)	0 (0.0%)	0	0	0	–	0	0	0	–
Total	9,557	3,732 (39.0%)	1,998 (20.9%)	694	578	116	83.3	458	419	39	91.5

Information for genes previously mapped as SNP or TDM was used to determine their positions relative to their eQTL. Gene locations relative to eQTL were considered 'local' or 'distant' when genes were located within 15 cM (on consensus map scale) or outside their respective eQTL

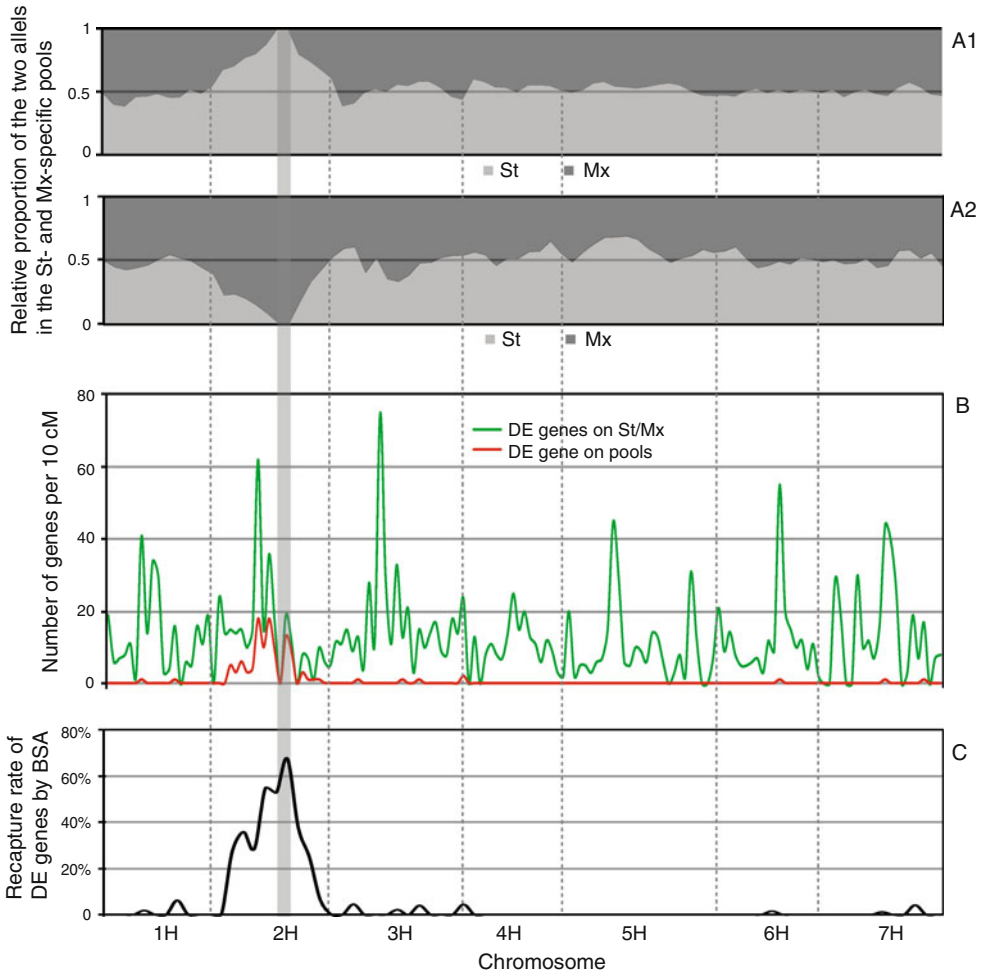

Fig. 2 An alignment across the 7 barley (H) chromosomes showing, **a** relative proportion of the St and Mx alleles in the St- and Mx-specific pools (a1 and a2 respectively), **b** distribution of the number of differentially expressed genes on St/Mx (*green line*) and pools (*red line*). The number of genes was counted on 10 cM sliding window along each chromosome and, **c** recapture rate of the genes with high LOD eQTL through bulked segregant analysis (BSA)

and 0.086) ranging between 0.38–0.59 and 0.31–0.67 (Fig. 2, a1, a2).

BSA for differential gene expression and their distribution across the genome

We performed BSA by co-hybridizing mRNA from the four pairs of allele-specific bulks to 44k barley Agilent arrays. A total of 411 genes were identified as significantly differentially expressed between the bulks at $P < 0.05$. Of these, 192 were represented by probes available on the 15K barley Agilent array which had previously been used for eQTL analysis and for assessing differential expression between the parents (Chen et al. 2010a, b). As was our intention, use of the 44K array, therefore significantly enriched the pool of potential candidate genes for *Rphq11*. Chen et al. (2010a, b) showed that eQTL with LOD scores of >10 almost always co-located with their corresponding genes and that high LODs were a feature of *cis*-eQTL. They identified LOD >10 eQTL for 2,582 genes, with 1,977 also significantly differentially expressed between parental lines. We exploited this eQTL dataset in the current study, finding eQTL information for 139 common genes. 113 of these 139 genes had eQTL LOD >10 (Chen et al. 2010a, b). Based on the previous conclusions about high LOD eQTL, we therefore, assumed that the map position of these 113 genes could be inferred directly from the location of their *cis*-eQTL. Figure 2b compares the genome-wide distribution of all DE genes between the parental lines (curves in green) and those detected here by BSA (curves in red). Of the 113 DE genes identified by BSA, 103 were located on chromosome 2H as three peaks that corresponded to the distribution of abundance of all mapped DE genes (curves in green) observed on parental lines in the same region. The remaining 10 genes were located on other chromosomes (Fig. 2b). We then

calculated the fraction of DE genes that were identified between the parents in our previous study (Chen et al. 2010a, b) that were also detected by BSA (Fig. 2c). We termed this the 'recapture rate'. The highest recapture rate (68%) was obtained within the 8 cM window at 85–93 cM where the two bulks had completely contrasting genotypes. The recapture rate decreased with distance from this genetic window and no DE gene could be detected when the average allele frequencies in both bulks fell in the range of 0.39–0.61. BSA based on RNA abundance therefore identifies genes that are predominantly located close to or within the target region and, as expected, the numbers decrease with distance.

Determining the map positions of remaining DE genes

To determine map positions for more DE genes identified by BSA, we exploited several other available 'gene-mapping' datasets. As *cis*-eQTL are tissue- and development-specific in many cases, the addition of a second tissue specific eQTL experimental dataset greatly increases the overall number of *cis*-eQTL (Potokina et al. 2008a, b). We therefore exploited an eQTL dataset generated on the same *St/Mx* DH mapping population, but using RNA from germinating embryos (Potokina et al. 2008a, b. As a different array platform was used in this experiment, we tested whether the LOD >10 threshold we used previously (Chen et al. 2010a, b) was also robust for this experiment. Potokina et al. (2008a, b) mapped 1,596 genes in their experiment and 1,116 were common with Chen et al. (2010a, b). Of these, 1,096 (98%) were located at similar positions (within 10 cM) (Table 2) indicating that LOD >10 was again a suitable threshold for *cis*-eQTL designation. We then used individual chromosome gene expression data derived from aneuploid genetic stocks (single barley chromosome addition lines in a wheat background) for chromosome assignments (Cho et al. 2006). And last, we included gene-based marker mapping data from two other studies, a SNP map (Close et al. 2009) and an RFLP-based transcript map (Stein et al. 2007). Using these combined datasets we were able to determine map positions for 181 DE genes identified by BSA and assign these to chromosomes. 162 (90%) were located on chromosome 2H and 22 on other chromosomes (see Supplementary Table).

Conservation of synteny shows that barley chromosome 2H is a composite of rice chromosomes Os04 (2H 0–20 cM and 60–160 cM) and Os07 (2H 27–59 cM). BLASTN searches using the 411 genes identified by BSA against rice genomic sequence identified 111 and 62 homologues located on chromosome Os04 and Os07, respectively. We positioned these genes 'virtually' onto the barley genetic map using the SNP map locations from Close et al. (2009) as references (Supplementary Table). Of the 173 genes with a predicted position based on conservation of synteny, 81 had also been determined by one of the methods referred to previously and 69 co-located. Overall, of the 411 DE genes identified by BSA, we were able to determine putative map positions for 276. Of these, 254 (92%) mapped to chromosome 2H with 22 located on different chromosomes (Supplementary Table). Our results suggest that the BSA using pooled RNA from individuals with opposite alleles at a target genetic locus is an effective approach for identifying tightly linked genes/markers.

Table 2 Number of genes detected as high LOD eQTL that map to a similar (<10 cM) location in experiments using RNA from geminating embryos (Aff) and pathogen-infected seedling leaves (Agi)

Chr.	Genes with eQTL LOD >10			Genes with eQTL detected same position LOD >10, both Aff and Agi
	Aff	Agi	Overlap	
1H	425	319	141	139
2H	570	391	184	180
3H	759	535	232	229
4H	359	251	107	104
5H	567	386	173	170
6H	294	275	113	109
7H	391	425	166	165
Total	3,365	2,582	1,116	1,096

Discussion

We report here the use of a combination of BSA and differential gene expression to locate genes to regions of the (unsequenced) barley genome that contain a QTL for leaf rust resistance. In a previous study we had defined the target locus using QTL analysis but wanted to obtain a better appraisal of the gene content of the region both for candidate gene identification based on differential gene expression signatures, and for assessing in detail the extent of regional conservation of synteny with model grass genomes, a conduit to additional genetic marker development and identification of positional gene candidates. Chen et al. (2010a, b) previously used pairs of NILs from different genetic backgrounds that were developed for different leaf rust resistance phenotypic QTL to achieve the same general objectives. While both approaches can be considered effectively similar, genetically they are fundamentally different, and these differences may have important consequences when differential gene expression (rather than bi-allelic markers) is the metric being used to differentiate the paired samples. In BSA the objective of pooling is to achieve a balanced 'effective heterozygosity' at all loci

across the genome, except for the locus containing the target gene which is fixed in opposite directions in each pool. In contrast, in NILs the only region of the genome that is (supposed to be) different is that surrounding the target locus.

Is one approach better than the other for meeting our stated objectives? For BSA, the individuals that are pooled are generally derived directly from a population that has already been used to map (or monitor) segregation of a target trait. BSA can be conducted as soon as that information is available. BSA has great versatility and is unlimited in making genotypically contrasting pools for different regions across the genome by using different individual lines from the same population. The resolution of BSA can be varied by increasing or decreasing the number of individuals included in each pool of samples or by narrowing the genetic interval when marker data is available. However, BSA can also be inherently biased if the amounts of substrate (i.e. RNA or DNA) are highly skewed towards one or a few individuals in each pool. This could have a particularly dramatic effect on the utility of RNA abundance analysis to differentiate the assembled pools because (unlike DNA) the latter are generally measured quantitatively. One result of an imbalance in pool membership could be eleborated as either high background outside the target region (which may lead to the suggestion that the target region contains trans-acting factors that influence gene expression at regions across the genome) or result in genes from outside the target interval being falsely included in it. While NILs may be genetically cleaner which helps to avoid some of these issues, their biggest problem is that they are limited by availability, taking many generations of backcrossing and selection followed by selfing to generate them. Furthermore, if flanking genetic markers have not been used in their development then the size of the introgression containing the target region is often large and as a result, resolution suffers.

We selected *Rphq11*, a QTL for resistance to barley leaf rust on chromosome 2H (Marcel et al. 2007) to investigate the feasibility of BSA of differential gene expressions to identify genes in a region of interest. Re-analysis of the data from Chen et al. (2010a, b) showed that over 83% of the differentially expressed genes ($P < 0.05$) were regulated in *cis*-eQTL (i.e. genes located close to their eQTL). In our previous analysis using nearly-isogenic lines (NILs) (Chen et al. 2010a, b), we discovered the same proportion (83%) of differentially expressed genes between *P. hordei*-infected NILs and their recurrent parents that fell into the target regions based on map information from high LOD (>10) eQTL in at least one of three eQTL studies (Potokina et al. 2008a, b; Chen et al. 2010a, b, R. Wise, unpublished data). We therefore conclude that differential gene expression observed between two lines is closely associated with *cis*-regulation. This relationship is reinforced by observations that almost all (over 98%) eQTL with LOD >10 were located close to their structural genes in experiments using both germinating embryos (Potokina et al. 2008a, b) and *P. hordei*-infected seedling leaves (Chen et al. 2010a, b).

The close association between DE and *cis*-regulation prompted us to investigate the feasibility of using BSA to identify and expand the number of genes that we could show were located in a specific region of the barley genome. We have successfully shown that 254 out of 276 (92%) genes (Supplementary Table) with a previously predicted map position were indeed located on the target chromosome 2H. Based on this, we expect that the remaining 135 genes identified by BSA (Supplementary Table) with a previously undetermined map position will also be predominantly located on chromosome 2H. The ability of BSA to identify differentially expressed genes covering a large chromosomal region within a single experiment highlights the efficiency of this as an approach for candidate gene identification.

We occasionally identified genes that were located on other chromosomes rather than the target chromosome 2H. While this is a common observation in BSA experiments with DNA based markers, and is most likely due to incomplete randomization across the genome in the bulks (as discussed above), this seems unlikely in our experiment given that we used four replicates each with over 12 independent individual plants and that no cluster of DE genes was observed in any region of the other six chromosomes. There is of course the possibility with an RNA abundance based assay that *trans*- acting factors on 2H are responsible for regulating genes on other chromosomes which then appear differentially expressed within the target interval.

While BSA based on RNA abundance allowed the chromosomal assignment of DE genes, to determine their exact map position still relies on other sources of information. This is no different from traditional BSA. Here, we took advantage of other publicly available gene marker mapping datasets (Close et al. 2009; Stein et al. 2007; Potokina et al. 2008a, b; Wise unpublished data), eQTL datasets (Potokina et al. 2008a, b; Chen et al. 2010a, b) and conservation of synteny, to determine locations. Mapping the respective genes using DNA-based genetic markers is preferred as it is the most reliable and accurate, although predicting map positions using high LOD *cis*-eQTL has proved robust for two previous eQTL studies (Potokina et al. 2008a, b; Chen et al. 2010a, b). The latter approach was generally successful in our study, where the locations of 47 genes inferred from *cis*-eQTL that were also mapped as gene-based SNP markers were in agreement. Only two exceptions where the SNP and eQTL data identified positions on different chromosomes were found for unigenes 7356 and 17572 (HarvEST25), (Supplementary Table).

In contrast to the high concordance between gene-based marker mapping and inference based on *cis*-eQTL, the prediction of genetic position based on conserved synteny with rice was less consistent. Previously, Cho et al. (2006) performed physical mapping and synteny analysis with rice using wheat-barley addition lines and revealed that 79% of physically mapped barley genes exhibited conserved synteny to rice at a chromosomal level, whereas Stein et al. (2007) found that only 50% of barley genes were collinear in both genomes. The accuracy of such predictions vary with factors such as density of reference genes and disruption of collinearity by ancestral rearrangements involving multiple linked or unlinked genes (Bilgic et al. 2007; Devos 2005). A consequence is that the orthologuous genes in two species many not lie in regions that otherwise exhibit significant levels of conservation of synteny (Tarchini et al. 2000; Li and Gill 2002). Alternatively a one-to-one relationship between orthologues does not exist between the species in question (Brueggeman et al. 2002; Griffiths et al. 2006). Predicting map position based on synteny alone can therefore be relatively inaccurate. We found that map locations predicted using synteny, when compared to map position determined by gene-based markers or inferred from *cis*-eQTL, showed 81% concordance. Thus, predictions based on synteny alone, could result in approximately 20% genes being erroneously assigned. In the work reported here, caution should therefore be exercised when considering the 92 genes determined solely on synteny.

BSA for differential RNA abundance using specific biological material will identify only a subset of putatively *cis*-regulated genes due to the tissue and developmental specificity of gene expression (Potokina et al. 2008a, b; Zhang and Borevitz 2009). For example, only 26% *cis*-eQTL were found to be in common to both germinating embryos and seedling leaves, and the addition of another tissue to eQTL analysis substantially increased the number of identified *cis*-eQTL (Potokina et al. 2008a, b). It is therefore conceivable that as more eQTL datasets are generated and become available from samples at different developmental stages or tissue types, or subjected to alternative biotic or abiotic stresses, the discovery of genes in regions of interest using the BSA approach described here could increase significantly.

We chose the *Rphq11* locus for differential expression by BSA, and identified a large collection of candidate genes. These are both positional- and expression-based candidates. It was reassuring that *HvPHGPx* (unigene727 and 2453 from HarvEST 35), a promising candidate gene identified in a previous eQTL study focused on *Rphq11* (Chen et al. 2010a, b), was reproduced in the current work as significantly differentially expressed, emphasizing the potential of RNA abundance-BSA for candidate gene identification. However, in cases where the exact location of a phenotypic QTL has not been defined accurately by flanking markers, linked genes could still be efficiently identified by creating bulks based on phenotype. These genes would serve as a valuable resource for targeted marker saturation, as candidate genes in their own right, and for comparative genetics.

Acknowledgments We gratefully acknowledge J. McNicol and C. Hackett for valuable discussions concerning the experimental and custom array design, T. Marcel, A. Vels, F. Yeo, A. Gonzalez, Z. Kohutova, F. Meijer-Dekens, R. Aghnoum, M. Macaulay and K. McLean for their kind help with sampling and Drs. G. Bryan and B. Thomas for their critical review of the manuscript. Funding for this experiment was provided by the European Union Bioexploit Grant No. 513959 (FOOD) to RW and RN (www.bioexploit.net) and by Scottish Government Rural and Environment Research and Analysis Directorate (RERAD) Programme 1, Work Package 1 (http://www.programme1.net/programmes).

References

Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond B Biol Sci 274:227–274

Bilgic H, Cho S, Garvin DF, Muehlbauer GJ (2007) Mapping barley genes to chromosome arms by transcript profiling of wheat–barley ditelosomic chromosome addition lines. Genome 50:898–906

Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J et al (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328–9333

Caldwell DG, McCallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R (2004) A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.). Plant J 40:143–150

Chen X, Niks RE, Hedley PE, Morris J, Druka A, Marcel TC, Vels A, Waugh R (2010a) Differential gene expression in nearly isogenic lines with QTL for partial resistance to *Puccinia hordei* in barley. BMC Genomics 11:629

Chen X, Hackett CA, Niks RE, Hedley PE, Booth C, Druka A, Marcel TC, Vels A, Bayer M, Milne I, Morris J, Ramsay L, Marshall D, Milne L, Waugh R (2010b) An eQTL Analysis of Partial Resistance to *Puccinia hordei* in Barley. PLoS One 5(1):e8598

Cho S, Garvin DF, Muehlbauer GJ (2006) Transcriptome analysis and physical mapping of barley genes in wheat–barley chromosome addition lines. Genetics 172:1277–1285

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szűcs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10: Article No. 582

Devos KM (2005) Updating the 'Crop Circle'. Curr Opin Plant Biol 8:155–162

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM (2001) Brachypodium distachyon: a new model system for functional genomics in grasses. Plant Physiol 127:1539–1555

Fernandez-del-Carmen A, Celis-Gamboa C, Visser RGF, Bachem CWB (2007) Targeted transcript mapping for agronomic traits in potato. J Exp Bot 58:2761–2774

Flavell RB, Bennett MD, Smith JB, Smith DB (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257–269

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752

Guo J, Jiang RHY, Kamphuis LG, Govers F (2006) A cDNA-AFLP based strategy to identify transcripts associated with avirulence in *Phytophthora infestans*. Fungal Genet Biol 43:111–123

Hori K, Kobayashi T, Shimizu A, Sato K, Takeda K, Kawasaki S (2003) Efficient construction of high-density linkage map and its application to QTL analysis in barley. Theor Appl Genet 107:806–813

Isidore E, Scherrer B, Bellec A, Budin K, Faivre-Rampant P, Waugh R, Keller B, Caboche M, Feuillet C, Chalhoub B (2005) Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region. Funct Integr Genomics 5:97–103

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17:388–391

Keurentjes JJB, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Basten SL, Peeters AJM, Vreugdenhil D, Koornneef M, Jansen RC (2007) Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. Proc Natl Acad Sci USA 104:1708–1713

Kirst M, Basten CJ, Myburg AA, Zeng ZB, Sederoff RR (2005) Genetic architecture of transcript-level variation in differentiating xylem of a eucalyptus hybrid. Genetics 169:2295–2303

Kleinhofs A, Graner A (2001) An integrated map of the barley genome. In: Phillips RL, Vasil IK (eds) DNA-based markers in plants, 2nd edn. Kluwer, Dordrecht, pp 187–199

Kloosterman B, Oortwijn M, Willigen J, America T, de Vos R, Visser RGF, Bachem CWB (2010) From QTL to candidate gene: genetical genomics of simple and complex traits in potato using a pooling strategy. BMC Genomics 11:158

Li W, Gill BS (2002) The colinearity of the *Sh2/A1* orthologous region in rice, sorghum and maize is interrupted and accompanied by genome expansion in the *Triticeae*. Genetics 160:1153–1162

Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, Hazendonk E, Prins P, Plasterk RHA, Jansen RC, Breitling R, Kammenga JE (2006) Mapping determinants of gene expression plasticity by genetical genomics in C-elegans. PLoS Genetics 2(12):e222

Lundqvist U, Franckowiak J, Konishi T (1996) New and revised descriptions of barley genes. Barley Genet Newsl 26:22–43

Luo ZW, Potokina E, Druka A, Wise R, Waugh R, Kearsey MJ (2007) SFP Genotyping from Affymetrix Arrays is robust but largely detects *cis*-acting expression regulators. Genetics 176:789–800

Marcel TC, Varshney RK, Barbieri M, Jafary H, de Kock MJD, Graner A, Niks RE (2007) A high-density consensus map of barley to compare the distribution of QTLs for partial resistance to *Puccinia hordei* and of defence gene homologues. Theor Appl Genet 114:487–500

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease resistance gene by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008a) eQTL analysis of 16, 000 barley genes reveals a complex pattern of genome wide transcriptional regulation. Plant J 53:90–101

Potokina E, Druka A, Luo Z, Moscou M, Wise R, Waugh R, Kearsey M (2008b) Tissue dependent limited pleiotropy affects gene expression in barley. Plant J 56:287–296

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson J, Wanamaker S, Walia H, Rodriguez E, Hedley P, Liu H, Morris J, Close T, Marshall D, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Gen Genomics 274:515–527

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lübberstedt T (2007) Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population. BMC Genomics 8:22

Sreenivasulu N, Graner A, Wobus U (2008) Barley genomics: an overview. Int J Plant Genomics Article ID 486258

Stein N, Prasad M, Scholz U, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney R, Perovic D, Grosse I, Graner A (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823–839

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice *Adh1–Adh2* region reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391

Varshney RK, Prasad M, Graner A (2004) Molecular marker maps of barley: a resource for intra- and interspecific genomics. In: Lörz H, Wenzel G (eds) Molecular marker systems. Springer, Berlin, pp 229–243

Varshney RK, Grosse I, Haehnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006) Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor Appl Genet 113:239–250

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101:9915–9920

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA (2007) Global eQTL mapping reveals the complex genetic architecture of transcript level variation in Arabidopsis. Genetics 175:1441–1450

Yu Y, Tomkins J, Waugh R, Frisch D, Kudrna D, Kleinhofs A, Brueggeman R, Muehlbauer G, Wise R, Wing R (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 101:1093–1099

Zhang X, Borevitz JO (2009) Global analysis of allele-specific expression in *Arabidopsis thaliana*. Genetics 182:943–954