Dichotomy in the NRT Gene Families of Dicots and Grass Species

Darren Plett¹, John Toubia¹, Trevor Garnett¹, Mark Tester¹, Brent N. Kaiser²*, Ute Baumann¹

¹ Australian Centre for Plant Functional Genomics, Waite Research Institute, University of Adelaide, Adelaide, South Australia, Australia, ² School of Agriculture, Food and Wine, Waite Research Institute, University of Adelaide, Adelaide, South Australia, Australia

Abstract

A large proportion of the nitrate (NO₃⁻) acquired by plants from soil is actively transported via members of the NRT families of NO₃⁻ transporters. In Arabidopsis, the NRT1 family has eight functionally characterised members and predominantly comprises low-affinity transporters; the NRT2 family contains seven members which appear to be high-affinity transporters; and there are two NRT3 (NAR2) family members which are known to participate in high-affinity transport. A modified reciprocal best hit (RBH) approach was used to identify putative orthologues of the Arabidopsis NRT genes in the four fully sequenced grass genomes (maize, rice, sorghum, Brachypodium). We also included the poplar genome in our analysis to establish whether differences between Arabidopsis and the grasses may be generally applicable to monocots and dicots. Our analysis reveals fundamental differences between Arabidopsis and the grass species in the gene number and family structure of all three families of NRT transporters. All grass species possessed additional NRT1.1 orthologues and appear to lack NRT1.6/NRT1.7 orthologues. There is significant separation in the NRT2 phylogenetic tree between NRT2 genes from dicots and grass species. This indicates that determination of function of NRT2 genes in grass species will not be possible in cereals simply on sequence homology to functionally characterised Arabidopsis NRT2 genes and that proper functional analysis will be required. Arabidopsis has a unique NRT3.2 gene which may be a fusion of the NRT3.1 and NRT3.2 genes present in all other species examined here. This work provides a framework for future analysis of NO₃⁻ transporters and NO₃⁻ transport in grass crop species.

Introduction

Nitrogen use efficiency (NUE) in plants is determined by the efficiency with which the plant acquires and uses nitrogen. Nitrate (NO₃⁻) is the primary nitrogen source for most plants in agricultural soils; cereal crops, however, access only 33–50% on average of the NO₃⁻ applied to the soil by farmers [1,2]. In order to improve this efficiency a more complete understanding of the transport of NO₃⁻ from the soil to the plant and within the plant itself is required. An important first step towards improving the NO₃⁻ uptake capacity and the NUE of crop plants would be characterisation of the transporters responsible for NO₃⁻ transport. Either the expression of the relevant genes or else the function of the proteins encoded by the genes could then be manipulated through traditional plant breeding or genetic engineering in order to improve NO₃⁻ uptake characteristics. With this goal in mind, the aim of this research was to identify the NO₃⁻ transporters in grass species.

The transport of NO₃⁻ and the transporters involved in this process have best been characterised in Arabidopsis due both to its amenability to physiological analyses and availability of genetic resources. The transport of NO₃⁻ is mediated largely by members of the NRT gene families which have recently been reviewed [3]. The Arabidopsis genome contains 33 NRT1/PTR family genes [3]. Only AtNRT1.1 to AtNRT1.8, however, have functional analyses indicating that these proteins do indeed transport NO₃⁻. The NRT1 family comprises predominantly low-affinity NO₃⁻ transporters, with the exception of AtNRT1.1 which appears to mediate dual-affinity NO₃⁻ transport [4,5] based on phosphorylation status of the amino acid residue T101 [6]. A recent study indicated that AtNRT1.1 may also function as an NO₃⁻ sensor [7]. The expression of AtNRT1.2 is constitutive and located predominantly in the root epidermis indicating that the encoded transporter may also be involved in NO₃⁻ uptake from the soil [8]. The expression of AtNRT1.3 in roots is repressed by exposure to NO₃⁻ and is induced by NO₃⁻ deprivation; its functional role, however, remains less clear [9,10]. AtNRT1.4 is expressed primarily in the leaf petiole and appears to be involved in NO₃⁻ storage [11]. AtNRT1.5 appears to mediate NO₃⁻ efflux and to have a role in the loading of NO₃⁻ into the xylem for transport to the shoot [12]. AtNRT1.6 is involved in transporting NO₃⁻ from maternal tissue to developing embryos [13]. AtNRT1.7 has been identified as playing role in the remobilisation of NO₃⁻ from older to younger leaves through facilitating phloem loading [14]. Very recently Li et al [15] have shown that AtNRT1.8 is responsible for retrieving NO₃⁻ from the xylem parenchyma in the roots and shoots, thus working synergistically with AtNRT1.5 to control long-distance NO₃⁻ transport.

The NRT2 family are high-affinity NO₃⁻ transporters comprising NO₃⁻ inducible and constitutively expressed members [3]. The
Results

A commonly accepted, extensively used and well documented method for the determination of genes that share a common evolutionary ancestor (orthologues, paralogues) across two genomes is the reciprocal best hit (RBH) [28,29]. The RBH approach assumes that orthologous sequences hit to each other as the best scoring hit in a pairwise search between two genomes. There are however, limitations in the RBH method; it does not take paralogy into account and the highest scoring protein reported by BLAST is often not the nearest phylogenetic neighbour (high false negative rate) [30]. In this study we elected to use the RBH method over other orthology detection strategies due to its high stringency and success in identifying orthologues with a low false positive rate [31,32]. To overcome the above mentioned shortcomings, a modified RBH, not restricted to the top hits alone and extended to include steps of refinement and validation was used (see Materials and Methods).

NRT1 family

As the genetic sequence(s) or protein motif(s) that separate the NRT1 genes from the PTR genes (the protein products of which transport peptides) are unknown, the analysis here was limited to determining putative grass orthologues of eight functionally characterised NRT1 genes (Table S1). Results from our analyses identified grass orthologues of AtNRT1.2, AtNRT1.3 and AtNRT1.4. However, for the remaining five members, a lack of resolution compounded by the analysis of only a subset of the 53 NRT1(PTR) genes and the complexity of ancestral events, rendered clear orthologous relationships between dicots and monocots, unattainable. RBH BLASTp E values in both forward and reverse direction for all eight NRT1 genes was less than \(10^{-159}\) with AtNRT1.1 – AtNRT1.3 and AtNRT1.8 showing greater sequence similarity to their grass counterparts (E values \(\leq 2 \times 10^{-159}\) = 63% to 65% sequence identity across \(\approx 85\%\) of the query sequence) than AtNRT1.6 and AtNRT1.7 (\(\approx 50\%\) sequence identity across \(\approx 80\%\) of the query sequence). Furthermore, RBH analyses returned near identical hits to the grass genomes for AtNRT1.5 and AtNRT1.8 as well as for AtNRT1.6 and AtNRT1.7. Graphical representation of the dicot – monocot RBH analysis can be seen in Figure S1. All-against-all RBH results for the NRT1 family is illustrated in Figure 1 (A–F).

The phylogenetic tree of the eight Arabidopsis NRT1 genes and their identified grass homologues (Figure 2) depicts these findings clearly. For AtNRT1.1, there is one gene in poplar; in the grasses, however, there are three clades of closely related NRT1.1-like genes all containing at least one representative from each grass species (subclade 3 has an extra Brachypodium and maize gene). As the genes fall into three subclades, it is likely that gene duplication events gave rise to the three groups after the dicot-monocot split. For AtNRT1.2, AtNRT1.3 and AtNRT1.4 we see a much clearer picture, all cluster with a single clade of grass genes containing at least one member from the four grass species (rice and maize have an extra representative in the 1.3 and 1.4 clade respectively). Notably, poplar has no AtNRT1.4-like gene. AtNRT1.5 and AtNRT1.8 sit together with two and one poplar orthologues respectively and branch off from the main tree with two separate but complete clades of grass genes. Due to a slightly higher dicot – monocot RBH score for 1.5 over 1.8, the grass members of these two clades were named 1.5A and 1.5B. Upon close scrutiny of this branch, we found that the grass 1.5B clade had a nearer phylogenetic neighbour (a PTR gene – At5g19640) even though BLAST result scores were higher for 1.5 and 1.8. Since both clades sit equidistant from AtNRT1.5 (and neither more closely related to AtNRT1.8) an unambiguous assignment of orthology is not possible. Similarly, AtNRT1.6 and AtNRT1.7 sit together on the tree but unlike 1.5 and 1.8, 1.6 and 1.7 share one poplar orthologue and branch off from the main tree with a single clade of grass genes missing both a maize and sorghum representative. Interestingly, the sorghum genome was found to contain a degraded pseudogene version that is related to NRT1.6 and NRT1.7, but this transcript is likely to be non-existent as no ESTs exist in any database.

Additionally, eight monocot sequences - three maize, three sorghum and one each from rice and Brachypodium - were initially included in the potential orthologue list after RBH analysis but rejected after refinement and validation. These sequences were not included as they were clearly seen to be nearest neighbours to three Arabidopsis PTR genes (Figure S2) and this was authenticated in RBH scores.

NRT2 family

The seven members of the NRT2 family in Arabidopsis possessed a greater sequence similarity with each other than did the members of the NRT1 family (Table S1). RBH analysis of the
NRT2 family returned identical results for AtNRT2.1 – AtNRT2.4 and AtNRT2.6 (E values = zero, 62% to 72% sequence identity across >90% of the query sequence). Results of the best hits for AtNRT2.5 were marginally lower (E values ranging between 10^{-162} and zero). Graphical representation of the dicot – monocot RBH analysis can be seen in Figure S1. All-against-all RBH results for the NRT2 family is illustrated in Figure 3.

Results from analyses on the NRT2 family painted a particularly interesting picture, exclusive of AtNRT2.5, the grass NRT2 genes sit entirely separate on the phylogenetic tree from the Arabidopsis NRT2 genes (Figure 4). Three poplar sequences (named PtNRT2.4A, PtNRT2.4B and PtNRT2.4C due to a higher dicot – monocot RBH score with AtNRT2.4) cluster with the Arabidopsis sequences suggesting that the NRT2 genes developed primarily following the divergence of the monocots and dicots. Thus identification of a clear grass orthologues was only achieved for AtNRT2.5. A close investigation of the genomic localisation of these genes revealed clustering of the genes somewhat reminiscent of genes involved in disease resistance. AtNRT2.1 and AtNRT2.2 are neighbouring genes in opposing orientation, AtNRT2.3 and AtNRT2.4 are tandem repeats and AtNRT2.6 (which peculiarly has the highest pairwise sequence similarity over all pairwise matches in the group, to AtNRT2.3 – 89%) was located on a completely separate chromosome to the others. Correspondingly, of the twelve related sequences in grass (2.1 – 2.4 in Figure 4), eleven are located in close proximity to another in their respective genomes. Maize has three genes related to this NRT2 branch, two of which are closely located on chromosome 4, the other gene being located on a separate chromosome (Figure S3). Similarly, sorghum has two closely located NRT2 genes with a third related gene located one gene to the upstream side of the closely located pair; this third related gene clusters in the phylogenetic tree with the third NRT2 gene from maize. Rice has a similar pair of closely located NRT2 genes (actually giving rise to the same protein sequence), while Brachypodium has two sets of closely located NRT2 genes which are immediately adjacent to each other. Interestingly, the pair of NRT2 genes found in each of the grass species is generally separated by a non-NRT2 gene. These non-NRT2 genes do not share any sequence similarity with one another, nor to any other functionally characterised proteins in the databases (BLASTx search against NCBI). The pairs of NRT2 genes in the grass species may have similar function to AtNRT2.1 and AtNRT2.2, although confirmation of this would require proper functional analysis. Analysis of the gene structure of the NRT2 genes showed that the dicot NRT2 genes all contain both exons and introns, while none of the grass NRT2 genes contain introns (Table S2); this would indicate an ancient divergence of the members of the NRT2 gene family.

There is one distinct NRT2.5-like gene in each grass genome; poplar, however, has two copies. There are no NRT2.7-like genes in any of the grass genomes, or in poplar. AtNRT2.7 is the most diverged of all the NRT2 sequences.

NRT3 family

The NRT3 family in Arabidopsis contains two members, AtNRT3.1 and AtNRT3.2. These genes are not NO₃⁻ transporters, but have been shown to be necessary for NO₃⁻ transport through interaction with other NRT2 transporters. The grass genomes were analysed for orthologues to both of these Arabidopsis NRT3 genes.

Genome analysis revealed the NRT3 family is best represented by individual splice forms. For NRT3.2 three splice variants have been predicted giving rise to two different amino acid sequences (www.arabidopsis.org, TAIR Acc# AT4G24730) (Figure S4 and
Figure 2. Phylogenetic relationship of the NRT1 family. Unrooted Neighbour-joining tree of NRT1 transporters in Arabidopsis (black), poplar (blue) and 4 grass species: rice (orange), sorghum (green), maize (red) and *Brachypodium* (purple). Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.2 estimated amino acid substitution per residue.
doi:10.1371/journal.pone.0015289.g002

Figure 3. NRT2 reciprocal BLAST polygons. Reciprocal best hits are connected by black lines, for Arabidopsis (black), maize (red), *Brachypodium* (purple), poplar (blue), sorghum (green) and rice (orange). Results are depicted for (A) AtNRT2.1 or AtNRT2.2 or AtNRT2.3 or AtNRT2.4 or AtNRT2.6 and (B) AtNRT2.5.
doi:10.1371/journal.pone.0015289.g003
In comparison to the 443 aa long protein encoded by the longest splice form (At4G24730.1, hereafter referred to as AtNRT3.2), the proteins translated from the other two splice forms (At4G24730.2 and At4G24730.3, which will be referred to as NRT3.2SF2/3) are 311aa in length, the first 281 aa being identical with that of the longest splice form. Furthermore, there is evidence (NCBI Acc# DQ492237) for another splice product from this locus, which gives rise to a protein of 209 aa overlapping with the C-terminal half of the longest splice form. This protein sequence (referred to here as NRT3.2CT) was described and compared to NRT3.1 by Okamoto et al [20], revealing that the two proteins share 61% amino acid sequence identity.

When the protein sequence of At4G24730.1 was used in BLAST searches it became apparent that in the other species in this study the N- and C terminal halves of At4G24730.1 are coded for by two distinct genes. Consequently, it was decided that the protein sequences of NRT3.2SF2/3 and NRT3.2CT would be used separately for orthology searches. Notably, three splice forms have been predicted for NRT3.1 but all three of these translate into the same protein (http://arabidopsis.org/servlets/TairObject?type = gene&name = AT5G50200.3).

The results from RBH analysis were the same for both AtNRT3.1 and AtNRT3.2CT, identifying a pair of co-orthologous genes in maize, sorghum and rice and a tandem repeat in Brachypodium (Figure S1). Interestingly, poplar has three representative homologues of AtNRT3.1/3.2CT. BLAST E-values in both forward and reverse direction ranged between 10^{-27} and 10^{-32}, with sequence identities ranging from 45% to 52% across 66% to 74% of the query sequence for AtNRT3.1 and ~41% across ~85% of the query sequence for AtNRT3.2CT (Graphical representation of the dicot – monocot RBH analysis can be seen in Figure S1). 

NRT Genes in Grass Species

![Figure 4. Phylogenetic relationship of the NRT2 family.](image)

Unrooted Neighbour-joining tree of NRT2 transporters in Arabidopsis (black), poplar (blue) and 4 grass species: rice (orange), sorghum (green), maize (red) and Brachypodium (purple). Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.08 estimated amino acid substitution per residue.

doi:10.1371/journal.pone.0015289.g004
second best hits in the forward direction which were reciprocal in reverse for both maize (GRMZM2G337128_T01) and sorghum (Sb02g024380.1), but not in rice. E values, sequence identity and alignment length for these hits were found to be just below those of the top hits (E value $10^{-117}$, sequence identity $≈62\%$ and alignment length of $≈99\%$). Although considered noteworthy, the maize and sorghum sequences were not used in any subsequent analysis. Figure 5B shows RBH results for AtNRT3.2 and AtNRT3.2SF2/3 and Figure 7 indicates the phylogenetic relationship of these protein sequences.

An attempt was made to explain our finding that At4G24730.1 appeared to be a fusion of what seemed to be two distinct genes in other species. This was done by comparing the genomic organisation of the NRT3 loci in Arabidopsis lyrata and in a further analysis.

Figure 5. NRT3 reciprocal BLAST polygons. Reciprocal best hits are connected by black lines, for Arabidopsis (black), maize (red), Brachypodium (purple), poplar (blue), sorghum (green) and rice (orange). Results are depicted for (A) AtNRT3.1 or AtNRT3.2CT and (B) AtNRT3.2 or AtNRT3.2SF2/3. doi:10.1371/journal.pone.0015289.g005

Figure 6. Phylogenetic relationship of the NRT3.1 or 3.2CT family. Unrooted Neighbour-joining tree of NRT3.1 or 3.2CT family in Arabidopsis (black), poplar (blue) and 4 grass species: rice (orange), sorghum (green), maize (red) and Brachypodium (purple). Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.1 estimated amino acid substitution per residue. doi:10.1371/journal.pone.0015289.g006
four dicot species each with a fully sequenced genome. As can be seen from Figure 8, only in Arabidopsis species is the organisation such that a di-cistronic mRNA can give rise to a fusion product through alternative splicing. In Ricinus communis, Vitis vinifera and Poplar trichocarpa the NRT3.2SF2/3 gene and the NRT3.2CT gene are encoded on the opposite strand of the double helix, and no additional separate NRT3.1-like gene could be found. The exception is Medicago trunculata which appears to lack any NRT3.2CT (or NRT3.1) orthologues in the genome at all. This organisation, with the genes being encoded on opposite strands in direct genomic vicinity, is true also for Canica papaya, Cucumis sativus, and Manihot esculenta (data not shown). These results clearly indicate that the genomic organisation found in Arabidopsis may be the exception rather than the rule for dicots.

Discussion

The purpose of this study was to determine the cereal orthologues of the characterised Arabidopsis NRT genes. Several studies have considered the NRT genes in grasses, but have resulted in some confusion, not least in the nomenclature ascribed to the various NRT genes identified. Much of this confusion was due, presumably, to the unavailability of fully sequenced grass genomes leading, for example, to the cloning of orthologues using degenerate primers. For instance, Lin et al [33] cloned an NRT gene in rice which was referred to as OsNRT1.1 by Tsay et al [3]. This gene is part of the NRT1/PTR family, but it is not a likely orthologue for AtNRT1.1. From the present analysis, and that of Tsay et al [3], it is evident that the rice genes Os08g05910 and Os10g06000 are much more likely to be the orthologues of AtNRT1.1. Recent work indicates that AtNRT1.1 (CHL1) may function as a nitrogen sensor [7]; should this be the case, the grasses may have more finely tuned nitrogen sensing, root tissue specific nitrogen sensors, a nitrogen sensor in the shoot tissue (depending on where these genes are expressed), or else may have a very different sensing mechanism altogether. Analysis of the NCBI Unigene database for Oryza sativa Build #80 (http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=4530) shows that OsNRT1.1A and OsNRT1.1B are expressed throughout the rice plant, however OsNRT1.1A is predominantly expressed in the root and is expressed more highly than OsNRT1.1B. This indicates that at least two genes potentially fill the same functional role of AtNRT1.1 in grasses. Conversely, the grass genomes lack certain NRT genes that have been characterised in Arabidopsis. Our analysis reveals that the Brachypodium and rice genomes contain only one protein similar to AtNRT1.7 and AtNRT1.6, or no orthologues in the case of maize and sorghum. AtNRT1.6 appears to be involved in transport of NO3− from the maternal tissue to the developing embryo [13] and AtNRT1.7 plays a role in the remobilisation of NO3− from the older leaves [14]. Again analysis of the NCBI Unigene database for Oryza sativa Build #80 shows OsNRT1.7 is expressed in flower, seed and panicle, perhaps indicating that NRT1.7 in the grasses fills a similar functional role to AtNRT1.6. Whether the proposed long distance NO3− transport functions of the AtNRT1.5 [12] and AtNRT1.8 [15] genes are indeed carried out by their closest grass homologues (NRT1.5A and NRT1.5B) or whether the grasses employ different genes remains to be investigated. The NCBI Unigene database

Figure 7. Phylogenetic relationship of the NRT3.2SF2/3 family. Unrooted Neighbour-joining tree of NRT3.2SF2/3 family in Arabidopsis (black), poplar (blue) and 4 grass species: rice (orange), sorghum (green), maize (red) and Brachypodium (purple). Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.04 estimated amino acid substitution per residue. doi:10.1371/journal.pone.0015289.g007
indicates that OsNRT1.5A is expressed predominantly in root and panicle, but also in stem, leaf and flower. OsNRT1.5B was mostly expressed in panicle, but also flower and stem.

Perhaps the most obvious difference between the grasses and Arabidopsis lies in the structure of the NRT2 gene family. The presence of closely located NRT2 genes is reminiscent of the AtNRT2.1/AtNRT2.2 cluster in Arabidopsis and functional analysis of these genes may show similar function between the NRT2 gene clusters. However, with extra duplication in Brachypodium, for example, it will be interesting to identify the functional role played by each of the repeats. Since the NRT2 genes in the grasses lack an intron, it would appear that development of the NRT2 family occurred following the split between the dicots and monocots. It is possible that when they diverged from primitive dicots, the early monocots lost the NRT2 intron(s) whilst the dicots retained the intron(s). However, determination of the development of the NRT2 gene family in plants requires significant further analysis including the genomes of species all along the evolutionary tree, especially in the case of genes with multiple candidates (e.g. NRT1.1). We have attempted to take this issue into account by of any common protein motifs or sequence differences that separate the dicot and grass NRT2s may provide important structure/function information of value in guiding biotechnological approaches to improving NO\textsubscript{3}\textsuperscript{-} transport in plants. Although not included in our full analysis (since the genome has not been fully sequenced) the four barley (Hordeum vulgare) NRT2s [35,36] group with the two pairs of Brachypodium NRT2 genes (BdNRT2.1/ BdNRT2.2 and BdNRT2.3/BdNRT2.4) (Figure S6); consequently, the functional analysis of these genes provides little insight into the potential function of the NRT2 genes from the other grasses except to indicate the cereal NRT2 transporters likely mediate high affinity NO\textsubscript{3}\textsuperscript{-} transport as the barley NRT2 transporters do.

Our analysis of the NRT3 family also indicates there may be fundamental differences between Arabidopsis and the grasses indicative of unique evolutionary events. Of particular importance, the significance of the fusion of the AtNRT3.2SF1/3 gene with the AtNRT3.1CT gene in Arabidopsis compared with the way this protein interacts with the high affinity NRT2 transporters in grasses remains unknown. It will be interesting to determine whether the AtNRT3.2SF2/3 orthologues in the grasses are involved in NO\textsubscript{3}\textsuperscript{-} transport. In barley (Hordeum vulgare), three NRT3 genes (NAR2.1, NAR2.2 and NAR2.3) have been identified [37], all of which are similar to the AtNRT3.1 or AtNRT3.2CT genes (Figure S7). It remains to be seen whether or not the NRT3 genes identified in this study play a functional role similar to that of the barley orthologues.

To assist future research we have developed a nomenclature for grass NRT genes (Table 1). Without functional characterisation of the grass NRT orthologues it is difficult to determine which grass orthologue will have similar function to a given Arabidopsis gene, especially in the case of genes with multiple candidates (e.g. NRT1.1). We have attempted to take this issue into account by

---

**Figure 8. Genomic organisation of the NRT3 genes in the dicots.** Results are depicted for Arabidopsis thaliana, Arabidopsis lyrata, Ricinus communis, Poplar trichocarpa, Vitis vinifera and Medicago trunculata. The genes represented by a similar colour are reciprocal top BLAST hits. Genes are labelled with Arabidopsis thaliana nomenclature: AtNRT3.1, AtNRT3.2SF1, AtNRT3.2SF2/3 and AtNRT3.2CT. Chromosome or scaffold numbers for each species are provided. Illustrations are not to scale. doi:10.1371/journal.pone.0015289.g008
| Arabidopsis thaliana | Populus trichocarpa | Oryza sativa | Zea mays | Sorghum bicolor | Brachypodium distachyon |
|----------------------|---------------------|-------------|--------|----------------|------------------------|
| Symbol               | TAIR ID             | Symbol      | JGI ID | Symbol          | Brachypodium.orgID      |
| AtNRT1.1             | AT1G12110           | PtNRT1.1    | POPTR_0003s11080.1 | SbNRT1.1A          | Bradi3g16670.1          |
| AtNRT1.2             | AT1G69850           | PtNRT1.2    | POPTR_0001s7250.1 | SbNRT1.1B          | Bradi3g33030.1          |
| AtNRT1.3             | AT1G21670           | PtNRT1.3A   | POPTR_0001s15580.1 | SbNRT1.1C          | Bradi3g33040.1          |
|                     |                     | PtNRT1.3B   | POPTR_0001s21630.1 |                    |                        |
| AtNRT1.5             | AT1G32450           | PtNRT1.5A   | POPTR_0001s02810.1 | SbNRT1.5A          | Bradi1g78330.1          |
| AtNRT1.6             | AT1G27080           | PtNRT1.5B   | POPTR_0001s08710.1 | SbNRT1.5B          |                        |
| AtNRT1.7             | AT1G69870           | PtNRT1.7    | POPTR_0001s116700.1 |                  |                        |
| AtNRT1.8             | AT1G2450            | PtNRT1.8    | POPTR_0001s0810.1  |                  |                        |
| AtNRT2.1             | AT1G12940           | PtNRT2.1A   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.2             | AT1G3000            | PtNRT2.1B   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.3             | AT1G8010            | PtNRT2.1C   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.4             | AT1G6780            | PtNRT2.2A   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.5             | AT1G6790            | PtNRT2.2B   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.6             | AT1G3470            | PtNRT2.2C   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.7             | AT1G3570            | PtNRT2.3A   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.8             | AT1G3570            | PtNRT2.3B   | POPTR_0001s1520.1  |                  |                        |
| AtNRT2.9             | AT1G3570            | PtNRT2.3C   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.1             | AT1G3570            | PtNRT3.1A   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.2             | AT1G3570            | PtNRT3.1B   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.3             | AT1G3570            | PtNRT3.1C   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.4             | AT1G3570            | PtNRT3.2A   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.5             | AT1G3570            | PtNRT3.2B   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.6             | AT1G3570            | PtNRT3.2C   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.7             | AT1G3570            | PtNRT3.3A   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.8             | AT1G3570            | PtNRT3.3B   | POPTR_0001s1520.1  |                  |                        |
| AtNRT3.9             | AT1G3570            | PtNRT3.3C   | POPTR_0001s1520.1  |                  |                        |

*GenBank identifier.

doi:10.1371/journal.pone.0015289.t001
naming all potential co-orthologues A, B, C, etc. However, in the case of the \textit{NRT2} gene family where dicot and grass genes do not cluster together in the phylogenetic tree, this approach becomes problematic. Therefore, we named family members \textit{NRT2.1}, \textit{NRT2.2}, etc, despite the fact that a grass \textit{NRT2.1} may not share a functional role with \textit{AtNRT2.1}.

In conclusion, the present analysis of the \textit{NRT} gene families in Arabidopsis and the grasses has revealed some striking differences in gene family structure. Important questions about the evolution of \textit{NRT} transporters in plants and, significantly, about the suitability of Arabidopsis as a model for \(\text{NO}_3^-\) transport in the grasses have also been posed. With the current exponential increase in the availability of molecular genetic resources for cereal crop plants it appears likely that the relevance of Arabidopsis research will decline. This analysis provides a framework for future studies of \(\text{NO}_3^-\) transporters and transport in the grasses, and potentially will guide strategies for improvement of NUE in cereal species through genetic manipulation of the \textit{NRT} genes.

Materials and Methods

Sequences and Databases

DNA and amino acid sequences of 17 \textit{AtNRT} family members (Table S1) were retrieved from TAIR (The Arabidopsis Information Resource), \textit{Arabidopsis thaliana} genome annotation database release 9 (http://www.arabidopsis.org/). The complete database of predicted amino acid sequences from \textit{Arabidopsis thaliana} and \textit{Populus trichocarpa} as well as from four monocot species; \textit{Zea mays}, \textit{Oryza sativa}, \textit{Brachypodium distachyon} and \textit{Sorghum bicolor} were downloaded from public databases (Table S3).

Bioinformatics

Identification of homologues. Identification of homologues was based primarily on sequence similarity between the 17 Arabidopsis \textit{NRT} genes and the predicted amino acid sequences of the four above mentioned monocots (maize, rice, \textit{Brachypodium} and sorghum). This was achieved by using BLASTp, standalone version 2.2.21 [38,39] and a modified reciprocal best hit (RBH) approach [28,29]. In brief, BLAST searches were performed which queried the set of 17 At\textit{NRT} protein sequences against each of the poplar/monocot databases in a pairwise manner (forward BLAST). Following this and deviating from the standard method, the top ‘cluster of best hits’ returned from each pairwise forward BLAST (based on E-value and sorted by score) was then used as queries in subsequent BLAST searches against the Arabidopsis database (reverse BLAST). Proteins from the reverse BLAST that returned as one of their best hits the original query protein from the forward BLAST (the relevant At\textit{NRT}), were then selected for further evaluation as homologues. For all forward and reverse BLASTp searches, E value cut-off was set to 1E-20 (>1E-20), output was set to tabular (-m 8), all other parameters were left as default. The aforementioned deviation introduced to the method used in this study was made in an attempt to resolve nearest phylogenetic neighbour and paralogy shortcomings when using BLAST and the strict RBH approach to identify homologues between species [30]. The list of homologues was then refined by removal of those candidates not specifically related to the At\textit{NRT}s of interest (Table S4). This was achieved via manual inspection of multiple sequence alignments and their corresponding trees. Throughout the analyses all splice variants of all identified homologues accepted for further analysis were used in subsequent rounds of RBH (eleven candidate homologues were found to possess splice variants, see Table S1). However, only the one member with the longest protein sequence from each splice variant group was used to build trees.

The remaining candidates were then used in all-against-all rounds of RBH analysis between the four monocot sequence databases. The RBH approach and criteria used for the original monocot to dicot search, as described above, was again applied here.

Multiple Sequence Alignment and Tree building. The RBH obtained sequences for \textit{NRT1}, were aligned by MAFFT version 6.240 using the L-INS-I method with associated default parameters at [http://align.genome.jp/mafft/] [40] and manual editing, for \textit{NRT2}, 3.1 and 3.2 by T-coffee::Advanced [41] using the server at the Swiss Institute for Bioinformatics, pairwise method set to ‘best_pair’, multiple method set to ‘mafft_msa’ and default for all other parameters (http://tcoffee.vital-it.ch/). Trees were built using various programs of the Phylogenetic Interference Package (PHYLIP, 3.63 J. Felsenstein, http://evolution.genetics.washington.edu/phylip.html). One thousand bootstrap datasets were generated with SEQBOOT to estimate the confidence limits of nodes. Protein distance matrices were calculated with PROTDIST using the PMB model [42]. Trees were generated with WEIGHBOR [43], and the majority rule consensus tree was generated by CONSENSE (default method used ‘Majority rule (extended)’ >50%). Trees were visualized using the software Genious v4.8 (Biomatters Ltd, New Zealand, http://www.genesius.com/). Information on genome organisation was obtained from the Phytozome (www.phytozome.org) and TAIR.

Supporting Information

Figure S1 At\textit{NRT1}, 2 and 3 reciprocal BLASTp results. Depiction of forward and reverse reciprocal BLAST\textit{p} hits is provided for (A) poplar, (B) \textit{Brachypodium}, (C) maize, (D) sorghum and (E) rice. BLAST\textit{p} hits of equal e-value in forward and reverse directions are depicted as forward facing and reverse facing arrows respectively. The colour code of all arrows represents the order of hits, as returned by the BLAST\textit{p} program, first best hit (red), second best hit (blue), third best hit (orange), fourth best hit (green), fifth best hit (black). Truncated arrows indicate a minor drop in the e-value score between hits. Also coded are the gene accession numbers; an asterisk (*) indicates the presence of alternative splice forms all scoring identical BLAST\textit{p} results, an upward facing arrow head (↑) indicates a tandem repeat with identical BLAST\textit{p} score, red accessions indicate neighbouring genes in opposing orientation with identical BLAST\textit{p} scores and blue accessions indicate genes in close proximity to each other with identical BLAST\textit{p} scores.

Figure S2 Phylogenetic relationship of potential grass PTR transporters orthologues to At\textit{NRT1.6} and At\textit{NRT1.7}. Unrooted Neighbour-joining tree of \textit{NRT1} transporters in Arabidopsis (black), poplar (blue) and 4 grass species: rice (orange), sorghum (green), maize (red) and \textit{Brachypodium} (purple). Highlighted in the box are the three closest PTR homologues to At\textit{NRT1.6} and At\textit{NRT1.7} (AT1G18880, AT3G47960 and AT5G62680) as well as orthologous grass PTR transporters (all in brown). This figure provides rationale for exclusion of grass transporters (brown) as orthologues to At\textit{NRT1.6} and At\textit{NRT1.7}. Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.2 estimated amino acid substitution per residue.
Figure S3  Conservation of closely localised NRT2 genes. Gene identifiers and chromosome number are provided for NRT2 genes (red) for (A) Arabidopsis, (B) poplar, (C) rice, (D) Brachypodium, (E) maize and (F) sorghum. Sorghum has a third closely localised NRT2 gene (blue). Also depicted are non-NRT2 genes (black) between NRT2 genes. Illustrations are not to scale. (TIF)

Figure S4  Depiction of AtNRT3 genes. Schematic of the AtNRT3.2 locus (AT4G24730) from TAIR 9 gBrowse (http://gbrowse.arabidopsis.org/). Represented are chromosome, BAC, locus, gene models from TAIR 8 and TAIR 9 and cDNA details. (TIF)

Figure S5  Alignment of AtNRT3 proteins. Proteins included are AtNRT3.1 and AtNRT3.2CT described previously Okamoto et al [20] and the new versions identified in TAIR9 (AtNRT3.2SF1 AtNRT3.2SF2 and AtNRT3.2SF3). Colour scheme for residue similarity (letter/background): black/white – non similar; blue/light blue – conservative; black/green – block of similar residues; red/yellow – identical; and green/white – weakly similar. Gene identifiers are provided in brackets. Refer to Figure 7 for genomic organisation of Arabidopsis genes. (TIF)

Figure S6  Phylogenetic relationship of the NRT2 family including barley family members. Unrooted Neighbour-joining tree of NRT2 transporters in Arabidopsis (black), poplar (blue) and 3 grass species: rice (orange), sorghum (green), maize (red), Brachypodium (purple) and barley (brown). The four barley members include HvNRT2.1 (HVU34198), HvNRT2.2 (HVU34290), HvNRT2.3 (AF091115) and HvNRT2.4 (AF091116). Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.08 estimated amino acid substitution per residue. (TIF)

Figure S7  Phylogenetic relationship of the NRT3.1 or NRT3.2CT family including barley family members. Unrooted Neighbour-joining tree of NRT3.1 or 3.2CT family in Arabidopsis (black), poplar (blue) and 3 grass species: rice (orange), sorghum (green), maize (red), Brachypodium (purple) and barley (brown). The three barley members include HvNRT3.1A (HVU33448), HvNRT3.1B (HVU33449) and HvNRT3.1C (HVU33450). Bootstrap values from 1,000 replicates were used to estimate the confidence limits of the nodes. The scale bar represents a 0.08 estimated amino acid substitution per residue. (TIF)

References
1. Raun WR, Johnson GV (1999) Improving nitrogen use efficiency for cereal production. Agron J 91: 357–363.
2. Sylvester-Bradley R, Kindred DR (2009) Analysing nitrogen responses of cereals to prioritize routes to the improvement of nitrogen use efficiency. J Exp Bot 60: 1939–1951.
3. Tsay Y-F, Chiu C-C, Tiai C-B, Ho C-H, Hsu P-K (2007) Nitrate transporters and peptide transporters. FEBS Lett 581(12): 2290–2300.
4. Wang R, Liu D, Crawford NM (1998) The Arabidopsis CHL1 protein plays a major role in high-affinity nitrate uptake. Plant Physiol 119(25): 15134–15139.
5. Liu K-H, Huang C-Y, Tsay Y-F (1999) CHL1 is a dual-affinity nitrate transporter of Arabidopsis involved in multiple phases of nitrate uptake. Plant Cell 11(5): 865–874.
6. Liu K-H, Tsay Y-F (2003) Switching between the two action modes of the dual-affinity nitrate transporter CHL1 by phosphorylation. EMBO J 22(5): 1005–1013.
7. Ho C-H, Lin S-H, Hu H-C, Tsay Y-F (2009) CHL1 functions as a nitrate sensor in plants. Cell 138(6): 1184–1194.
8. Huang NC, Liu KH, Ho LJ, Tsay YF (1999) Cloning and functional characterization of an Arabidopsis nitrate transporter gene that encodes a constitutive component of low-affinity uptake. Plant Cell 11(8): 1391–1392.
9. Ores M, Krapp A, Daniel-Vedele F (2002) Analysis of the NRT2 nitrate transporter family in Arabidopsis. Structure and gene expression. Plant Physiol 129(2): 889–896.
10. Okamoto M, Vidmar JJ, Glass ADM (2003) Regulation of NRT1 and NRT2 gene families of Arabidopsis thaliana: responses to nitrate provision. Plant Cell Physiol 44(3): 304–317.
11. Chin C-C, Lin C-S, Hoa A-P, Su R-C, Lin H-L, et al. (2004) Mutation of a nitrate transporter, AtNRT1:4, results in a reduced petide nitrate content and altered leaf development. Plant Cell Physiol 45(9): 1139–1148.
12. Lin S-H, Kuo H-F, Canivenc G, Lin C-S, Lepeit M, et al. (2008) Mutation of the Arabidopsis NRT1.5 nitrate transporter causes defective root-to-shoot nitrate transport. Plant Cell 20(9): 2514–2528.
13. Almeida A, Lin SH, Tsay YF (2008) Characterization of the Arabidopsis nitrate transporter NRT1.6 reveals a role of nitrate in early embryo development. Plant Cell 20(12): 3289–3299.
14. Fan S-C, Lin C-S, Hsu P-K, Lin S-H, Tsay Y-F (2009) The Arabidopsis nitrate transporter NRT1.7, expressed in phloem, is responsible for source-to-sink remobilization of nitrate. Plant Cell 21(9): 2750–2761.
15. Li J-Y, Fu Y-L, Pike SM, Bao J, Tian W, et al. (2010) The Arabidopsis nitrate transporter NRT1.8 functions in nitrate removal from the xylem sap and mediates cadmium tolerance. Plant Cell 11:073242.
16. Li W, Wang Y, Okamoto M, Crawford NM, Siddiqui MY, et al. (2007) Dissection of the AtNRT2.1/AtNRT2.2 inducible high-affinity nitrate transporter gene cluster. Plant Physiol 143(1): 425–433.
17. Wirth J, Chopin F, Santoni Yr, Viennois GI, Tillard P, et al. (2007) Regulation of root nitrate uptake at the NRT2.1 protein level in Arabidopsis thaliana. J Biol Chem 282(32): 23541–23552.

18. Chopin F, Orsel M, Dorbe M-F, Chardon F, Truong H-N, et al. (2007) The Arabidopsis ATNRT2.7 nitrate transporter controls nitrate content in seeds. Plant Cell 19(3): 1509–1602.

19. Orsel M, Chopin F, Leleu O, Smith SJ, Krapp A, et al. (2006) Characterization of a two-component high-affinity nitrate uptake system in Arabidopsis. Physiology and protein-protein interaction. Plant Physiol 144(3): 1304–1317.

20. Okamoto M, Kumar A, Li W, Wang Y, Siddiqi MY, et al. (2006) High-affinity nitrate transport in roots of Arabidopsis depends on expression of the NRT2.1 gene. J Biol Chem 282(32): 23541–23552.

21. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296: 92–100.

22. Yu J, Hu S, Wang J, Wong GK-S, Li S, et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296: 79–92.

23. The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463(7282): 763–768.

24. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956): 1112–1115.

25. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457(7229): 551–556.

26. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, et al. (2006) The genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science 313(5793): 652–656.

27. Bruno WJ, Socci ND, Halpern AL (2000) Weighted neighbour joining: a likelihood-based approach to distance-based phylogeny reconstruction. Mol Biol Evol 17: 189–197.
