GWAS of anion content in *Brassica napus*

Corresponding author:
Stanislav Kopriva
Institute of Botany
University of Cologne
Zülpicher Str. 47b
50674 Köln
Germany
Tel. +49 221 4708530
Fax +49 221 4701516
e-mail: skopriva@uni-koeln.de

Research area: Ecophysiology and Sustainability
Dissection of control of anion homeostasis by Associative Transcriptomics in *Brassica napus*

Anna Koprivova¹, Andrea L. Harper², Martin Trick, Ian Bancroft², Stanislav Kopriva¹*

John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

¹ Current address: Institute of Botany, University of Cologne, 50674 Köln, Germany
² Current address: Department of Biology, University of York, Heslington, YO10 5DD, UK

One sentence summary:
Associative Transcriptomics in *Brassica napus* identified candidate genes for control of variation in nitrate, phosphate, and sulfate contents.
Footnotes

S.K.’s and I.B.’s research is supported by grants BB/J004561/1, BB/H004351/1 and BB/L002124/1 from BBSRC, and the John Innes Foundation.
Abstract
To assess the variation in nutrient homeostasis in oilseed rape and to identify genes responsible for this variation we determined foliar anion levels in a diversity panel of *Brassica napus* accessions, 84 of which had been genotyped previously using mRNA sequencing. We applied Associative Transcriptomics (AT) to identify sequence polymorphisms linked to variation in nitrate, phosphate, or sulfate in these accessions. The analysis identified several hundred significant associations for each anion. Using functional annotation of Arabidopsis homologues and available microarray data we identified 60 candidate genes for controlling variation in the anion contents. To verify that these genes function in control of nutrient homeostasis we obtained Arabidopsis T-DNA insertion lines for these candidates and tested them for accumulation of nitrate, phosphate, and sulfate. Fourteen lines differed significantly in levels of the corresponding anions. Several of these genes have been shown previously to affect accumulation of the corresponding anions in Arabidopsis mutants. These results thus confirm the power of AT in dissection of the genetic control of complex traits and present a set of candidate genes for use in improvement of efficiency of *B. napus* mineral nutrition.
INTRODUCTION
Plants require 14 essential mineral nutrients (Marschner, 2012). The nutrients most commonly limiting plant growth are nitrogen (N), phosphorus (P), potassium (K), and sulfur (S), which are present in relatively large amounts in plant tissues and therefore named macronutrients (Maathuis, 2009; Marschner, 2012). N, P, and S are taken up as oxygenic anions and either stored in the vacuoles or assimilated into organic compounds. The availability of these minerals in soil is usually low, so for intensive agriculture it has to be improved by fertilization, adding significant monetary and environmental costs. A major target for crop improvement is, therefore, to relieve their dependence on high levels of mineral fertilizers and improve nutrient use efficiency (Parry and Hawkesford, 2012).

Nutrient use efficiency (NUE) is defined as yield per unit of input (Good et al., 2004). NUE depends on the ability to efficiently take up the nutrient from the soil, but also on transport, storage, mobilization, usage within the plant, and even on the environment (Good et al., 2004; Rengel and Marschner, 2005). The partitioning of the nutrients between vacuolar storage and assimilation is thus an important contributing factor of their use efficiency.

Several approaches have been taken to understand the genetic basis of nutrient homeostasis. Firstly, the response of plants to nutrient deficiency stress has been explored to identify processes affected by such stress and regulatory networks (Hammond et al., 2003; Hirai et al., 2003; Wang et al., 2003; Wu et al., 2003; Wang et al., 2004; Hirai et al., 2005; Nikiforova et al., 2005; Krouk et al., 2010).

Another major approach to dissect control of complex traits, such as NUE, makes use of natural genetic variation (Loudet et al., 2003; Gallais and Hirel, 2004; Chardon et al., 2012; Weigel, 2012). These traits can be analysed through quantitative trait locus (QTL) analysis (Loudet et al., 2003; Harada et al., 2004; Reymond et al., 2006; Habash et al., 2007; Loudet et al., 2007; Ding et al., 2010) or genome-wide association studies (GWAS) (Atwell et al., 2010; Chan et al., 2011; Harper et al., 2012). The usefulness of GWAS has been demonstrated by capturing numerous well-characterised candidate genes (Aranzana et al., 2005; Atwell et al., 2010). Traits connected with accumulation of mineral elements have also been successfully analysed using GWAS (Atwell et al., 2010; Chao et al., 2012).

Understanding of the control of nutrient homeostasis is particularly important for crop plants as it may contribute to improving NUE and reduction of fertiliser use. Both QTL and GWAS have been used not only in model species, but also directly in crops, such as oilseed rape (Ding et al., 2010; Harper et al., 2012). Due to its polyploidy, B. napus presents a significant
challenge for GWAS, which has however, been successfully circumvented by using mRNA sequencing for identification of the polymorphic molecular markers in an approach termed Associative Transcriptomics (AT) (Harper et al., 2012). The validity of the approach was demonstrated by identification of a polymorphism in a MYB28 gene, encoding a transcription factor controlling synthesis of aliphatic glucosinolates, being responsible for accumulation of the corresponding glucosinolates in seeds (Harper et al., 2012).

Here we show results of application of the AT approach to dissect genetic control of variation in nitrate, phosphate, or sulfate in leaves of *B. napus*. Candidate genes indicated by GWAS were tested using Arabidopsis T-DNA insertion lines resulting in 14 lines affected in accumulation of the corresponding anion. These genes include genes known to affect nutrient homeostasis as well as those, which have never been associated with nutrition. The study thus confirms the power of AT and presents a set of interesting genes for detailed analysis of their roles in control of plant mineral nutrition and potentially improvement of NUE of crops.

RESULTS AND DISCUSSION

Anion content in *B. napus* varieties

The *Brassica napus* core diversity set, comprising 99 varieties (Table S1), was planted in the field at John Innes Centre site. Two leaf disks from the youngest fully developed leaves of eight week old plants were sampled and used for determination of nitrate, phosphate, and sulfate. The anion concentrations in the leaves varied substantially in the different varieties (Figure 1). Nitrate was found to be the most variable anion, its levels varying from 0.31 to 25.7 μmol/g FW, whereas phosphate was found in the range of 2.1 to 10.4 μmol/g FW and sulfate levels were typically the highest between 12.5 and 41.7 μmol/g FW (Table S1).

Interestingly, whereas phosphate levels showed normal distribution among the accessions and sulfate levels were also close to normal distribution, nitrate concentrations showed a very different pattern. More than a third of the accessions contained very low nitrate levels, under 2 μmol/g FW, whereas ten accessions, i.e. 12%, contained more than 10 μmol/g FW nitrate (Figure 1). In most varieties the most abundant anion was sulfate (Table S1), which accumulates to high levels in various Brassica species (Blake-Kalff et al., 1998), and not nitrate as typically observed in Arabidopsis (Figure 2D, (Mugford et al., 2009)). The difference in nitrate concentration is remarkable. It is not linked to a specific group of *B. napus*, since varieties very low (less than 1 μmol/g FW) and very high (more than 10 μmol/g FW) are found among spring oilseed rapes, winter oilseed rapes, or swedes. Low nitrate
content is a desirable trait, as it is connected with high nitrogen utilisation efficiency, however, the significance of such contrasting accumulation of this anion is unknown and represents an intriguing question for further research.

**Associative Transcriptomics**

To determine the regions of genome involved in control of anion homeostasis we made use of the AT approach recently developed for *B. napus* (Harper et al., 2012). This approach correlates variation in quantitative traits with markers derived from transcript sequencing. In a proof of concept, known QTLs for erucic acid content were correctly identified showing thus the utility of the use of transcript-derived markers for GWAS (Harper et al., 2012). We used the diversity panel of 84 accessions for which the functional genotypes (i.e. single nucleotide polymorphism, SNP, scores in expressed sequences) were reported in Harper et al. (2012). The analysis, using a linear mixed model to correct for population structure, revealed a large number of SNPs significantly associated with variation in the anion content. For nitrate, 151 SNPs were linked with significance of P<0.005, whereas for phosphate and sulfate the analysis revealed 146 and 168 significant markers, respectively (Table S2).

**Identification of candidate genes**

Utilising the colinearity between *B. napus* and Arabidopsis chromosomes, GWAS output identifies Arabidopsis orthologs of genes in which the SNPs are located (Harper et al., 2012). This allows immediate access to functional annotation of the genes and to a large variety of Arabidopsis resources. The Arabidopsis orthologs were, therefore, used as a basis to identify suitable candidate genes for causing the variation in anion content. Several candidate genes were immediately obvious. For example, among SNPs identified in the nitrate dataset, the SNP with 5\(^{th}\) lowest P-value is localized in a *B. napus* homologue of the CLC-A nitrate channel (At5g40890) (Table S3), which was previously shown to contribute to control of nitrate levels in Arabidopsis leaves (De Angeli et al., 2006). Also among the candidate genes containing significantly associated SNPs were genes for enzymes involved in nitrate assimilation aspartate aminotransferase 5 (At4g31990) and glutamine synthetase 2 (At5g35630). In the phosphate dataset genes for H(+-)-ATPase 1 (At2g18960) and hypothetical phosphate/phosphoenolpyruvate translocator (At5g33320) were found to contain the associated markers. Also in the dataset based on sulfate levels several clear candidates were identified directly, such as S-adenosylmethionine synthetase (At4g01850), cysteine synthase C (At3g59760), or a 3(2),5 -bisphosphate nucleotidase SAL2 (At5g54390) (Table
However, the significant markers themselves are often not the causal polymorphisms affecting the traits, but are only linked (Myles et al., 2009). Thus, the genes underlying the variations and containing the causal polymorphic SNPs are found within the linkage disequilibrium of the markers. For Arabidopsis, such linkage disequilibrium is about 10 kbp (Kim et al., 2007), whereas for *B. napus* it is 300-1000 kbp (Delourme et al., 2013). We therefore again exploited the colinearity of Arabidopsis and *B. napus* genomes and continued the analysis on the Arabidopsis genome. We inspected Arabidopsis genes within +/- 15 kbp from each significant marker for potential candidate genes controlling nutrient homeostasis using two complementary criteria. Firstly, the functional annotations were used as a guide to search for genes involved in nutrient uptake or assimilation and signalling. Secondly, we hypothesised that genes controlling nutrient levels in the leaves are likely to be regulated by the deficiency of the corresponding nutrient and therefore assessed the expression of these genes in Arabidopsis microarray data for nitrate, phosphate, and sulfate deficiency (Wang et al., 2003; Wu et al., 2003; Maruyama-Nakashita et al., 2006). Together these analyses resulted in 60 candidate genes, 21 of which directly included the SNPs used as GWAS markers (Table S3). Thirty genes were identified in the nitrate dataset, 11 in phosphate set and 19 genes potentially control variation in sulfate content. The expression analysis contributed only 6 genes to the candidate list (Table S3).

**Analysis of candidate genes**

The next step of analysis was based on the hypothesis that mutants in genes controlling variation of anion content will be affected in the levels of such anions. We obtained T-DNA insertion lines for the candidate genes available from the Nottingham Arabidopsis Stock Centre and recovered homozygous mutants by standard PCR genotyping. Fourteen mutants were obtained for genes potentially controlling nitrate levels, nine and fifteen mutants were recovered from the phosphate and sulfate sets, respectively (Table S4). For the CLC-A gene, a previously described *clea-1* mutant in Ws background (Geelen et al., 2000) was provided by Sébastien Thomine, CNRS Gif-sur-Yvette. These mutants were grown in controlled environment conditions and anion levels in leaves were determined (Figure 2). From the 39 mutant lines tested, 12 showed significantly different nitrate levels and 15 each differed in phosphate and sulfate. Among these, two lines were simultaneously affected in nitrate and phosphate, eight lines in nitrate and sulfate, and 6 lines in phosphate and sulfate. Interestingly, when a second anion was affected in the lines with different phosphate levels, the changes were always in the same direction, i.e., for lines with higher phosphate levels
also nitrate or sulfate were higher. On the other hand, in the lines where both nitrate and sulfate were different from wildtype, the changes were complementary, increase in nitrate was accompanied by decrease in sulfate and vice versa. Since many of the candidate genes are involved in pathways important for general growth and development, it is possible that some of the nutrient effects, particularly when multiple nutrients are affected, are due to pleiotrophic effects of the mutations.

However, while relatively high proportion of the selected lines, 31% or 38%, were indeed affected in anion contents, only a few lines showed changes in the same anion for which they were selected as candidates. Thus, among the 14 candidates for control of nitrate content, only three differed in nitrate, while among the nine candidates for regulation of phosphate homeostasis only three were confirmed. On the other hand, in the set of 15 candidates for control of sulfate content eight showed correspondingly lower sulfate levels. Altogether the analysis resulted in identification of 14 genes potentially responsible for variation in anion content in Brassica napus (Table 1).

**Genes potentially affecting anion homeostasis**

The different levels of the respective anions in the 14 mutants were independently confirmed in 13 of them (Supplemental Figure S1). The 13 genes include several genes for which an evidence for or a clear link to causing variation in anion levels were described before, as well as genes that were not previously associated with plant nutrition. Apart from the CLC-A, which was shown to affect nitrate homeostasis in Arabidopsis (De Angeli et al., 2006), the 3(2),5'-bisphosphate nucleotidase SAL2 also has a direct relevance for control of anion content, as mutants of the closely related paralog FIERY1 possess significantly lower sulfate levels (Lee et al., 2012). From several candidate genes functioning in nitrate metabolism (Table S4) only the ASPARTATE AMINOTRANSFERASE 3 has been confirmed to affect nitrate levels when disrupted. The H(+-)ATPase 1 (At2g18960) may contribute to maintaining phosphate levels as H(+-)ATPases were associated with response to phosphate deficiency (Tomasi et al., 2009). Two genes with a direct relevance to sulfur metabolism were identified in the sulfate dataset. Cysteine synthase C catalyses the last step of primary sulfate assimilation, synthesis of cysteine (Heeg et al., 2008; Takahashi et al., 2011). The At3g48990 gene is annotated as functioning in response to cadmium, where sulfur compounds, such as phytochelatins, play a very important role (Sarry et al., 2006). Finally, the At3g06420 gene encodes one isoform of the ATG8 protein involved in autophagy (Thompson et al., 2005). Autophagy has been recently revealed as an important process in
nutrient homeostasis, particularly at nutrient limitation, as shown in studies with nitrate (Guiboileau et al., 2012) and sulfate (Alvarez et al., 2012). It is thus possible that it also contributes to nutrient homeostasis during vegetative growth, since the mutant in At3g06420 accumulated nitrate, however, the mechanisms of such contribution remains to be elucidated. Other genes, however, do not have obvious functions in control of plant nutrition. Particularly interesting is the identification of two genes involved in ethylene signalling \textit{ETHYLENE INSENSITIVE 2} (\textit{EIN2}) and \textit{EIN5} associated with variation in sulfate and phosphate levels, respectively. Ethylene signalling is involved in several responses to phosphate deficiency, regulation of gene expression, production of acid phosphatases and accumulation of anthocyanins (Lei et al., 2011). Even more evidence is present on role of ethylene in regulation of sulfate starvation response. The central regulator of sulfate starvation response, SLIM1, is a member of EIN3-like family of transcription factors (Maruyama-Nakashita et al., 2006). Ethylene induces directly activity of the key enzyme of sulfate assimilation, adenosine 5'-phosphosulfate reductase, and ethylene signalling is important for regulation of this enzyme by salt (Koprivova et al., 2008). Arabidopsis mutants in ethylene signalling are more sensitive to selenium, which primarily affects sulfate uptake and assimilation (Van Hoewyk et al., 2008). Ethylene production is affected by nitrate supply and ethylene signalling is involved in regulation of nitrate transporters (Zheng et al., 2013), showing a general role of this phytohormone in control of plant nutrition. The findings of altered anion content in \textit{ein2} and \textit{ein5} mutants and the links between \textit{EIN2} and \textit{EIN5} and variation in sulfate and phosphate levels in the \textit{B. napus} accessions thus support this conclusion and point to the necessity of more detailed investigations of the role of ethylene in control of plant mineral nutrition.

Two genes from the candidate list are part of the CUL4 RING ubiquitin ligase complex. The complex catalyses ubiquitination of a variety of proteins in different cellular pathways to control their degradation (Angers et al., 2006). The CUL4 RING ligase has been associated for example with DNA repair and control of photomorphogenesis (Chen et al., 2006; Kapetanaki et al., 2006). The two genes identified in the study belong to DWD proteins, substrate receptors of the CUL4 RING complex, which often regulate development (Lee et al., 2008). However, protein degradation leads to increased amino acid recycling and may so modulate the demand for reduced nitrogen or sulfur and in this way affect the homeostasis of the stored anions nitrate and sulfate. How the two DWD proteins are associated with plant nutrition, however, needs further study.
The At4g24350 gene encodes an uncharacterised protein belonging to phosphorylase gene family with a predicted nutrient reservoir activity. However, the corresponding protein contains a nucleoside phosphorylase domain and so may be involved in turnover of nucleotide phosphates and so affect phosphate homeostasis. The last gene from the list, At4g01010, encodes a cyclic nucleotide gated channel AtCNGC13. These calmodulin binding channels have a plethora of functions in plants, from transport of potassium and other monovalent cations, over disease response signalling, to thermal sensing (Maser et al., 2001; Ali et al., 2007; Finka et al., 2012). AtCNGC13 has not been characterised so far, however, reduced expression of its closest paralogue AtCNGC10 resulted in many developmental and growth defects, such as early flowering, starch accumulation, and growth retardation (Borsics et al., 2007). Given the number of functions members of this gene family have in plants it is not possible to predict function of AtCNGC13, but its interesting phenotype in sulfate accumulation encourages detailed investigation.

Power of Associative Transcriptomics

The number and nature of genes potentially controlling nutrient content confirms the value of AT in genetic dissection of complex traits. First, previously known QTLs for erucic acid content were confirmed using this approach (Harper et al., 2012). Further, analysis of association based on gene expression levels revealed that seed glucosinolate content in *B. napus* is controlled by expression of *MYB28* (Harper et al., 2012). Interestingly, we observed no significant associations between gene expression levels and anion content at the selected high level of significance, $10^{-6}$, necessary to avoid false positives (Harper et al., 2012), all significant associations were found among the SNP markers. In the proof of concept experiments (Harper et al., 2012), the expression markers were identified with significance between $10^{-7}$ and $10^{-9}$, whereas the most highly significant SNP markers reached only $p=10^{-5}$ (Harper et al., 2012). The presence of *CLC-A* and *SAL2* among the candidate genes showing significant effects on anion contents is further confirmation of the power of this approach. The analysis, however, revealed other gene candidates that were not linked to mineral nutrition before. The putative function of these genes was confirmed by analysis of T-DNA lines, showing that AT has the potential to uncover new links between metabolic and/or regulatory pathways. In Arabidopsis GWAS approaches also recovered previously known candidate genes (Aranzana et al., 2005; Atwell et al., 2010; Chan et al., 2011; Chao et al., 2012), but reports of new discoveries, are still scarce (Angelovici et al., 2013; Meijon et al., 2014; Verslues et al., 2014).
The candidate genes were found in the analysis of 84 *B. napus* accessions, while previous study showed a strong association based on only 53 varieties. First Arabidopsis GWAS experiments were based on 96 accessions, but it has been calculated that at least 192 accessions are needed for a comprehensive sampling of genetic diversity (Atwell et al., 2010). The population structure of the *B. napus* diversity set might thus be more suitable for these studies as fewer accessions are needed to obtain robust results. It should also be noted that the plant material used for AT originated from a field trial and not from controlled conditions. This confirms the robustness of the genome wide methods even for nutrition traits that are naturally dependent on microenvironment. In addition, the AT approach was able to cope with the nitrate dataset, which is far from normal distribution of the values, since in nitrate and phosphate datasets the number of significantly associated markers was similar and the number of confirmed candidates was identical. The prediction success was, however, not as high in nitrate (20%, 3 out of 15 tested) as for sulfate (47%) and phosphate (33%).

To test, whether candidate genes from AT have higher probability to be affected in nutrient content than genes of similar annotation but picked randomly we performed an *in silico* experiment. Since no anion dataset is publically available we used the total element data in the ionomics resources at www.ionomicshub.org (Baxter et al., 2007). We searched for data from T-DNA lines in genes between At1g60000 and At1g69000 with a similar annotation as we used for our selection, and compared concentration of total phosphorus, selenium, as its distribution strongly correlates with sulfur, and iron in leaves. This interval was selected because it contains two genes with known effects of accumulation of these elements, *APR2* for sulfur (and Se) and iron transporter *IRT3*. Indeed both mutants showed significant effects on Se and Fe concentration, respectively. From the twenty analysed mutants, two were affected in P (p<0.05), four in Se, and three in Fe (Table S5). Thus, the success of the random set to identify genes affected in nutrient homeostasis was with 10-20% clearly significantly lower than the success rate of the candidates from AT analysis of sulfate and phosphate datasets and similar to the nitrate set.

From the three datasets analysed in this study, most candidates were recovered for control of sulfate levels. Interestingly, the two genes shown to control variation in sulfate levels in Arabidopsis have not been found among the candidate genes. These two genes were shown to underlie QTLs from analysis of Bay-0 x Shahdara population and they encode isoforms of two consecutive enzymes of the sulfate assimilation pathway, ATP sulfurylase and adenosine 5′-phosphosulfate (APS) reductase (Loudet et al., 2007; Koprivova et al., 2013). The *APR2* gene encoding an isoform of APS reductase in Shahdara contained a unique SNP that resulted
in an amino acid substitution near the active centre and inactivation of the enzyme (Loudet et al., 2007). ATP sulfurylase affects sulfate levels due to variation in transcript levels of the main isoform ATPS1, caused by deletion in intron 1 found in a few accessions (Koprivova et al., 2013). Unique or very rare haplotypes, such as the Sha allele of APR2, are almost impossible to identify through GWAS and it is thus no surprise that the gene was not recovered. For ATPS1, the number of accessions studied may not have been high enough to identify the small variation in expression levels observed in Arabidopsis. Thus, although the number of candidates for genes controlling trait variation is large, the list is still not complete and other genes can be identified by other methods, e.g. QTL analysis.

CONCLUSIONS
We have performed a genome wide association study of leaf anion content in B. napus using AT. The analysis resulted in identification of 13 genes, potentially involved in control of sulphate, nitrate, or phosphate levels. First, genes within linkage disequilibrium of the significant markers were inspected. Genes annotated as involved in the nutrient uptake and metabolism, signalling networks, and transcription factors were marked for further analysis, as well as genes affected in expression by the particular nutrient deficiency. Homozygous T-DNA insertion lines for these genes were obtained and analysed for anion content. This pipeline can be generally adopted to facilitate identification of new genes controlling trait variation from GWAS results not only in Arabidopsis but also B. napus. The presence of several genes known to affect anion levels in the resulting gene list confirms the suitability of the approach. The mechanisms of action and the causal polymorphisms of these candidates will be subjects of future more detailed studies.

MATERIAL AND METHODS
Plant material and growth conditions
The plant material for initial anion content measurements was derived from a field grown diversity set of 99 Brassica napus lines, as listed in Supplemental Table S1. The seeds were germinated and grown in long-day glasshouse conditions (16 h photoperiod) at 15 °C (400 W HQI metal halide lamps). Plants were pricked out after 11 d and transferred into a field at JIC site, arranged into a four block, one-way randomized design with one plant of each of the accessions per block and randomized within each block. Leaf disks were cut from mature leaves of 8 weeks old plants ca. 4 hours into the light period and immediately frozen in liquid nitrogen.
For functional analysis of the candidate genes *Arabidopsis thaliana* ecotype Col-0 was used as wild type, except for the clcA-1 mutant (FLAG_171A06), obtained from S. Thomine, CNRS Gif-sur-Yvette, which is in Ws background. T-DNA lines disrupting the candidate genes were obtained from the Nottingham Arabidopsis Stock Centre (NASC) and genotyped by PCR (for accession numbers and primers see Table S4) to obtain homozygous mutants. Plants for analysis were grown for 5 weeks in controlled environment room under a short day 10 h light/14 h dark cycle at constant temperature of 22°C, 60% relative humidity, and light intensity of 160 µE s⁻¹ m⁻². Four individual plants from each genotype were analysed and the experiment was independently repeated.

**Measurement of anion content**

For anion measurements ca. 50 mg of frozen plant material was homogenised in 1 mL of deionised water and the anions, nitrate, phosphate, and sulfate, separated by HPLC on an IC-PAK ion exchange column as described in (Scheerer et al., 2010).

**Associative transcriptomics**

The previously developed SNP dataset (Supplementary Data File 6 of Harper et al., 2012) had previously been entered into the program STRUCTURE 2.3.3 (Pritchard et al., 2000) for Bayesian population structure analysis. An admixture model with independent allele frequencies was used and K was set between 1 and 5, each repeated three times with a burn-in length of 100,000 followed by 100,000 iterations of the Monte Carlo Markov Chain algorithm. The method of (Evanno et al., 2005) was employed to estimate the number of clusters that best represents the dataset. Once the optimal number of K populations was established, K-1 Q matrix scores for each individual could be used as a fixed effect in the subsequent association analysis.

The trait data, STRUCTURE Q matrix, and SNP data was entered into the program TASSEL V3.0 (Bradbury et al., 2007) (Bradbury et al., 2007). Minor allele states below 0.05 were removed from the SNP dataset leaving 62,980 SNPs (Harper et al., 2012), and a kinship (K) matrix was calculated to estimate the pairwise relatedness between individuals. These datasets were entered into a Mixed Linear Model (MLM) with optimum compression and P3D variance component estimation to decrease computing time for the large dataset.

Using methods and scripts described in (Bancroft et al., 2011) and (Higgins et al., 2012) along with sequence datasets previously submitted to SRA (under accessions ERA122949,
ERA036824 and ERA063602), transcript quantification was undertaken for the *B. napus* accessions used in this study. The sequence reads were mapped to the “cured” reference described in Harper et al., (2012), which comprised the A and C genome versions of each unigene (189,116 unigene sequences in all). Transcript abundance was quantified and normalized as reads per kb per million aligned reads (RPKM) separately for the A and C genome versions of each unigene. The relationship between gene expression and the trait values were calculated by linear regression using R (http://www.R-project.org/). For each unigene, previously calculated RPKM values (Harper et al., 2012) were regressed as the dependent variable on the trait value as the independent variable, and $R^2$ and significance values were calculated for each unigene.

Statistical analysis

The normality of distribution of anions in *B. napus* varieties was tested using the Kolmogorov-Smirnov Test (http://www.physics.csbsju.edu/stats/KS-test.html). Significant differences between the WT and the T-DNA lines were analysed according to the Student’s t-test for $P < 0.05$.

SUPPLEMENTAL MATERIAL

Supplemental Table S1. Anion contents in leaves of *Brassica napus* varieties.
Supplemental Table S2. Markers significantly associated with variation in anion levels in *B. napus*.
Supplemental Table S3. Candidate genes for analysis of T-DNA lines.
Supplemental Table S4. T-DNA lines used for analysis.
Supplemental Table S5. Variation in content of phosphorus, selenium, and iron in leaves of randomly selected T-DNA from www.ionomicshub.org

ACKNOWLEDGEMENTS

This work was supported by UK Biotechnology and Biological Sciences Research Council grants BB/H004351/1, BB/J004561/1 and BB/L002124/1, and the John Innes Foundation.
FIGURE LEGENDS

Figure 1. Distribution of the anion contents among the *B. napus* varieties.
A nitrate, B phosphate, and C sulfate contents were determined in leaves of 99 *B. napus* varieties by HPLC. Shown is the frequency of anion levels among the 99 accessions.

Figure 2. Anion contents in the candidate genes.
A nitrate, B phosphate, and C sulfate contents were determined in leaves of Arabidopsis T-DNA lines for genes potentially affecting anion homeostasis by HPLC. Lines originating from a nitrate screen are presented in black, those from phosphate dataset are in light grey and lines from a sulfate screen are shown in dark grey. D Anion levels in *clcA* mutant and its corresponding wild type Ws. Asterisks mark values significantly different from wild type at P<0.05 (T-test).
References


Table 1. Candidate genes for control of anion content.

<table>
<thead>
<tr>
<th>dataset</th>
<th>gene</th>
<th>annotation</th>
<th>P values compared to Col-0</th>
<th>anion levels compared to Col-0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>nitrate</td>
<td>phosphate</td>
</tr>
<tr>
<td>nitrate</td>
<td>At5g40890</td>
<td>nitrate transporter CLC-A</td>
<td>0.000003</td>
<td>0.2699</td>
</tr>
<tr>
<td>nitrate</td>
<td>At3g06420</td>
<td>autophagy related protein</td>
<td>0.0016</td>
<td>0.4297</td>
</tr>
<tr>
<td>nitrate</td>
<td>At5g11520</td>
<td>aspartate aminotransferase 3</td>
<td>0.0344</td>
<td>0.2792</td>
</tr>
<tr>
<td>phosphate</td>
<td>At2g18960</td>
<td>H(+-)-ATPase 1</td>
<td>0.2424</td>
<td>0.0414</td>
</tr>
<tr>
<td>phosphate</td>
<td>At1g54490</td>
<td>ethylene insensitive 5</td>
<td>0.1843</td>
<td>0.0008</td>
</tr>
<tr>
<td>phosphate</td>
<td>At4g24350</td>
<td>nutrient reservoir activity</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>sulfate</td>
<td>At3g48990</td>
<td>response to cadmium ion</td>
<td>0.0385</td>
<td>0.1206</td>
</tr>
<tr>
<td>sulfate</td>
<td>At3g18060</td>
<td>CUL4 RING ubiquitin ligase complex</td>
<td>0.0404</td>
<td>0.1091</td>
</tr>
<tr>
<td>sulfate</td>
<td>At3g59760</td>
<td>cysteine synthase C</td>
<td>0.3775</td>
<td>0.0139</td>
</tr>
<tr>
<td>sulfate</td>
<td>At5g03280</td>
<td>ethylene insensitive 2</td>
<td>0.1146</td>
<td>0.0059</td>
</tr>
<tr>
<td>sulfate</td>
<td>At2g46560</td>
<td>CUL4 RING ubiquitin ligase complex</td>
<td>0.0517</td>
<td>0.0045</td>
</tr>
<tr>
<td>sulfate</td>
<td>At4g01010</td>
<td>cyclic nucleotide-gated channel 13</td>
<td>0.0016</td>
<td>0.1984</td>
</tr>
<tr>
<td>sulfate</td>
<td>At5g54390</td>
<td>3(2),5 -bisphosphate nucleotidase SAL2</td>
<td>0.0080</td>
<td>0.0969</td>
</tr>
</tbody>
</table>