Plett, Darren Craig; Toubia, John; Garnett, Trevor Paul; Tester, Mark Alfred; Kaiser, Brent Norman; Baumann, Ute
Dichotomy in the NRT gene families of dicots and grass species, PLoS One, 2010; 5(12):e15289

© 2010 Plett et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PERMISSIONS

http://www.plosone.org/static/license

Open-Access License

No Permission Required

PLOS applies the Creative Commons Attribution License (CCAL) to all works we publish (read the human-readable summary or the full license legal code). Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLOS journals, so long as the original authors and source are cited. No permission is required from the authors or the publishers.

In most cases, appropriate attribution can be provided by simply citing the original article (e.g., Kaltenbach LS et al. (2007) Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. PLOS Genet 3(5): e82. doi:10.1371/journal.pgen.0030082). If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then please indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared. For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

This broad license was developed to facilitate open access to, and free use of, original works of all types. Applying this standard license to your own work will ensure your right to make your work freely and openly available. Learn more about open access. For queries about the license, please contact us.

4th July 2013

http://hdl.handle.net/2440/64139
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
The best characterised members are AtNRT2.1 and AtNRT2.2, which are located next to each other on chromosome 1 and appear to encode proteins with similar function [16]. AtNRT2.1 seems to be more crucial for NO\textsubscript{3} uptake and is expressed in the root cortex and epidermis [17]. Both AtNRT2.1 and AtNRT2.2, however, are inducible by provision of NO\textsubscript{3}\textsuperscript{−} to NO\textsubscript{3}\textsuperscript{−} starved plants [10], and compensate for one another in that expression of either increases when the other is reduced [16]. Little is known about AtNRT2.3 other than its expression may increase and decrease in cycles over the life cycle in the roots and shoots (mostly shoot) [9,10]. AtNRT2.4 is expressed predominantly in the root, and expression appears to decrease following exposure of plants to NO\textsubscript{3}\textsuperscript{−} [9,10]. Similarly, AtNRT2.5 is expressed in the root and shoot (mostly root) and is repressed by the provision of NO\textsubscript{3}\textsuperscript{−} [9,10]. The expression of AtNRT2.6 remains relatively unchanged in roots and shoots (mostly root) following exposure of plants to NO\textsubscript{3}\textsuperscript{−} [9,10]. AtNRT2.7 appears to have a role in storage of NO\textsubscript{3}\textsuperscript{−} in seeds [18].

The NRT3 genes in Arabidopsis play a role in NO\textsubscript{3}\textsuperscript{−} transport through regulating the activity of NRT2 genes, but are not themselves transporters [19,20]. The two NRT3 genes appear to be closely related, but NRT3.1 (NAR2.1) appears to play the more significant role in high-affinity NO\textsubscript{3}\textsuperscript{−} uptake [20]. Although the recent annotation of the Arabidopsis genome has indicated that AtNRT3.2 gene is larger than originally published [20], the significance of this fact is unknown (http://www.arabidopsis.org/).

Here, bioinformatic analyses are presented of the NRT1, NRT2 and NRT3 gene families in the four fully sequenced grass genomes of rice [21,22,23], Brachypodium [24], maize [25] and sorghum [26]. Also included is an analysis of poplar as a further fully sequenced dicot species [27], with the purpose of strengthening observations made on the dichotomy between Arabidopsis and the grass species. The analyses were limited to fully sequenced genomes to ensure completeness and to increase the utility of the work for informing further research into NO\textsubscript{3}\textsuperscript{−} transporters in grass species. The evolution from a common ancestor of the four species studied is such that they provide a good indication of the diversity of genomes within the grass species; maize and sorghum are the most closely related, having diverged an estimated 12 million years ago [24]. Also, in order to clarify past identification of the relevant genes and to provide a standardised framework for future researchers, a nomenclature for the grass NRT genes is presented.

**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 $10^{-19}$ with AtNRT1.1 – AtNRT1.3 and AtNRT1.8 showing greater sequence similarity to their grass counterparts (E values $<2\times10^{-15}$ to $63\%$ sequence identity across $\approx85\%$ of the query sequence) than AtNRT1.6 and AtNRT1.7 ($\approx50\%$ sequence identity across $\approx80\%$ 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 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 NRT1.1 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 1. NRT1 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) AtNRT1.1, (B) AtNRT1.2, (C) AtNRT1.3, (D) AtNRT1.4, (E) AtNRT1.5 and AtNRT1.8 and (F) AtNRT1.7. doi:10.1371/journal.pone.0015289.g001

---

PLoS ONE | www.plosone.org 3 December 2010 | Volume 5 | Issue 12 | e15289
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). Closer investigation of the BLAST results to maize, rice and sorghum revealed a pattern; the genomic organisation of all these genes was well conserved between species. We postulate that either; a duplication event took place before the monocot – dicot split giving rise to co-orthologous pairs of genes of which one was eventually lost in Arabidopsis, or that, recent but separate duplication events have occurred in both the grass species and poplar, independent to Arabidopsis. RBH results for AtNRT3.1 and AtNRT3.2CT are illustrated in Figure 5A. Figure 6 shows the phylogenetic relationship of these protein sequences.

Similarly, AtNRT3.2 and AtNRT3.2SF2/3 produced the same RBH results, identifying a single NRT3 orthologue in maize, sorghum and rice and a tandem repeat in Brachypodium (Figure S1). BLAST E values in both forward and reverse direction ranged from $10^{-119}$ to $10^{-130}$, scoring ≈67% sequence identity. Alignment length differed between the two RBH analyses and was shown to be higher for NRT3.2SF2/3 (≈98%) compared with NRT3.2 (≈68%) (Graphical representation of the dicot – monocot RBH analysis can be seen in Figure S1). This difference can be explained by the fact that all BLAST hits landed within the aligned section between AtNRT3.2 and AtNRT3.2SF2/3 and never in the extended region of the former. For this reason, all subsequent investigations with respect to NRT3.2 were undertaken with only the AtNRT3.2SF2/3 splice form. Also similar to the NRT3.1 genes, close scrutiny of the AtNRT3.2SF2/3 RBH results revealed
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 $\approx 62\%$ and alignment length of $\approx 99\%$). Although considered noteworthy, the maize and sorghum sequences were not used in any subsequent analysis. Figure 5B shows RBH results for AtNRT3.2 or AtNRT3.2SF2/3 and Figure 7 indicates the phylogenetic relationship of these protein sequences.

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

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

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 Carica 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 Os10g40600 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 NO$_3^-$ from the maternal tissue to the developing embryo [13] and AtNRT1.7 plays a role in the remobilisation of NO$_3^-$ from the older leaves [14]. Again analysis of the NCBI Unigene database for Oryza sativa Build #80 shows that 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 NO$_3^-$ 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. NRT2 genes identified in this study play a functional role similar to that of the barley orthologues).

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 AtNRT3.2SF2/3 gene with the AtNRT3.1CT gene in Arabidopsis compared with the way this protein interacts with the high affinity NRT2 transporters 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$_3^-$ transport as the barley NRT2 transporters do.

Our analysis of the NRT3 family also indicates there may be 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$_3^-$ 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$_3^-$ transport as the barley NRT2 transporters do.

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
<table>
<thead>
<tr>
<th>Arabidopsis thaliana</th>
<th>Populus trichocarpa</th>
<th>Oryza sativa</th>
<th>Zea mays</th>
<th>Sorghum bicolor</th>
<th>Brachypodium distachyon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol</td>
<td>TAIR ID</td>
<td>Symbol</td>
<td>JGI ID</td>
<td>Symbol</td>
<td>MSU ID</td>
</tr>
<tr>
<td>AtNRT1.1</td>
<td>AT1G2110</td>
<td>PINRT1.1</td>
<td>POPTR_0003s11080.1</td>
<td>OsNRT1.1A</td>
<td>LOC_Os09g05910.1</td>
</tr>
<tr>
<td>AtNRT1.2</td>
<td>AT1G69850</td>
<td>PINRT1.2</td>
<td>POPTR_0012s07250.1</td>
<td>OsNRT1.2</td>
<td>LOC_Os09g38294.1</td>
</tr>
<tr>
<td>AtNRT1.3</td>
<td>AT3G21670</td>
<td>PINRT1.3A</td>
<td>POPTR_0014s15580.1</td>
<td>OsNRT1.3A</td>
<td>LOC_Os04g39030.1</td>
</tr>
<tr>
<td>AtNRT1.4</td>
<td>AT2G26690</td>
<td>N/A</td>
<td>N/A</td>
<td>OsNRT1.4</td>
<td>LOC_Os10g40600.1</td>
</tr>
<tr>
<td>AtNRT1.5</td>
<td>AT1G32450</td>
<td>PINRT1.5A</td>
<td>POPTR_0013s02810.1</td>
<td>OsNRT1.5A</td>
<td>LOC_Os04g39030.1</td>
</tr>
<tr>
<td>AtNRT1.6</td>
<td>AT1G27080</td>
<td>N/A</td>
<td>N/A</td>
<td>OsNRT1.6</td>
<td>LOC_Os09g38294.1</td>
</tr>
<tr>
<td>AtNRT1.7</td>
<td>AT1G69870</td>
<td>PINRT1.7</td>
<td>POPTR_0015s04370.1</td>
<td>OsNRT1.7</td>
<td>LOC_Os01g68510.1</td>
</tr>
<tr>
<td>AtNRT1.8</td>
<td>AT4G21680</td>
<td>PINRT1.8</td>
<td>POPTR_0014s17750.1</td>
<td>OsNRT1.8</td>
<td>LOC_Os01g68510.1</td>
</tr>
<tr>
<td>AtNRT2.1</td>
<td>AT1G08990</td>
<td>PINRT2.1A</td>
<td>POPTR_0009s01410.1</td>
<td>OsNRT2.1A</td>
<td>LOC_Os02g46460.1</td>
</tr>
<tr>
<td>AtNRT2.2</td>
<td>AT1G08100</td>
<td>PINRT2.1B</td>
<td>POPTR_0003s08710.1</td>
<td>OsNRT2.1B</td>
<td>LOC_Os02g48570.1</td>
</tr>
<tr>
<td>AtNRT2.3</td>
<td>AT5G60780</td>
<td>PINRT2.1C</td>
<td>POPTR_0014s02000.1</td>
<td>OsNRT2.1C</td>
<td>LOC_Os02g48570.1</td>
</tr>
<tr>
<td>AtNRT2.4</td>
<td>AT5G60770</td>
<td>PINRT2.1D</td>
<td>POPTR_0015s09290.1</td>
<td>OsNRT2.1D</td>
<td>LOC_Os02g48570.1</td>
</tr>
<tr>
<td>AtNRT2.5</td>
<td>AT1G2940</td>
<td>PINRT2.1E</td>
<td>POPTR_0015s09290.1</td>
<td>OsNRT2.1E</td>
<td>LOC_Os02g48570.1</td>
</tr>
<tr>
<td>AtNRT2.7</td>
<td>AT5G4570</td>
<td>N/A</td>
<td>N/A</td>
<td>OsNRT2.7</td>
<td>LOC_Os09g30820.1</td>
</tr>
<tr>
<td>AtNRT3.1</td>
<td>AT5G60200</td>
<td>PINRT3.1A</td>
<td>POPTR_0012s09110.1</td>
<td>OsNRT3.1A</td>
<td>LOC_Os09g30820.1</td>
</tr>
<tr>
<td>AtNRT3.2C</td>
<td>AT5G49237</td>
<td>PINRT3.1B</td>
<td>POPTR_0015s09660.1</td>
<td>OsNRT3.1B</td>
<td>LOC_Os09g30820.1</td>
</tr>
<tr>
<td>AtNRT3.2SF2/3</td>
<td>AT4G24730.2</td>
<td>PINRT3.1C</td>
<td>POPTR_0015s09660.1</td>
<td>OsNRT3.1C</td>
<td>LOC_Os09g30820.1</td>
</tr>
</tbody>
</table>

*GenBank Identifier.
doi:10.1371/journal.pone.0015289.t001
naming all potential co-orthologues A, B, C, etc. However, in the case of the 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 NRT2.1, NRT2.2, etc. despite the fact that a grass NRT2.1 may not share a functional role with AtNRT2.1.

In conclusion, the present analysis of the NRT gene families in Arabidopsis and the grasses has revealed some striking differences in gene family structure. Important questions about the evolution of NRT transporters in plants and, significantly, about the suitability of Arabidopsis as a model for NO\textsubscript{3}{−} transport in the grasses have also been posed. With the current exponential increase in the availability of molecular genetic resources for cereal crops it appears likely that the relevance of Arabidopsis research will decline. This analysis provides a framework for future studies of NO\textsubscript{3}{−} transporters and transport in the grasses, and potentially will guide strategies for improvement of NUE in cereal species through genetic manipulation of the NRT genes.

**Materials and Methods**

**Sequences and Databases**

DNA and amino acid sequences of 17 AtNRT family members (Table S1.) were retrieved from TAIR (The Arabidopsis Information Resource), Arabidopsis thaliana genome annotation database release 9 ([http://www.arabidopsis.org/](http://www.arabidopsis.org/)). The complete database of predicted amino acid sequences from Arabidopsis thaliana and Populus trichocarpa as well as from four monocot species; *Zea mays*, *Oryza sativa*, *Brachypodium distachyon* and *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 NRTs and the predicted amino acid sequences of the four above mentioned monocots (maize, rice, *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 AtNRT 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 AtNRT), were then selected for further evaluation as homologues. For all forward and reverse BLASTp searches, E value cut-off was set to 1E-20 (-e 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 AtNRTs 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 (even 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 NRT1, were aligned by MAFFT version 6:240 using the L-INS-I method with associated default parameters at [http://align.genome.jp/mafft/](http://align.genome.jp/mafft/) ([40]) and manual editing, for 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_pairwise’, multiple method set to ‘mafft_multi’ and default for all other parameters ([http://tcoffee.vital-it.ch/](http://tcoffee.vital-it.ch/)). Trees were built using various programs of the Phylogenetic Inference Package (PHYLIP, 3.63 [J. Felsenstein, [http://evolution.genetics.washington.edu/phylip.html](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 WEIGHBON ([43]), and the majority rule consensus tree was generated by CONSENSE (default method used ‘Majority rule (extended)’ >50%). Trees were visualized using the software Geneious v4.8 ([Biomatters Ltd, New Zealand, [http://www.geneious.com/](http://www.geneious.com/)]. Information on genome organisation was obtained from the Phytozome ([www.phytozome.org](http://www.phytozome.org)) and TAIR.

**Supporting Information**

**Figure S1** AtNRT1, 2 and 3 reciprocal BLASTp results. Depiction of forward and reverse reciprocal BLASTs is provided for (A) poplar, (B) *Brachypodium*, (C) maize, (D) sorghum and (E) rice. BLASTp 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 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 BLASTp results, an upward facing arrow head (†) indicates a tandem repeat with identical BLASTp score, red accessions indicate neighbouring genes in opposing orientation with identical BLASTp scores and blue accessions indicate genes in close proximity to each other with identical BLAST scores.

**Figure S2** Phylogenetic relationship of potential grass PTR transporters orthologues to AtNRT1.6 and AtNRT1.7. 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). Highlighted in the box are the three closest PTR homologues to AtNRT1.6 and AtNRT1.7 ([AT1G18880, AT3G47960 and AT5G62680](http://www.arabidopsis.org/)) as well as orthologous grass PTR transporters ([all in brown](http://www.arabidopsis.org/)). This figure provides rationale for exclusion of grass transporters from AtNRT1.6 and AtNRT1.7 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 5 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 (AY253448), HvNRT3.1B (AY253449) and HvNRT3.1C (AY253450). 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

Table S1  Identifiers, annotations and family similarities for the Arabidopsis NRT genes described in this study. Provided are the TAIR annotation and identifier, UniProtKB identifier and RefSeq identifier for each NRT gene. Amino acid length of each protein is provided as well as the amino acid identity (%) between the NRTx.1 protein and the other family members. (XLS)

Table S2  Comparison of the NRT2 gene intron and exon numbers for dicots and grasses. Gene identifiers and exon and intron numbers are provided for the dicots Arabidopsis thaliana, Populus trichocarpa, Manihot esculenta, Ricinus communis and Vitis vinifera; and for the grasses Oryza sativa, Zea mays, Sorghum bicolor and Brachypodium distachyon. (XLS)

Table S3  Web addresses for the genome databases searched during this study including Arabidopsis thaliana, Brachypodium distachyon, Oryza sativa, Populus trichocarpa, Sorghum bicolor and Zea mays. (XLS)

Table S4  List of candidate homologues excluded from analysis after manual inspection of multiple sequence alignments and their corresponding trees. (XLS)

Acknowledgments
We wish to thank Carl Simmons, Antoni Rafalski and Kannwarpal Dhugga for their comments on the manuscript. Thanks also to Christina Morris for critical comments on the manuscript.

Author Contributions
Conceived and designed the experiments: DP JT TG MT BNK UB. Performed the experiments: DP JT UB. Analyzed the data: DP JT UB. Wrote the paper: DP JT TG BNK UB.