World Academy of Science, Engineering and Technology
International Journal of Agricultural and Biosystems Engineering
Vol:5, No:11, 2011

Identification of PIP Aquaporin Genes from Wheat

Sh. A. Yousif and M. Bhave

***Abstract*—**There is strong evidence that water channel proteins 'aquaporins (AQPs)' are central components in plant-water relations as well as a number of other physiological parameters. We had previously reported the isolation of 24 plasma membrane intrinsic protein (PIP) type AQPs. However, the gene numbers in rice and the polyploid nature of bread wheat indicated a high probability of further genes in the latter. The present work focused on identification of further AQP isoforms in bread wheat. With the use of altered primer design, we identified five genes homologous, designated PIP1;5b, PIP2;9b, TaPIP2;2, TaPIP2;2a, TaPIP2;2b. Sequence alignments indicate PIP1;5b, PIP2;9b are likely to be homeologues of two previously reported genes while the other three are new genes and could be homeologs of each other. The results indicate further AQP diversity in wheat and the sequence data will enable physical mapping of these genes to identify their genomes as well as genetic to determine their association with any quantitative trait loci (QTLs) associated with plant-water relation such as salinity or drought tolerance.

***Keywords*—**Aquaporins, homeologues, PIP, wheat

I. Introduction

It was long thought that water crosses biological membranes simply via osmosis. However, this process is slow, non-regulated and inefficient for rapid or regulated water flux. Koefoed-Johnsen and Ussing [1] suggested for the first time that pores or channels could exist in biomembranes. In the late 1980s, CHIP28, a 28-kDa Channel-forming Integral Membrane Protein, a member of the MIP (Major Intrinsic Protein) family was isolated from erythrocytes where it is highly abundant [2]. Later, expression in frog oocytes and subsequent enhanced swelling of the cells in hypotonic solutions showed that CHIP28, or aquaporin 1 (AQP1), facilitated water transport across membranes [3]. Protein conformation analysis of the amino acid sequences indicates the aquaporin family members exhibit six trans-membrane ∝ helices (TMH1-6), connected by five loops (loops A–E) and two highly conserved asparagine-proline-alanine (NPA) motifs (one each, in loops B and E) [4]. The transport substrate specificity appears to be applied primarily at the NPA motifs and the aromatic/Arg (ar/R) selectivity filter, the latter comprised of one residue each from TMH2 and TMH5 and two from loop E (residues LE1, LE2) [5]. Plant aquaporins appear to be much more abundant and diverse in nature compared to those in other organisms. They are commonly classified into seven different subfamilies: plasma membrane intrinsic proteins (PIPs) tonoplast intrinsic proteins (TIPs), NOD26-like intrinsic proteins (NIPs), small basic intrinsic proteins (SIPs), GlpF-like intrinsic proteins (GIPs), Hybrid Intrinsic Protein (HIP) and X Intrinsic Proteins (XIPs) [6]. Several MIP members had been identified in plants, e.g., 33 genes in maize [7], 38 in Arabidopisis [8], and 39 in rice [9]. The plant MIPs are also found to conduct a number of substrates other than water, e.g., glycerol, ammonia, urea, CO_2, boron, or antimony [10,11 and 12], some of which may have physiological functions and/or adverse effects at different internal concentrations. It is thus essential to identify the MIP gene family members in major crop plants for a number of agricultural applications in plant selection , breeding or transgenics. We had previously identified 24 PIP genes from wheat, comprised of 19 from gene isolations (PIP1;2 to PIP1;12, PIP2;5 to PIP2;12) and 5 only as EST (TaPIP1;1, TaPIP2;1, TaPIP2;2, TaPIP2;3, TaPIP2;4) based on cDNA sequences available in Genbank at the time [13]. By mining the PIP aquaporin data from wheat (http://www.ncbi.nlm.nih. gov), nine complete coding sequences (CDS; accession numbers AF139814, AF366564, AF366565, DQ867075.1, DQ867076.1, DQ867077, DQ867078, DQ345447, DQ345446, GQ452384.1). Alignments and sequences comparisons showed that DQ867078 is similar to PIP2;8, while GQ452384.1, AF139814 and DQ867077 are similar to TaPIP1;1, TaPIP2;1, and TaPIP2;3, respectively. This raised the question whether the equivalents of rice PIP2;2 and PIP2;4 and/or other genes might exist in wheat; leading to the present study.

II. Materials and Methods

Genomic DNA (gDNA) was extracted from leaves of common wheat (*Triticum aestivum* L., 6n) cultivar Cranbrook (AUS accession #22660) using the Wizard Genomic DNA Purification Kit (Promega Australia). Forward and reverse primers (Table I) were designed base on DNA sequence data of the 24 of wheat PIP genes [13]. PCR amplifications were performed with 2X Biomix (Bioline, Australia; contains dNTPs, $MgCl_2$ and Taq DNA polymerase), 200ng gDNA of cv. Cranbrook, 0.3 µM primers, in 50 µL volumes. The amplifications involved heating the reaction mix at 94°C for 4 min for initial denaturation; followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 52 - 55 °C

Sh. A. Yousif is with the Agricultural Research Directorate/Ministry of Science and Technology, PO Box 765, Baghdad- Iraq (phone: +9647709960308; e-mail: yousifshatha@ yahoo.com).

M. Bhave is with Environment and Biotechnology Centre, Faculty of Life and Social Sciences, Swinburne University of Technology, Hawthorn, PO Box 218, Melbourne 3122, Australia. (e-mail: mbhave@swin.edu.au).

TABLE I
DEGENERATE PRIMERS USED FOR PCR OF GDNA FOR PIP GENE ISOLATIONS

Primer[a]	Sequence (5′–3′)
PIPF4	CCACSBTCCTCTTCCTCTACG
PIPF5	CCACGCTSCTCTTCCTCTACA
PIPR6	GSCCCACCCAGAAGATCC
PIPR7	GGCCGACCCAGAAGATCCAC
PIPR8	GACCATGTAGAGCAGCGCGC

a: 'F' in the name of the primer indicates forward primers; 'R' indicates reverse primers. S= C + G, B= T + C + G

(primer pair-specific temperatures) for 45 s, and extension at 72 °C for 1 min; and a final extension step of 72 °C for 10 min. The amplification products were separated on a 1% agarose gels, and visible bands of the expected size were purified using a Perfectprep® Gel Cleanup Kit (Eppendorf) and cloned into pGEM®-T Easy vector (Promega Australia) using the manufacturer's protocol. Chemically competent *E. coli* JM109 cells were transformed with the ligation mixes and cultures from individual colonies grown overnight were used in plasmid DNA preparations for sequencing. The sequencing reactions were conducted using 300 ng plasmid DNA, 3.2 pmol of vector-based primers T7 (5′GTAATACGACTCAGGGC 3′) or SP6 (5′ TTTAGGTGA CACAGAATC 3′), 1.5 µL of ABI BigDye Terminator reagent v3.1 (Applied Biosystems, Foster City, CA, USA), and 3.5 µL of the supplied 5X dilution buffer, in 20-µL volumes, per the instructions of the Australian Genome Research Facility, St Lucia, Australia (http://www.agrf.org.au). They were purified using the AGRF protocol and subjected to capillary separation at AGRF using a 3730xl DNA Analyzer (Applied Biosystems). Sequence manipulations were conducted in BioEdit Sequence Alignment Editor v.7.0.5.3 (http://www.mbio.ncsu.edu/Bio Edit/page2.html) [14].

III. RESULTS

Multiple polymerase chain reaction (PCR) products, ranging in size from 400 to 1600 base pairs (bp), resulted from the primer pairs PIPF5/R, PIPF4/R6, PIPF4/R7 and PIPF4/R8. Sequences were obtained for at least two recombinant plasmids for each cloned fragment. The sequences showed high similarity to PIP genes after BLAST analysis (http://blast.jcvi.org/euk-blast/plantta_blast.cgi). Several sequences showed several single-nucleotide polymorphisms and could be grouped together into eight main types of putative wheat PIP genes.

The representatives of the eight types ranged in size from 245 to 848bp, and were aligned in ClustalW (http://www.ebi.ac.uk/ClustalW) with the known wheat PIPs (http://www.ncbi.nlm.nih.gov), comprised of nine complete CDSs (AF139814, AF366564, AF366565, DQ867075.1, DQ867076.1, DQ867077, DQ867078, DQ345447, DQ345446, GQ452384.1) and 19 partial CDSs (PIP1;2 to PIP1;12 and PIP2;5 to PIP2;12) from previous work [13]. This allowed their grouping into two PIP1 (called TaPIPA and B) and six PIP2 (TaPIP C-H) sequences and prediction of introns and exons, and the exons were joined to obtain putative cDNA contigs. Sequence comparisons showed that TaPIPA, TaPIPG and TaPIPH are PIP1;5, PIP2;9 and PIP2;10, respectively, while TaPIP B, C, D, E and F are different from all PIP genes identified in NCBI, indicating that they are likely to be novel members of wheat aquaporin.

Based on the ClustalW alignment with the known wheat PIPs and the phylogenic trees constructed from the ClustalW data using TreeView (data not shown), TaPIPB was designated as PIP1;5b because showed high similarity in sequences of all exons with previously reported TaPIP1;5 (it is sharing 100% identity in predicted amino acid sequences) and intronI but varied in the sequences of intron II and III.(Table II).TaPIPF was named PIP2;9b, and varied from the previously published PIP2;9 in the sequence and length of their single intron (intronIII), the length of this intron is 108 and 107 in PIP2;9 and PIP2;9b respectively (Table II).The sequences of TaPIPC, D and E compelled us to ask whether they may represent TaPIP2;1 TaPIP2;2, TaPIP2;3 and TaPIP2;4 (GenBank Acc. CK163432, CK163244, CK162927 and CD872490 respectively) which were originally identified only as ESTs [13] and have no gene structures available. Putative cDNA contigs and the resulting putative protein sequences of TaPIPC, D and E were aligned in ClustalW with those of ESTs, and the phylogenic trees constructed from the alignment data (not shown) exhibited highest sequence identity to TaPIP2;2 (Table III) and we suggested that TaPIPC, D, E could be called TaPIP2;2, TaPIP2;2a and TaPIP2;2b respectively (Fig.1) because they Sharing 93-95% identity in predicted amino acid sequences of CK163432 (TaPIP2;2 EST) (Table III) but they are different from each other in the length of intron I which is 82,112 and 107bp respectively (Table II) and the location of a Glycine residue of loop B agreeing with Forrest and Bhave [13] . It is presently unclear length of exon I and II and if they contain additional introns, due to the amplified section being a partial gene.All sequenced introns exhibited typical GT/AG splice junctions except TaPIP2;10 (TaPIPH) showing GC/AG (Table II). TaPIPs contain the signature sequences including two highly conserved NPA motifs (Table IV), while the partial sequences of TaPIP2;2, TaPIP2;2a and TaPIP2;2b showed only the first NPA (Fig.1 and Table IV). A second narrower constriction, the ar/R region, is formed toward the extracellular vestibule. All ar/R filter residues could be ident- fied in TaPIP1;5b and TaPIP2;9b, and were found to be Phenylalanine– Histidine – Threonine– Arginine (Table IV).

TABLE II
INTRON SEQUENCES OF PIP GENES

Gene	Intron I	Intron II	Intron III
PIP2;10	GCACGCACGCACGCACGCA CATATGCATACGTCCATCTC CCTAGAACAACTTAACCCAC ACATGGATACGATTCGATAC GTGAACTGTCGAGTGTACTG CTGCAG	0	_____
PIP2;9	0	0	GTAATGAAGCTTCTTCCTCTCTCTCTCTACCTC CCGCCGTCTCCATCTGTTTTCTTCTTCTACCTC TTGCAACTTGCAAGAAGAAGCTAACCTCCCAT GCCTTNNNNN
PIP2;9b	0	0	GTAATTAAGCTTCTTCCTCTCTCTACCTTACGT TTTCTTCTTCTACCTAATCTTGCAAGAAGAAG AAGAAGCTAACTAACTTGGCATGCTTTGAACA TGACATGCAG
TaPIP2;2	GTCCGTCCTTCGACCACTGG TTCTCTTCTCGCGCGTGCAT ATGCGTGTGGCTGACTGTTC TCTCTTCACGATGTATTCAT AG	____	_____
TaPIP2;2a	GTCCGTCCATCGATCGACCA CTAGTTCTCTTCTCGCGCGT GCATAACTGCACATGCATGC ATCATGCATGGGTCATGCGT GAGACTGACTGTTCTCTCTT CACGATGCATAG	____	_____
TaPIP2;2b	GTCCGTCCTTCGTCCACTGG TTCTCTTCTCGCGCATGCAT ACATAACTCCATATGCATGC ATCATGCATTCGTGTGAGAC AGACTGTCCTCTGTTCACGA TGCATAG	____	_____
TaPIP1;5	GTACCGCTTCTTCCTTGCTCT CTGCTTCTTTCTTGTAGCTTC CTACAGTATGTGGTGCTAAC TGGTAAGGGATGGGTGGAT GCAG	GTGAGTACCATCCCATCACCG CCGCTCTGCTCTTCCCCATGTT GAATTTCCTTACTGTGCTTAC TTAAAACTTAATGGTGTGTGG AACGTCGACAG	GTGAGTGAACTAAAACCGAAACTGAAACCTA ACACCCCTCCTTTCCTTTGTTCTGAACTTGAAA GCATGGCACTTGCTTGTTTTCTGCACAGTAGT AACTGCAGGCATGTGGTGGCAGTAGTATTCAG TAACTTAAAGCTGATGCCGACGCATAGCTACT CTCTATTGCACACCTAATTACCCGATGTAAAT TGGGAAGTTTATCAGTCTGGTGAACTGATGTG ATAGTAGCTGCAATAATTGAGCAGTTGTAGCT GATGATAGTAGTGTATCATCTGCTGGTTATTC AAATATGGGAAGTCTACCCTGATCTCAAGGGA AATGCTGAGAGACTTGCCTCGGCGTAGCACGC AACCACTGCCTTCTGGTAGTAGGACAGTGACT TGAGTGTCCATTTAGTAGATTATTATCTGGGA AATTGTTTGGCATTGCTGTGCTTATAGTAGTA ATAAGGGCTGACATATCCATCAAATTTTCGCG TCTGCAG
TaPIP1;5b	GTACCGCTTCTTCCTTGCTCT CTGCTTCTTTCTTGTAGCTTC CTACAGTATGTGGTGCTAAC TGGTAAGGGATGGGTGGAT GCAG	GTGATGCCATCACCGCCGTTC TACTTCTTTCCCCCGTCGACA GTCAGTAGTAAAAAAAGATG TTCAAATTTCCTTACTGTGCTT AGTTAATGACTACTTATTGCC TGTGGAACGTCGACAG	*GTGAGTGAACCGAAACCAAACACCCCTCCTT TCCTCTGTCTGAACTTGAAAACA

*Indicates partial sequence of intron III; ⎯ indicates absence of sequence data; splice junctions are marked with bold letters.

IV. Discussion

The present work is the first report of the genomic sequences of the wheat TaPIP2;2 genes. The identity in exon sequences but variations in intron sequences indicate they could be homeologous; this can be tested using nullisomic-tertrasomic lines. Similarly, the genes TaPIP1;5b and TaPIP2;9b may also represent homeologs of the previously identified genes. TaPIP1;5b (TaPIPB) shows intron II to be 121bp, but the length of intron III is unclear because due to the sequence being partial, while the intron III in PIP1;5 is 488 .TaPIP2;9b (TaPIPG) contains intron III only, of 107 bp (Table II). Only intronI had been identified in previous work [13] in the partial genomic sequence of PIP2;10; the present work isolated further genomic sequence (TaPIPH) for this sixth gene, and no other intron was found, at least in the section sequenced (Table II). The NPA motif is present once in a loop between the second and third membrane-spanning domains (loop B) and once in a loop between the fifth and membrane-spanning domains (loop E).

These two loops are thought to be involved in forming the pore through which water molecules move [15]. A second narrower constriction, the ar/R selectivity filter is formed toward the extracellular vestibule, above the NPA [16], and has a role in specific transport of water molecules as well as possibly in transport of other substrates. The ar/R filter of TaPIP1;5b and TaPIP2;9b showed the combination F-H-T-R, which has been reported for major intrinsic protein (MIPs) encoded by PIP genes in many crops including wheat which have been shown to be water transporters. This suggests these two wheat genes may also have a similar function.

The sequence data obtained in this work on three new TaPIP2;2 members and one likely new homeologue each of TaPIP1;5 and TaPIP2;9 thus provide further tools for the physical and genetic mapping of these important genes, for identifying their chromosomal locations or genetic linkage to water homeostasis-related traits, respectively.

Acknowledgment

The authors are grateful to Dr Kerrie Forrest (current address Department of Primary Industries, Victoria, Australiafor helpful discussions and advice on sequence analyses. This work was partially supported by a visiting post-doctoral through the Endeavour research programme of 2007 Endeavour AusAID Iraq Research Fellowship**,** Department of Education, Employment and Workplace, Government of Australia.

References

[1] V. Koefoed-Johnsen and H. Ussing, "The contributions of diffusion and flow to the passage of D2O through living membranes. " Acta Physiol Scand, vol. 28, pp. 60–76, 1953.

[2] B. Denker, B. Smith, F. Kuhajda, and P. Agre, "Identification, purification and partial characterization of a novel Mr 28,000 integral membrane protein from erythrocytes and renal tubules. " J Biol Chem., vol. 263, pp.15634–15642, 1988.

[3] G. Preston, J. Jung, W. Guggino, and P. Agre, "Membrane topology of aquaporin CHIP – analysis of functional epitope-scanning mutants by vectorial proteolysis. "J Biol Chem., vol. 269, pp. 1668– 1673, 1994.

[4] J. Reizer, A. Reizer, and M. Saier, "The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins." Crit Rev Biochem Mol Biol., vol. 28, pp.235–257, 1993.

[5] H. Sui , B. Han, J. Lee, P. Walian and B. Jap, "Structural basis of water-specific transport through the AQP1 water channel." Nature, vol. 414, pp. 872–878, 2001.

[6] J.Danielson and U. Johanson, " Unexpected complexity of the Aquaporin gene family in the moss Physcomitrella patens" BMC Plant Biol. Vol.8: 45 , 2008. doi: 10.1186/1471-2229-8-45.

[7] F. Chaumont, F. Barrieu, E. Wojcik, M. Chrispeels and R. Jung, "Aquaporins constitute a large and highly divergent protein family in maize. "Plant Physiol.,vol. 125, pp.1206–1215, 2001.

[8] R. Kaldenhoff, A. Bertl, B. Otto, M. Moshelion, and N. Uehlein, "Characterization of Plant Aquaporins. " Methods in Enzymology, vol.428, pp. 505-526, 2007.

[9] A. Bansal and R. Sankararamakrishnan, "Homology modeling of major intrinsic proteins in rice, maize and Arabidopsis: comparative analysis of transmembrane helix association and aromatic/ arginine selectivity filters. "BMC Struct Biol.,vol. 7,pp. 27, 2007.

[10] R. Kaldenhoff, M .Fischer, "Aquaporins in plants." Acta Physiol.,vol. 187, pp.169–176, 2006.

[11] F.Chaumont, M. Moshelion and M.Daniels, "Regulation of plant aquaporin activity." Biol. Cell , vol.97,pp 749−764, 2005.

[12] S.Mandal , M. Maharjan, S. Singh, M. Chatterjee and R. Madhubala, "Assessing aquaglyceroporin gene status and expression profile in antimony-susceptible and -resistant clinical isolates of Leishmania donovani from India." J Antimicrob Chemother. VOL.65, PP. 496-507, 2010.

[13] K. Forrest, and M. Bhave, "The PIP and TIP aquaporins in wheat form a large and diverse family with unique gene structures and unctionally important features. " Funct Integr Genomics, vol. 8, pp. 115-33, 2008.

[14] T. Hall, " BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symp Ser., vol. 41, pp.95–98, 1999.

[15] M. Chrispeels, N. Crawford and J. Schroeder, "Proteins for Transport of Water and Mineral Nutrients across the Membranes of Plant Cells." Plant Cell, vol. 11,pp. 661-676, 1999.

[16] I. Wallace and D. Roberts, "Homology modeling of representative subfamilies of arabidopsis major intrinsic proteins. classification based on the Aromatic/Arginine selectivity filter." Plant Physiology, vol. 135, pp. 1059-1068, 2004.

```
CK163432      RVPLQQTPX KLQGTSPELA LAKDIEAAPQ GGEFSTKDYS DPPPAPIVDF 50
TaPIP2;2b
TaPIP2;2
TaPIP2;2a

CK163432     EELTKWSLYR AVIAEFVATL LFLYITVATV IGYKHQSDPT VNTTDAACSG 100
TaPIP2;2b                    TL LFLYITVATV IGYKQQSDPT VNTTDVACSG
TaPIP2;2                     TL LFLYVTVATV IGYKHQSDPT VNTTDAACSG
TaPIP2;2a                    TF LFLYVTVATV IGYKHQSDPT VNTTDAACSG
                             *  **** ***** **** ***** ***** ****

CK163432     VGILGIAWAF GGMIFVLVYC TAGVSGGHIN PAVTFGLFLA RKVSLIRALL 150
TaPIP2;2b    VGILGIAWAF GGMIFVLVYC TAGVSGGHIN PAVTFGLFLA RKVSLIRALL
TaPIP2;2     VGILGIAWAF GGMIFVLVYC TAGVSGGHIN PAVTFGLFLA RKVSLIRALL
TaPIP2;2a    VGILGIAWAF GGMIFVLVYC TAGVSGGHIN PAVTFGLFLA RNVSLIRALL
             ********** ********** ********** ********** * ********

CK163432     YIIAQCLGAI CGVGLVKGFQ SSYYVRYGGG ANELSAGYSK GTGLAAEIIG 200
TaPIP2;2b    YMA
TaPIP2;2     YMA
TaPIP2;2a    YMD
             *

CK163432     TFVLVYTVFS ATDPKRNARD SHIPVLAPLP IGFAVFMVHL ATIPITGTGI 250
TaPIP2;2b
TaPIP2;2
TaPIP2;2a

CK163432     NPARSLGAAV IYNTDKAWDD QWIFWVGALI XAXIAAXYHQ YVLRASAAKL 300
TaPIP2;2b
TaPIP2;2
TaPIP2;2a
Clustal Co

CK163432     GFYRSNSHGR SGRDVICAKK KAALSKCAMV ARTTPLSCSV SRVFASSAAT 350
TaPIP2;2b
TaPIP2;2
TaPIP2;2a

CK163432     FLXLCICEVT AHRTSTVPDF ASL 370
TaPIP2;2b
TaPIP2;2
TaPIP2;2a
```

Fig. 1 Alighment of the amino acid of GenBank Acc. CK163432 (TaPIP2;2 EST), TaPIP2;2,TaPIP2;2a and TaPIP2;2b. The numbers refer to the respective amino acid sequence. The conserved motifs (NPA)are marked with bold letters

TABLE III
% IDENTITY OF GENES ISOLATED FROM gDNA TO EST

Gene name	Isolated gene	% identity to CK163432		% identity to CK163244		% identity to CK162927		% identity to CD872490	
		cDNA	protein	cDNA	protein	cDNA	protein	cDNA	protein
TaPIP2;2	TaPIPC	97	95	87	88	86	86	86	87
TaPIP2;2a	TaPIPD	95	93	89	88	88	86	87	87
TaPIP2;2b	TaPIPE	96	94	81	80	81	77	79	79

TABLE IV
THE ar/R AND NPA MOTIF RESIDUE IN THE PUTATIVE PIP PROTEINS

Gene	H2 and Loop B	H5 and Loop E
TaPIP1;5b	VGIQGIAWS**F**GGMIFALVYCTA GIS▲ GGHI**NPA**VTFGLF	GFAVFLV**H**LATIPITG**T**GI**NPAR**SLGAAIIY
TaPIP2;2	VGILGIAWA**F**GGMIFVLVYCTA GVS▲ GGHI**NPA**VTFGLF	-
TaPIP2;2a	VGILGIAWA**F**GGMIFVLVYCTA GVS▲ GGHI**NPA**VTFGLF	-
TaPIP2;2b	VGILGIAWA**F**GGMIFVLVYCTA GVS▲ GGHI**NPA**VTFGLF	-
TaPIP2;9b	VGILGIAWA**F**GGMIFVLVYCTA GVS▲ GGHI**NPA**VTFGLL	GFAVFMV**H**LATIPITG**T**GI**NPAR**SLGAAVIY

▲indicates intron position; - indicates absence of sequence data.