Genetic variants associated with the root system architecture of oilseed rape (Brassica napus L.) under contrasting phosphate supply

Xiaohua Wang1,2, Yanling Chen1,2, Catherine L. Thomas3, Guangda Ding1,2, Ping Xu1, Dexu Shi1,2, Fabian Grandke4, Kemo Jin1,2, Hongmei Cai2, Fangsen Xu1,2, Bin Yi1,†, Martin R. Broadley3,†, and Lei Shi1,2,*,†

1National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China, 2Key Laboratory of Arable Land Conservation (Middle and Lower Reaches of Yangtze River), Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China, 3Plant and Crop Sciences Division, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK, and 4Department of Plant Breeding, IFZ Research Centre for BioSystems, Land Use and Nutrition, Justus Liebig University, Giessen 35392, Germany

*To whom correspondence should be addressed: Tel. +86 27 87286871. Fax. +86 27 87280016.
Email: leish@mail.hzau.edu.cn
†These authors jointly supervised and contributed equally to this work.

Edited by Dr. Masahiro Yano

Received 25 November 2016; Editorial decision 15 March 2017; Accepted 29 March 2017

Abstract

Breeding crops with ideal root system architecture for efficient absorption of phosphorus is an important strategy to reduce the use of phosphate fertilizers. To investigate genetic variants leading to changes in root system architecture, 405 oilseed rape cultivars were genotyped with a 60K Brassica Infinium SNP array in low and high P environments. A total of 285 single-nucleotide polymorphisms were associated with root system architecture traits at varying phosphorus levels. Nine single-nucleotide polymorphisms corroborate a previous linkage analysis of root system architecture quantitative trait loci in the BnaTNDH population. One peak single-nucleotide polymorphism region on A3 was associated with all root system architecture traits and co-localized with a quantitative trait locus for primary root length at low phosphorus. Two more single-nucleotide polymorphism peaks on A5 for root dry weight at low phosphorus were detected in both growth systems and co-localized with a quantitative trait locus for the same trait. The candidate genes identified on A3 form a haplotype ‘BnA3Hap’, that will be important for understanding the phosphorus/root system interaction and for the incorporation into Brassica napus breeding programs.

Key words: oilseed rape, root system architecture, single-nucleotide polymorphism, genome-wide association study, BnA3Hap
1. Introduction

The demand for oilseed rape (Canola; Brassica napus L.) as a food and fuel crop has increased global production to 24.6 Mt yr⁻¹ in 2013 (FAO 2013; http://faostat.fao.org/). Phosphorus (P) is critical in the production of oilseed crops, and with worldwide demand for P fertilizer rising, breeding P use efficiency into crops can be an effective strategy. Root system architecture (RSA) plays an important role in P utilization. This is due to low soil P mobility resulting in heteromorphic spatial variation within the soil, but also from variation in bio-availability due to varying soil pH, microbial activity, and colloid chemistry.²

Improved RSA traits have been identified in Brassica which deliver greater resistance to P starvation.³ For example, under low P in B. napus⁴ and B. oleracea⁵ primary roots become shorter and lateral roots increase in number and length. In A. thaliana, quantitative trait locus (QTL) LPR1 (Low Phosphate Root 1) is mapped to a 36-kb region on chromosome 1 in a recombinant inbred line (RIL) population.⁶ In rice (Oryza sativa), a major QTL for P uptake (Pup1) is mapped to a 150-kb region on chromosome 12 in an F₂ near-isogenic line (NIL) population.⁷ The gene conferring the phenotype (PSTOL1) is attributed to this region, and is a protein kinase in this QTL region that acts as an enhancer of early root growth, thereby enabling plants to acquire more phosphorus and other nutrients.⁸ In B. napus, a total of 62 significant QTL for total root length, root surface area, root volume, total dry weight and plant P uptake at two P levels were detected in three independent experiments with a F₁₀ RIL population under low and high P conditions. Two QTL clusters, uq.C3a and uq.C3b, were identified as specifically for root traits and P uptake under LP stress.⁹ A total of 131 QTL are identified for RSA under low and normal phosphate supply using a ‘pouch and wick’ system and an agar based system. QTLs for root and shoot biomass co-localized on chromosome A3, and a QTL for lateral root emergence colocalized with chromosomes A4/C4 and C8/C9 in a doubled haploid population (BnaTNDH).³,⁹

Genome-wide association studies (GWAS) have been successfully applied to the genetic dissection of complex traits in crops, such as drought resistance in Oryza sativa¹⁰,¹¹ and flowering time and leaf architecture in Zea mays.¹²,¹³ Compared to QTL mapping, GWAS can achieve a higher resolution of the underlying genetic loci and does not require a mapping population.¹⁴ A total of 40 SNP peaks were detected for symbiotic nitrogen fixation traits in bean (Phaseolus vulgaris).¹⁵ Six SNPs were attributed to root efficiency of P uptake in traditional rice.¹⁶ Seventy-four significant SNPs associated with P-efficiency were detected in soybean (Glycine max).¹⁷ Candidate genes for P efficiency in Arabidopsis,¹⁸ rice¹² and maize¹⁹ have been identified. GWAS has also been applied to the dissection of complex traits in B. napus, e.g. flowering time,¹⁴ seed weight and seed quality,²⁰ seed oil content,²¹ stem resistance to Sclerotinia sclerotiorum²² and seedling ionomic traits.²³

Haplotypes comprise SNPs or other markers on the same chromosome that are typically inherited together with little chance of recombination.²⁴ To understand the effects of allelic variation on seed oil content, 324 accessions of B. napus with favourable and undesirable haplotypes at ≥15 loci were grouped into four classes.²¹ Other important haplotypes were discovered for resistance to Sclerotinia stem rot in B. napus²⁵ and P efficiency in soybean.²⁶

In the present study, an association panel of 405 B. napus cultivars was genotyped with the 60K Brassica Infinium SNP array. RSA traits were investigated in ‘pouch and wick’ and hydroponic system experiments,²⁵ under low and high P treatments. The aims of this study were (i) to determine the genetic diversity and RSA of the population, (ii) to investigate the genetics of RSA under high and low P, and (iii) to predict candidate genes associated with the genetic loci to (iv) reveal a favourable haplotype for breeding P-efficient B. napus.

2. Materials and methods

2.1 Plant material

The association panel comprised 405 B. napus cultivars and inbred lines, including 342 semi-winter, 34 spring, 26 winter and 3 unknown types, collected from major breeding centers across China.²⁶ 370 lines originated in China, 20 from Europe, 5 from Canada, 4 from Korea and 2 from Japan (Supplementary Table S1).

2.2 Phenotyping root system architecture

The root system architecture (RSA) of the panel was screened at low P (LP: 5 µmol l⁻¹ P) and normal P (NP: 250 µmol l⁻¹ P) using a modified standard Hoagland’s solution in 2013-14. The seeds were sterilized using 70% (v/v) ethanol and NaOCl (2.5% active chlorine), and then germinated on gauze with pure water for three days. The seedlings were transplanted to a hydroponic growth tank and grown for the first week with a 25% standard Hoagland’s solution, and then with a 50% standard Hoagland’s solution followed by a 100% standard Hoagland’s solution for a further five days. After 12 days the primary root length (PRL) was measured by a ruler. Seedlings samples were dried at 80°C for seven days for dry weight measurements (shoot dry weight – SDW; root dry weight – RDW). The panel was also screened at LP (LP: 0 µmol l⁻¹ P) using a modified Hoagland’s solution whereby the KH₂PO₄ was replaced with KCl. The seedlings were grown using a ‘pouch and wick’ system,²⁶ adapted by Thomas.²⁷ Briefly, this method comprises growth pouches composed of germination paper down with the seedling growing between sheets of black polythene, the pouch is hung in a tank and suspended in nutrient solution, and each tank can accommodate 96 pouches (192 seedlings, one on each side of the paper). For each line, 16 seeds were sown across eight independent replicates distributed across four different aluminum frames. The nutrient solution was changed every 3 days. The RSA of samples were imaged at 14 days as previously.²⁵ The RSA including PRL (primary root length), LRN (lateral root number), MLRL (mean lateral root length = LRN/LRN), LRL (lateral root length), TRL (total root length = PRL + LRL) and LRD (lateral root density = LRN/PRL) was measured using RootReader 2D (http://www.plantmineralnutrition.net/r2d.php). Shoot and root samples were dried at 80°C and dry weights (shoot, SDW; root, RDW) were determined (Supplementary Fig. S1).

2.3 Data analysis

Restricted maximum likelihood (REML) procedures were used to estimate line means and sources of variation in root traits.²⁵ Means were estimated using the [Genotype] term as a fixed factor, retaining [Replicate/Seed of paper/Tray/Frame/Room] as random factors.³ Random terms and no defined fixed factor were used to estimate sources of variation. Statistical analysis was completed using GenStat15th Edition (VSN International, Oxford, UK).

2.4 Genotyping of the association panel and in silico mapping of SNPs

Four young leaves from different individuals of each variety were planted in Wuhan and were sampled. DNA was extracted using a
2.5 Population structure, relative kinship and linkage disequilibrium analysis of the association panel

The population structure was estimated using the Bayesian Markov chain Monte Carlo model (MCMC) in STRUCTURE V. 2.3.4. Each K value (the number of real populations) was obtained with five independent runs with a putative number of populations set from 1 to 10, based on a model for admixture and correlated allele frequencies. The length of the burn-in period and number of MCMC replications after burn-in were each set to 100,000. Based on the rate of change in log probability of the data (LnP(D)) and an ad hoc statistic AK between successive K values, the true K value was determined. The Q matrix, which integrates the cluster membership coefficient matrices of replicate runs from STRUCTURE, was determined using CLUMPP (http://www.stanford.edu/group/ro nerlan/glaba/clumpp.html). Genetic distances among varieties were calculated using PowerMarker version 3.25 (http://www.statgen.ncsu.edu/powermarker/downloads.htm). Based on the analysis of molecular variance, AMOVA, the subgroup differentiation, pairwise FST was measured using Arlequin V. 3.5.2.2 (http://cmpg.unibe.ch/software/arlequin35/). After double centering, distance matrices were used to obtain 'eigenvectors' in NTSYS V. 2.2 (http://www.exetersoftware.com/cat/ntsyspc/ntsyspc.html).

2.6 Genome-wide association analysis of RSA traits

Population structure and kinship were used to correct for spurious association created by genotype-phenotype covariance. GLM (generalized linear model) and mixed linear model (MLM) calculations were performed in TASSEL to determine the best model for association analyses. The GLM included a naive model without controlling for population structure, and the Q model which controlled for population structure using the Q matrix to identify populations. The MLM included the K model which assesses inter-individual relative kinship using the kinship matrix obtained in SPAGeDi software, and the Q + K model which controlled for population structure with the Q matrix, and kinship with the kinship matrix. Quantile-quantile plots were created with a negative log10 (P) value of the expected P-value from the genotype-phenotype association and the expected P-value from the assumption that no association exists between genotype and phenotype. The threshold of significance was set to P < 5.15 × 10⁻⁵ (1/total SNPs used).

2.7 Differential expression genes underlying the QTL and significant SNP loci

The B. napus cultivar 'Eyou Changjia' were planted at normal P (250 μmol l⁻¹) for ten days, and then half of the plants were transferred to a nutrition solution with zero P (0 μmol l⁻¹) for five days. The shoot (S) and root (R) tissues of three plants at normal P and zero P were separated and immediately placed in liquid nitrogen and stored at 80°C. The sequencing libraries of twelve RNA samples: S1_LP, S2_LP, S3_LP, S1_NP, S2_NP, S3_NP, R1_LP, R2_LP, R3_LP, R1_NP, R2_NP and R3_NP, were generated using the Illumina RNA Library Prep Kit and sequenced on an Illumina Hiseq 2000 platform with 100-bp paired-end reads. The B. napus reference genome was constructed using bowtie 2 (Broad Institute, Cambridge, MA, USA). Sequencing reads were aligned to the B. napus reference genome with bowtie2 and then assembled using TopHat 2.0.0 and Cufflinks. Gene expression was estimated using FPKM (fragments per kilobase of exon per million mapped fragments). Differentially expressed genes (DEGs) between two samples were identified with Cuffdiff, based on criteria False Discovery Rate (FDR) < 0.01 and |log2 (FPKM_LP/FPKM_NP)| > 2 (Ding, unpublished data).

The QTLs for root traits previously detected in the Tapidor x Ningyou7 doubled haploid (BnaTNDH) population were explored for colocalization with SNP marker associations in the present study. The SNP markers flanking co-located QTL and SNP regions on chromosomes A3 were mapped onto the genome of cultivar Darmor-bzh, respectively. The intervals of a significant locus ranged from loci minus LD to loci plus LD. Genes/transcription factors which were located in the homologous physical regions of the reference map and previously identified as DEGs (differentially expressed genes) responding to high and low P in the B. napus cultivar ‘Eyou changjia’, were selected as putative candidate genes: BnaA03g05800D, BnaA03g05820D, BnaA03g05830D, BnaA03g05850D, BnaA05g22460D, BnaA05g22070D.
Table 1. Mean, standard deviation (SD), minimum (min), maximum (max) and coefficient of Variation (CV, %) of the shoot dry weight (SDW, g), root dry weight (RDW, g), R/S ratio and primary root length (PRL) in the OSR panel, in the hydroponic system experiments in 2013 (HSI) and 2014 (HSII), under low phosphorus (LP) and normal phosphorus (NP) treatments

| Trait      | P Level | Mean ± SD 2013 | Min 2013 | Max 2013 | CV (%) 2013 | Mean ± SD 2014 | Min 2014 | Max 2014 | CV (%) 2014 |
|------------|---------|----------------|----------|----------|-------------|----------------|----------|----------|-------------|
| SDW (g)    | LP      | 0.129 ± 0.026  | 0.063    | 0.213    | 20.0        | 0.105 ± 0.028  | 0.030    | 0.209    | 26.3        |
|            | NP      | 0.357 ± 0.136  | 0.086    | 0.834    | 38.1        | 0.372 ± 0.144  | 0.069    | 1.266    | 38.8        |
| RDW (g)    | LP      | 0.041 ± 0.008  | 0.022    | 0.065    | 19.7        | 0.032 ± 0.008  | 0.012    | 0.0623   | 25.1        |
|            | NP      | 0.032 ± 0.012  | 0.008    | 0.081    | 36.4        | 0.030 ± 0.014  | 0.006    | 0.147    | 41.9        |
| R/S ratio  | LP      | 0.322 ± 0.049  | 0.198    | 0.502    | 15.2        | 0.312 ± 0.055  | 0.178    | 0.499    | 17.5        |
|            | NP      | 0.096 ± 0.031  | 0.023    | 0.213    | 32.2        | 0.091 ± 0.028  | 0.038    | 0.249    | 31.1        |
| PRL (cm)   | LP      | 31.00 ± 2.53   | 23.02    | 42.05    | 8.2         | 32.40 ± 3.33   | 21.69    | 41.78    | 10.38       |
|            | NP      | 25.8 ± 2.73    | 17.37    | 37.67    | 14.5        | 24.2 ± 2.85    | 15.61    | 37.36    | 11.8        |

Values with different letters show significant differences at P<0.05.

Table 2. Mean, standard deviation (SD), minimum (min), maximum (max) and coefficient of variation (CV, %) of the shoot dry weight (SDW, g), root dry weight (RDW, g), R/S ratio, primary root length (PRL), lateral root length (LRL), total root length (TRL), lateral root density (LRD), lateral root number (LRN) and mean lateral root length (MLRL) in the OSR panel, in the hydroponic system experiment, under a low phosphorus (LP) treatment

| Trait      | Mean ± SD 2013 | Min 2013 | Max 2013 | CV (%) 2013 |
|------------|----------------|----------|----------|-------------|
| SDW (g)    | 0.006 ± 0.002  | 0.002    | 0.011    | 28.4        |
| RDW (g)    | 0.002 ± 0.001  | 0.001    | 0.005    | 31.3        |
| R/S ratio  | 0.322 ± 0.068  | 0.177    | 0.585    | 21.2        |
| PRL (cm)   | 12.80 ± 2.70   | 4.50     | 22.70    | 21.2        |
| LRL (cm)   | 34.50 ± 9.63   | 10.51    | 71.16    | 27.9        |
| TRL (cm)   | 47.30 ± 11.70  | 18.45    | 90.49    | 24.7        |
| LRD (cm/cm²) | 1.30 ± 0.210  | 0.53     | 2.17     | 16.3        |
| LRN (N)    | 16.4 ± 4.3     | 5.0      | 33.5     | 26.4        |
| MLRL (cm)  | 0.65 ± 0.050   | 0.46     | 0.81     | 7.2         |

Table 3. SNP number, LD decay (physical distance on the genome when the value of r² is 0.2) and SNP density each chromosome in the OSR panel

| Chromosome | No. of SNPs | LD decay (kb) | SNP density per chromosome (SNP/kb) | SNP density per LD region (SNP/LD,100) |
|------------|-------------|---------------|-------------------------------------|---------------------------------------|
| A1         | 815         | 150           | 28                                   | 5                                     |
| A2         | 555         | 165           | 45                                   | 4                                     |
| A3         | 1224        | 125           | 24                                   | 5                                     |
| A4         | 837         | 125           | 23                                   | 5                                     |
| A5         | 886         | 200           | 25                                   | 8                                     |
| A6         | 898         | 125           | 27                                   | 5                                     |
| A7         | 1051        | 130           | 23                                   | 6                                     |
| A8         | 565         | 3300          | 33                                   | 100                                   |
| A9         | 916         | 600           | 37                                   | 16                                    |
| A10        | 895         | 475           | 19                                   | 25                                    |
| C1         | 1626        | 8250          | 24                                   | 347                                   |
| C2         | 1586        | 7125          | 29                                   | 250                                   |
| C3         | 1754        | 400           | 35                                   | 12                                    |
| C4         | 2131        | 6000          | 23                                   | 262                                   |
| C5         | 505         | 575           | 84                                   | 7                                     |
| C6         | 675         | 840           | 55                                   | 15                                    |
| C7         | 1016        | 775           | 39                                   | 20                                    |
| C8         | 980         | 1800          | 38                                   | 47                                    |
| C9         | 482         | 1350          | 98                                   | 14                                    |
| A genome   | 8642        | 250           | 27                                   | 35                                    |
| C genome   | 10755       | 1100          | 37                                   | 10                                    |
| Whole genome| 19397     | 1065          | 33                                   | 18                                    |

3 Results

3.1 Phenotypic variation in root system architecture at low and normal P

Extensive phenotypic variation was observed in all traits under the two growth systems (Supplementary Fig. S1). For example, in the hydroponic system (HS) under normal P (NP) in 2013, mean values ranged as follows: SDW 0.086–0.834 g, RDW 0.080–0.81 g, R/S ratio 0.023–0.213 and PRL 13.73–37.67 cm; and under low P (LP), SDW 0.063–0.213 g, RDW 0.022–0.065 g, R/S ratio 0.198–0.502 and PRL 23.02–42.05 cm. In the HS the RDW, R/S ratio and PRL were all greater at LP as compared with NP in both experimental years (HSI; HSII) (Table 1), and the coefficient of variation (CV) at NP was higher. In the ‘pouch and wick’ system (PS) experiment at LP, the panel phenotypes are shown in Table 2. All traits showed continuous variation with an approximate normal distribution (Supplementary Fig. S1).

3.2 LD decay, population structure and relative kinship

The 60K Brassica Infinium SNP array contains 52,157 SNPs. A total of 30,976 matched to a unique location in the reference genome of cultivar darmor-bzh and 19,397 high-quality SNPs with MAF > 0.05 and call frequencies of SNPs < 0.8 (Supplementary Table S3) were selected to assess the population structure, relative kinship, LD, and to conduct association analyses. SNP number on each chromosome ranged from 482 on C9 to 2131 on C4. The density of SNPs on each chromosome ranged from 98.31 kb/SNP on C9 to 19.23 kb/SNP on A10 (Table 3). LD decay on each chromosome ranged from 8250 kb on C1 to 125 kb on A3, A4 and A6 when r² was 0.2. SNP density in the LD region ranged from 4 SNP/LD chrA2 on A2 to 347 SNP/LDchrC1 on C1 (Table 3). The LD decay on the A genome, C genome and the whole genome was 1065 kb, 250 kb and
SNP density on chromosomes and in LD regions on the A genome were 27 SNP/kb and 35 SNP/LD_A, respectively, on the C genome were 36.97 SNP/kb and 10 SNP/LD_C, respectively and on the whole genome were 33 SNP/kb and 18 SNP/LD_C, respectively (Table 3).

The population structure of the *B. napus* panel was assessed. The K values increased continuously with no obvious inflexion points of LnP(K) values (Fig. 1a). The \( D_K \) values indicated that the rate of change adhoc statistic was equal to 2, therefore the population could be divided into two clusters; Subgroup1 and Subgroup2 (Fig. 1e). The \( F_{ST} \) of Subgroup1 and Subgroup2 was 0.027 (\( P < 0.05 \)). A total of 66 lines were in Subgroup1; 56, 3, 2, 2, 2, and 1 line(s) originated from China, Korea, Australia, Denmark, Germany, and Sweden, respectively. A total of 339 lines were in Subgroup 2; 314, 2, 1, 2, 5, 9, 2, 2, 1 and 1 line originated from China, Japan, Korea, Australia, Canada, Germany, France, Sweden, Poland and Denmark, respectively (Supplementary Table S1). The LD decay distance in Subgroup2 (875 kb) was longer than in Subgroup1 (812.5 kb) and shorter than in the whole panel (1065 kb). Gene diversity of Subgroup2 (0.4010) was higher than in Subgroup1 (0.3847). The PIC in Subgroup2 (0.3175) was higher than in Subgroup1 (0.3057) (Fig. 1d). The Kinship coefficient of Subgroup2 (0.0697) was higher than Subgroup1 (0.0540). The \( r^2 \) in Subgroup 2 (0.3452) was higher than in Subgroup1 (0.3122) (Fig. 1d).

The relative kinships of the association panel were evaluated. The lines with kinship coefficients < 0.05 accounted for 67.9, 69.5 and 70.6%, and the lines with kinship coefficients equal to 0 accounted for 57.4, 51.3 and 62.4% in the whole panel, Subgroup 1 and Subgroup 2, respectively (Fig 1d), indicating that the majority of the accessions have a weak relationship with one another (Fig. 1c).

3.3 Genome wide association mapping of RSA

The Naïve model, Q model, K model and Q + K model were used for association mapping. The deviations of observed values from the expected values were shown in the QQ plots (Supplementary Fig. S2) and were applied to select the most suitable model for each trait in the different experiments. Significant associations between SNPs and traits were identified by scanning the genome with a \( P < 5.15 \times 10^{-5} \) threshold. In the hydroponic experiment I (HSI) there were 41...
significant SNPs at LP and 9 at NP. Under LP; 7 SNPs for PRL were detected with the Q model; 4 SNPs for RDW were detected with the Q + K model; 6 SNPs for R/S ratio and 24 SNPs for SDW were detected with the Q model. Under NP; 1 SNP for PRL, 2 for RDW, 3 for R/S ratio and 3 for SDW were detected with the Q model. In the hydroponic experiment II (HSII), there were 31 significant SNPs at LP and 58 at NP. Under LP; 6 SNPs for PRL were detected with the Q + K model; 7 SNPs for RDW were detected with the naive model; 6 SNPs for R/S ratio were detected with Q model and 12 SNPs for SDW detected with the Q model. At NP; 9 SNPs for PRL, 2 SNPs for R/S ratio and 37 SNPs for SDW were detected with the Q model; and 10 SNPs for RDW with the naive model with low P treatment as the Q model (Supplementary Table S4). In the ‘pouch and wick’ system (PS) 165 significant SNPs were identified at LP; 28 SNPs for PRL, 33 for RDW, 22 for R/S ratio, 32 for TRL, 5 for RRD, 31 for LRL and 7 for MLRL were detected with the Q model and 3 SNPs for SDW and 4 SNPs for LRN were detected with the Q + K model (Supplementary Fig. S3 and Table S4).

3.4 Co-locating SNPs and QTLs

Nine significant SNPs detected with GWAS analyses co-localized with the intervals of the QTLs for the same traits in a previous linkage analyses of the BnaTNDH population,2 including one SNP on A3 for PRL, two SNPs on A5 for RDW (Supplementary Fig. S4a-c), three SNPs on A7 and C4 for SDW, and one SNP on A3 for TRL detected at LP; two SNPs on A3 for SDW detected at NP (Table 4). Therefore, the co-localized loci were the same traits under the same P treatment in the two studies. The two nearly significant marker positions on A3; Bn-scaffold_16893_1-p69538 and Bna-A03-p18820057 at 16.76–17.25 Mb for RDW at LP, were detected in both systems employed in this study (Supplementary Fig. S4b and c). Moreover, there was a significant SNP marker Bna-A03-p3001467 at 2.49–2.72 Mb on A3 detected for all root traits measured in the PS experiment at LP (Fig. 2b–h). Bna-A03-p3001467 could explain phenotypic variation (PVE) of PRL (9.2%), LRN (6.0%), LRL (12.4%), MLRL (6.3%), TRL (13.8%), RDW (14.5%) and SDW (5.9%) at LP (Table 3), and thus was the main locus responding to P deficiency.

3.5 Gene expression profile of candidate genes

A total of 88 candidate genes were within the QTL confidence interval on A3 (Fig. 2i and Supplementary Table S5); BnaA03g05800D, BnaA03g05820D, BnaA03g05830D and BnaA03g05850D were located in the LD decay region of the peak SNP Bna-A03-p3001467 at 2.49–2.72 Mb, and formed the haplotype blocks on A3 (BnaA3Hap).

Therefore, the co-localized loci were detected for the same traits under the same P treatments employed in this study (Supplementary Fig. S4a–c). Moreover, there was a significant SNP marker Bna-A03-p3001467 at 2.49–2.72 Mb on A3 detected for all root traits measured in the PS experiment at LP (Fig. 2b–h). Bna-A03-p3001467 could explain phenotypic variation (PVE) of PRL (9.2%), LRN (6.0%), LRL (12.4%), MLRL (6.3%), TRL (13.8%), RDW (14.5%) and SDW (5.9%) at LP (Table 3), and thus was the main locus responding to P deficiency.
at both P levels, the relative expression of \( BnaA03g05820D \) was higher in the shoot at NP and in the root at LP, the relative expression of \( BnaA03g05830D \) was higher in the root at LP, and the relative expression of \( BnaA03g05850D \) was lower in shoot at LP (Fig. 3). This provides further evidence that the four putative genes were closely associated with P efficiency.

3.6 BnA3Hap haplotypes associated RSA traits and geographical distribution

There were 21 SNPs located in the QTL confidence intervals on A3 (Fig. 4h). Eight successive SNPs located near SNP Bn-A03-p3052882 at 25.6 kb and the SNP Bn-A03-p3001467 at 75.2 kb formed nine haplotypes (Fig. 4g), designated 'BnA3Hap haplotypes'. The varieties with haplotype 6 (GAAAAAGG) had higher average values in all the investigated RSA traits than those with the other haplotypes (Fig. 4a–f); two of the three varieties with haplotype 6 originated from European countries. Whereas, the varieties with haplotype 2 (AGCCGAGG) had lower average values in all of the investigated RSA traits than those with the other haplotypes (Fig. 5). Fifty-one of the 53 varieties with haplotype 2 originated from China, Shanxi (18), Hunan (17), Jiangsu (41), Chongqing (55), Hubei (131), and Qinghai (9) provinces (Fig. 5). Tajima’s D test of each SNP loci of BnA3Hap in B. napus indicated varieties from Gansu, Shanxi and Jiangsu provinces of China and Canada showed haplotype-BnA3Hap selection (Fig. 6). The alignment of \( BnaA03g05800D, BnaA03g05820D, BnaA03g05830D \) and \( BnaA03g05850D \) in B. napus was identical to that of their orthologous genes in B. rapa, however, they differed from orthologs genes in A. thaliana. One gene was located on chromosome 3, and three were located on chromosome 5 of A. thaliana, suggesting hybridization between B. rapa and B. oleracea, while the BnA3Hap was conserved in the A genome (Supplementary Fig. S6).

4 Discussion

4.1 Genetic diversity, relative kinship and LD

In previous B. napus association panels, two subgroups were identified in 192 B. napus accessions\(^{21}\) and 523 B. napus accessions\(^{21}\) but three subgroups in 472 B. napus accessions.\(^{20}\) In the present study, 405 varieties of the panel of Liu\(^{21}\) were used and similarly divided
into two subgroups (Fig. 1). The LD, gene diversity, PIC, and relative kinship were slightly different between Subgroup1 and Subgroup2 (Fig. 1). As compared with European B. napus, the Chinese B. napus accessions showed a more diverse genetic background (Supplementary Table S1), due to greater recombination in the A subgenome introduced through inter-specific hybridizations between B. napus and B. rapa. The LD in the C genome is greater than in the A genome (Table 3), but lower in PIC and gene diversity, indicating that the C genome may contribute more valuable alleles to adapt to different growth environments. In addition, the LD decay observed here- 1.06 Mb was different from previous B. napus panels, which observed 0.5–1 cM in 192 inbred lines,32 2.0 Mb in 472 inbred lines33 and 6.5 Mb in 523 inbred lines,14 and larger than that of other species, e.g. 100 kb–1 Mb in rice,31 1–100 kb in maize,12 and 250 kb in Arabidopsis.34 The comparatively high resolution provided by association mapping is dependent upon the LD of genome.3 The low LD detected in this study indicated that this B. napus panel is suitable for association analysis and has the potential to identify SNPs in a narrow intervals equivalent to the distance of LD decay (1.06 Mb). In addition, the number of SNPs, the length of LD decay, and the SNP density on each chromosome were different in this study (Table 3), similar to the studies with 52314 and 47230 accessions.

In this study, we did not investigate the association between the population structure and the breeding history, as the cultivars in this panel lack breeding information history. The subpopulations in a panel of 70 tomato lines corresponded to historical patterns of breeding conducted for specific production environments.35 Second, RSA traits in the BnaTNDH population were screened at low and high P with a ‘pouch and wick’ and an agar culture system, respectively. RSA traits of GWAS population were screened using a solution culture system at both low and high P and a ‘pouch and wick’ system at low P. The growth conditions influenced genotypic ranking and the range of variability.39 Third, uncorrelated data may be the result different environmental parameters between experiments.

The response of RSA traits to low Pi availability in B. napus has been well characterized.3,8 With low P availability, typically a reduction in the development of the primary roots occurs,1,4,9 and increases the number and length of lateral roots are observed.3,35 LPR2 (At1g71040) was shown to regulate primary root development in Arabidopsis under low Pi availability. It is a paralogue of Arabidopsis LPR1 (At1g23010), which encodes a multicopper oxidase protein, which are essential for the reducing root growth observed when root tips contact low-Pi media.3,35 The RACK1 orthologous gene in Arabidopsis regulates lateral root formation by changing the sensitivity of the plant to gibberellin, brassinosteroid, abscisic acid, and auxin.40

The genes associated with significant SNPs for root traits in the loci of target QTL were predicted to be the candidate genes for P

4.2 Candidate genes identified for RSA

In this study, only 9 out of 285 SNPs colocalized with previously identified QTLs for RSA traits, which can be attributed to three reasons. First, as compared with QTL mapping, GWAS can identify relatively higher resolution genetic loci from underlying traits because natural populations incorporate ancient recombination events. Second, RSA traits in the BnaTNDH population were screened at low and high P with a ‘pouch and wick’ and an agar culture system, respectively. RSA traits of GWAS population were screened using a solution culture system at both low and high P and a ‘pouch and wick’ system at low P. The growth conditions influenced genotypic ranking and the range of variability.39 Third, uncorrelated data may be the result different environmental parameters between experiments.
Figure 4. The haplotypes associated with RSA (root system architecture) traits in the ‘pouch and wick’ system (PS) at low P (LP). (a). Lateral root length (LRL). (b) Lateral root number (LRN). (c) Mean lateral root length (MLRL). (d) Primary root length (PRL). (e) Root dry weight (RDW). (f) Total root length (TRL). (g) BnA3Hap haplotypes. (h) The BnA3Hap haplotypes on A3 (darker colour trilateral).

Figure 5. The geographical distribution of the BnA3Hap haplotypes in the B. napus panel. The percentage of the varieties in a region accounted for in the panel is shown in the circle.
efficiency. BnaA03g05800D, BnaA03g05820D, BnaA03g05830D and BnaA03g05850D were predicted to be closely associated with P efficiency, which were also located in the confidence interval of the QTL for primary root length at low P as well as the LD decay region of the peak SNP Bna-A03-p3001467 at 2.49–2.72 Mb on A3 (Fig. 2). All genes had significantly different expression in the P-efficient cultivar ‘Eyou Changjia’ between low P and high P (Fig. 2). The differential expression of the putative candidate genes in response to low-P stress was also confirmed in the P-efficient and P-inefficient accesses (Fig. 4). BnaA03g05800D in B. napus is orthologous to AT5G16010 in Arabidopsis, which encodes a 3-oxo-5-alpha-steroid 4-dehydrogenase family protein that moves between various organs under normal or nutrient-liming conditions.\(^{41}\) BnaA03g05820D is orthologous to AT3G02330 encoding a TCP family transcription factor.\(^{52}\) BnaA03g05830D is homologous to AT3G16110, which regulates lateral root development.\(^{53}\) Along with MEX1, the AT3G16150- plastidic GLC translocator (PGLCT) is essential for the growth and development of starch degradation products from chloroplasts and starch-mediated pathway for photoassimilate export in Arabidopsis.

BnaA03g05800D, BnaA03g05820D, BnaA03g05830D and BnaA03g05850D had differential SNPs within the 5' UTR or introns; the A/G alleles of lead SNP Bna-A03-p3073797 were located in the introns of BnaA03g05800D, A/G alleles of lead SNP Bna-A03-p3084593 were located in the introns of BnaA03g05820D, A/G alleles of lead SNP Bna-A03-p3091011 were located in the introns of BnaA03g05830D, and the A/G alleles of lead SNP Bna-A03-p3097703 were located in introns of BnaA03g05850D. These polymorphisms may cause differential gene expression between P-efficient and P-inefficient accesses (Figs. 3 and 4h). The characterization of the genetic elements associated with these traits could therefore provide elite alleles for the breeding of crops adapted to low P.

### 4.3 BnaA3Hap haplotypes

Breeders have matched functionally different P-tolerance alleles to specific environments during selection. Oilseed rape was selected as a crop 300–400 years ago.\(^{39}\) In this study, the SNPs within candidate genes (BnaA03g05800D, BnaA03g05820D, BnaA03g05830D and BnaA03g05850D) on A3 formed a haplotype ‘BnaA3Hap’, contributes to B. napus tolerance to low P stress (Fig. 4g and h). BnaA3Hap haplotypes were detected at low P under the ‘pouch and wick’ system (Fig. 4a–f), and verified under the hydroponic system for two years.

In addition, the Tajima’s D values of BnaA3Hap haplotypes in B. napus varieties were strongly associated with the distribution of the agronomic P surpluses in soil\(^{44}\) (Figs. 5 and 6). Compared to ancestor species (B. rapa and A. thaliana), the order of the four orthologous genes of BnaA3Hap haplotypes in A3 in B. napus showed collinearity with that in B. rapa, and were slightly different to that in A. thaliana (Supplementary Fig. S6), which indicates that this haplotype contributed to six RSA traits at LP (Fig. 4a–f) that were conserved during B. napus breeding. The discovery of the optimal haplotype of BnaA3Hap can now enable the accurate selection of B. napus with higher P efficiencies and improve our understanding of the molecular mechanisms underlying P efficiency in plants.

### Supplementary data

Supplementary data are available at DNARES Online.

### Conflict of interest

None declared.

### Funding

The authors acknowledge the financial support from the National Nature Science Foundation of China (Grant No. 31471933), the UK Biotechnology and Biological Sciences Research Council (BBSRC) (Crop Improvement Research Club, CIRC, Project BB/J019631/1; International Partnership Award BB/J020443/1), New Century Excellent Talents in University of Ministry of Education of China (Grant No. NCET-13-0809), Natural and Fundamental Research Funds for the Central Universities of China (Grant No. 104PY20; 2662013PY05).

### References

1. Kochian, L.V. 2012, Plant nutrition, rooting for more phosphorus. Nature, 488, 466–7.
2. Lynch, J.P. 2015, Root phenes that reduce the metabolic costs of soil exploration, opportunities for 21st century agriculture. Plant Cell Environ., 38, 1775–84.
3. Shi, L., Shi, T., Brodley, M.R., et al. 2013, High-throughput root phenotyping screens identify genetic loci associated with root architectural traits in Brassica napus under contrasting phosphate availabilities. Ann. Bot.-London, 112, 381–9.
