Research Article

Characterization of a gene encoding ornithine carbamoyltransferase from rice

Md. Shafiqul Islam Sikdar · Jung-Sup Kim

Received: 3 December 2009 / Accepted: 17 December 2009 © Korean Society for Plant Biotechnology

Abstract Ornithinine carbamoyltransferase (OTC) is an enzyme that catalyzes the key step in arginine biosynthesis in bacteria and plants. OTC is also involved in the urea cvcle and deficiency of the enzyme in human leads to disease. The argF gene encoding OTC has been reported in many bacteria and few plants. Here we report the characterization of a gene encoding OTC from rice (OsOTC). Analysis of a cDNA sequence from rice revealed that the full-length open reading frame of OsOTC consisted of 367 amino acids, corresponding to a protein of approximately 39.7 kDa. The predicted amino acid sequence of OsOTC harbor distinct five OTC signature sites and is highly homologous to that of enzymes of plants, animals and many bacterial OTCs. Expression of OsOTC in argF mutants of Escherichia coli showed that the gene was able to functionally complement to the mutant. These results suggest that the OsOTC encode a protein for ornithine carbamoyltransferase in rice.

Keywords Arginine, Rice (*Oryza sativa*), Ornithine carbamoyltransferase. Gene cloning

Introduction

Arginine (Arg) and its metabolism are of central importance in plant biology (Slocum, 2005). Arg is one of the most metabolically versatile amino acids and serves as a precursor for the synthesis of urea, nitric oxide, polyamines, proline, glutamate, creatine, and agmatine. Arg is metabolized through a complex in animals and highly regulated set of pathways that

Md. Shafiqul Islam Sikdar · J.-S. Kim (⊠) (Faculty of Biotechnology, Jeju National University, Jeju, 690-756, Korea) e-mail: biotech2020@jejunu.ac.kr remain incompletely understood at both the whole-body and the cellular levels (Morris, 2006). The main pathway for Arg metabolism in animals is the urea or Krebs-Henseleit cycle. Arg may represent as much as 40% of the total nitrogen in seed storage proteins (Vanetten *et al.* 1963) and is the most abundant free amino acid in cotyledons of pea seeds, where it is catabolized as a nitrogen source during the early stages of germination (deRuiter and Kollöffel, 1985).

The first step in this pathway is the reaction between ornithine and carbamoyl-phosphate to form citrulline. This reaction catalyzed by an enzyme ornithine carbamoyltransferase (OTC, EC 2.1.3.3) that is a ubiquitous and essential enzyme involved in the key step of the Arg biosynthesis (Fig. 1). This reaction is the essential step in the *de novo* synthesis of Arg in prokaryotes and eukaryotes (Thomson, 1980; Acaster et al. 1989; Huygen et al. 1987). L-citrulline is a substrate for argininosuccinate synthase, the enzyme involved in the final step in arginine biosynthesis. In animals, OTC is involved in the urea cycle and deficiency of the OTC enzyme in human leads to the broad range of diseases (Tuchman et al. 1998). The OTC genes have been characterized from all kinds of organisms including bacteria, fungi, plants and animals such as Escherichia coli (Jin et al. 1997; Ha et al. 1997; Langley et al. 2000), Mycobacterium tuberculosis (Sankaranyanan et al. 2008; Danielle et al. 2008), Pseudomonas aeruginosa (Villeret et al. 1995), Pseudomonas syringae (Mosqueda et al.1990), Pyrococcus furiosus (Villeret et al.1998; Massant et al. 2003), Saccharomyces cerevisiae (Eisenstein et al. 1984); Arabidopsis thaliana (Slocum et al. 2000), Pisum sativum (deRuiter and Kollöffel, 1985; Eid et al. 1974; Slocum et al. 1991; Taylor et al. 1981; Williamson et al. 1996), sugarcane (Glenn and Maretzki, 1977), Phaseolus vulgaris (Lee et al. 1997), Pyrus malus (Spencer and Titus, 1974) and also characterized in humans (Shi et al. 1998). However, there is



Fig. 1 Schematic diagram of arginine biosynthesis pathway in plants. The gene names from bacteria and yeast are provided in italic letters (adopted and modified from Slocum, 2005)

little information regarding the role of OTC in regulation of the biosynthesis of Arg in plants, coordination of Arg and pyrimidine pathway activities at the level of carbamoyl-phosphate utilization, or related metabolic process (Slocum et al. 2000). There are little reports of Arg synthesis related in OTC in rice. The results related OTC with different locations and chemical properties have been reported (deRuiter and Kollöffel, 1985; Eid et al. 1974; Glenn and Maretzki, 1977). The enzymes of plant Arg biosynthesis and their metabolic control have been incompletely characterized, but appear to be similar to those of most prokaryotes and lower eukaryotes. Subcellular compartmentation of Arg synthesis is still poorly understood. With few exceptions, plant genes encoding enzymes in Arg synthesis have not been identified and their gene products have not been characterized (Slocum, 2005). Major cereal crops such as rice, wheat and corn contain lower prot ein amounts compared to leguminous crops. To improve nutritional qualities in cereal crops, it is necessary to investigate the Arg biosynthesis pathway in crop plants. We studied here the characterization and expression of a gene encoding OTC from rice that is such an important crop plant.

Methods and Materials

Strains and Plasmids

Two *E. colis* trains, *H1238* and *Gif41*, were used in the study. The genotype of the strains is as follows: *H1238* [Hfr(PO100), *thr-25, fhuA49, argF58, relA1, spoT1, purA54, argI61*] and *Gif41* [Hfr(PO1), *thrC1001, LAM-, e14-, relA1, spoT1, thi-1*] (Theze *et al.*1974). The source of both strains was the *E. coli* Genetic Stock Center (CGSC) in Yale University, USA. The GenBank accession number of the rice EST expected to encode for putative ornithine carbamoyltransferase is AK243192 and the clone name is J100039M06. The rice EST was ordered from Rice Genome Resource Center (RGRC), National Institute of Agro biological Science (NIAS), Japan.

DNA sequence analysis

An EST clone (GenBank accession no. AK243192) was derived from rice cDNA library (Osato et al. 2002) from developing seeds prepared in pBluescript SK-. DNA sequencing was performed by an automatic sequencer (A1Fexpress DNA sequencer, Pharmacia Biotech. Inc., UK) with synthetic oligonucleotide primers. Nucleotide sequences and amino acid sequences were compared with sequences present in the Gen-Bank and EMBL databases and analyzed using BLAST (Wheeler et al. 2003) and CLUSTAL W multiple sequence alignment program (Thompson et al. 1994) or Biology WorkBench 3.2 (http://workbench.sdsc.edu; San Diego Supercomputer Center; University of California San Diego, USA). Comparison of sequences was performed at the nucleotide and amino acid level. Motifs were searched by GenomeNet Computation Service at Kyoto University (http://www.genome.ad.jp) and phylogenetic tree with bootstrap value was prepared by using the Mega 4.1 program (Kumer *et al.*, 2008).

Polymerase chain reaction (PCR) and recombinant construct

To amplify the full-length open reading frame (ORF) and overexpress the gene product in E. coli, the OsOTC specific primers were designed from the sequence information around the translational start and stop codons of OsOTC: OsOTC-F (5'-AGG ATC CAC AGG TTG AAA GAG AGC TGA TG C-3') and OsOTC-R (5'-AGG ATC CTC ACA AAG ACC TGG GTT ACA AGC-3'). Polymerase chain reaction (PCR, Sambrook et al. 2001) was conducted to amplify the ORF of OsOTC. After a plasmid in rice EST clone was purified from a pellet harvested from liquid culture containing Amp, the ORF was amplified using the designed primers from OsOTC sequence. The PCR reaction was performed using My Cycler TM PCR system (BioRad, U.S.A) for 40 cycles with 95° C for 1 min, 45 °C for 1 min, and 72 °C for 2 min, with 10 μ M primers. The PCR products were analyzed on 1% (w/v) agarose gel. The 1.1 kb PCR fragment was subcloned into a pMPM-K2 cloning vector and digested with BamHI and inserted into the same site of pBluescript II KS+ to construct pB

:: OsOTC. Restriction analysis was performed to confirm the construct.

Functional Complementation

The strains, *argF* mutant *H1238* and wild-type *Gif41*, were transformed with *pB::OsOTC* and pBluescript II KS+ as a control by electroporation (ECM399, BTX, USA), respectively after producing competent cell by washing with water and glycerol (Kim and Leustek, 1996) using the cuvette with 0.1 cm electrode gap and then plated on LB medium (20 g/L) with Amp (100 µg/ml). Amp-resistant colonies were then replica plated onto M9 minimal medium [(5 × M9 salts (200 ml/L), 1 MMgSO₄ (2 ml/L), 1 MCaCl₂ (0.1 ml/L)], containing 1 mM Isoptopyl β-D-thiogalactopyranoside (IPTG), 20% glucose (20 ml/L), Amp (25 µg/ml), and 19 amino acids (Sigma, Germany) each at a concentration of 25 µg/ml, excluding Arg. The plate was incubated at 37°C for 2-days. The growing colonies were retested for growth on Arg free medium (Kim and Leustek, 1996).

Growth assay in E. coli

The *argF* mutant *E. coli* harboring the *pB::OsOTC* construct, control plasmid or wild-type *E. coli* with control plasmid was used in growth assay. The 1 ml of overnight culture in LB medium with Amp of the *E. coli* strains were inoculated in 100 ml of M9 minimal medium containing 1 mM IPTG, 20% glucose, Amp (25 µg/ml) and 19 amino acids excluding Arg. The bacterial cell growth at 37° C was monitored through optical density measurements every hour using the spectrophotometer (UV1101, Biochrom, England) at 595 nm (OD₅₉₅). After 12 hrs, the diluted culture was plated and incubated overnight at 37° C to check the survival of *E. coli* cells.

Results and Discussion

Sequence analysis of OsOTC

An EST clone (GenBank accession number AK243192 and clone name J100039M06) obtained from Rice Genome Resource Center (RGRC) was analyzed to determine the nucleotide sequence using designed primers. The EST (*OsOTC*) sequence contained a full-length open reading frame consisted of 1084 bp, encoding for a protein of approximately 39.7 kDa. The expected isoelectric point of the protein was 6.4. Data analysis indicates that the *OsOTC* sequence was identical to the genomic region located in chromosomes II in rice. Alignments and comparison of the predicted amino acid sequence





Fig. 2 Amino acid sequence alignment of OTCs using Boxshade program after CLUSTAL W alignment. Completely conserved, identical and similar residues are visually shown as yellow, green and cyan, respectively. Accession numbers are as follows: AK24 3192 (OsOTC from *Oryza* sativa, this study), EDN63490 (ScOTC from *Saccharomyces cerevisiae*), CAA30121 (MmOTC from *Mus musculus*), NP_850730 (AtOTC from *Arabidopsis thaliana*) and NP_416973 (EcOTC from *Escherichia coli*)

for the *OsOTC* with other similar sequences have been highly conserved among diverse monocot and dicot plant species exhibiting between identity 73% to 88% such as *Zea mays* (88%), *Arbidopsis thaliana* (73%), *Ricinus communis* (76%) and *Canavalia lineata* (77%) and also showed identity 30% to 55% with different bacteria like *E. coli* (40%) and *Bacillus subtilis* (46%), respectively.

Analysis of the amino acid sequence of *OsOTC* in BLOCKS and PRINTS database revealed that there are 5 signature sites for OTC found in that amino acid sequence such as (110-124) (MIFAKPSMRTRVSFE), (143-156) (QMGKREETRDVA RV), (183-197) (VPVINGLTDYNHPCQ), (282-292) (VY TDVWASMGQ), (344-355) (SIVFPQAENRMH)(Fig. 2). Phylogenic analysis based on comparison of the related sequences further indicated that OsOTC is divergent and classified in plant OTCs distinct from bacterial or animal OTC groups and proposed to evolve from ancestral bacterial OTC. Numbers at nodes indicate levels of bootstrap support based at neighbor-joining analysis of 1000 re-sampled data set by using Mega 4.1. Numbers on branches are percentage of bootstrap analysis supporting the grouping of each branch (Fig. 3).



Fig. 3 Phylogenetic analysis of OsOTC related proteins using Clustal W and Mega 4.1 program. GenBank accession numbers are as follows: AK243192 (OsOTC from, Oryza sativa), YP 00 1124801 (GtOTC, Geobacillus thermodenitrificans), YP 002316 490 (AfOTC, Anoxybacillus flavithermus), NP 389007 (BsOTC, Bacillus subtilis), ZP_02327162 (PIOTC, Paenibacillus larvae), YP 001111650 (DrOTC, Desulfotomaculum reducens), YP 645 612 (RxOTC, Rubrobacter xylanophilus), YP 590894 (CkOTC, Candidatus koribacter), YP 002306817 (ToOTC, Thermococcus onnurineus), NP 578323 (PfOTC, Pyrococcus furiosus), YP 00 1130397 (PvOTC, Prosthecochloris vibrioformis), XP 0014154 32 (OlOTC, Ostreococcus lucimarinus), CAA04115 (AtOTC, Arabidopsis thaliana), NP 001151252 (ZmOTC; Zea mays), XP 002510416 (RcOTC, Ricinus communis), AAF17705 (CIOTC, Canavalia lineata), ACM47214 (MtOTC, Medicago truncatula), NP_001157474 (SsOTC, Sus scrofa), AAI07155 (HsOTC, Homo sapiens), CAA30121 (MmOTC, Mus musculus), NP 037210 (Rn OTC, Rattus norvegicus), ACX40954 (EcOTC, Escherichia coli)



Fig. 4 Functional complementation assay. The *argF* mutant *E. coli* strain H1238 containing *pB::OsOTC* and pBluescript II KS+; and Gif41 wild-type *E. coli* containing pBluescript II KS+ as a control

Functional complementation

The recombinant DNA, *pB::OsOTC*, was constructed using ORF of PCR-amplified *OsOTC* fragment. After transformation to *E. coli*, *OsOTC* activity *in vivo* was monitored a medi-



Fig. 5 Growth curves of *E. coli* mutant *H1238* harboring *pB:: OsOTC* and pBluescript II KS+; and *Gif41* containing control plasmid. Bacterial cells were grown at 37°C in M9 minimal medium containing 19 amino acids except Arg. Growth was monitored via optical density measurements at 595 nm (OD₅₉₅). Symbols: \blacklozenge , *H1238* + *pB::OsOTC* \blacktriangle , Wild type + pBluescript SK+ \blacksquare , *H1238* + pBluescript SK+

um containingIPTG and 19 amino acids excluding Arg. The functional complementation was performed using the *argF* mutants of *E. coli* to confirm the enzyme activity by the gene product of *OsOTC*. To check the viability of *E. coli* cells by *OsOTC* activity, the cells harboring *pB::OsOTC* were cultured for 12 hrs with shaking and the diluted portion was plated on agar medium containing the 19 amino acids and Amp (100 mg/ml) without Arg (Fig. 4). The viable colonies appeared in the *argF* mutants of *E. coli* harboring *pB::OsOTC* otherwise the same *E. coli* with control plasmid could not. These results indicated that the mutant *E. coli* viability with *pB::OsOTC* was an evidence of functional OTC activity by complementation.

Expression of OsOTC in E. coli

A growth study was performed to determine whether it would increase the sensitivity of bacterial cells to Arg by the *OsOTC* or not. The *pB::OsOTC* construct was transformed into *argF* mutant *E. coli H1238*. The pBluescript II KS+ as control plasmid were also transformed into wild type (Gif41) and *argF* mutant *H1238*. The *OsOTC* activity was monitored through a growth assay in the absence of Arg. Bacterial cells were grown in M9 minimal medium with 19 amino acids excluding Arg, containing IPTG and Amp. The wild type *E. coli* strain *Gif41* harboring control plasmid grew normally and showed classical S-shape growth curve in the medium lacking Arg (Fig. 5). An explanation is that the wild type *E. coli* strain could synthesize Arg itself thus it grew normally in the medium. The *argF* mutant strain *H1238* expressing *OsOTC* also grew normally and showed S-shape growth curve in the same medium (Fig. 5). When the mutant *H1238* with control plasmid was showed dramatic retardation in the medium due to lack of Arg. In this case, the *argF* mutant *E. coli* strain *H1238* could not synthesize Arg itself (Fig. 5). From these results, it was concluded that the expression of the OsOTC is able to functionally complement and this is an outcome of OsOTC activity. It would deduce that the OsOTC is functioning in Arg biosynthesis in rice.

We are now trying to investigate some important clues about substrate specificity of the enzyme by purifying recombinant OsOTC in *E. coli* and the physiological functions of this novel enzyme for Arg metabolism by screening T-DNA insertion mutants with the expression of OsOTC. These results may constitute a starting point at the molecular level to investigate Arg biosynthesis in rice, which might eventually be applied to modify the nutritional compositions of crop plants.

Acknowledgements

We wish to thank Rice Genome Resource Center (RGRC), National Institute of Agro biological Science (NIAS), Japan and *E. coli* Genetic Stock Center (CGSC) at Yale University for providing an EST clone AK243192 and *argF* mutant (CGSC # 5408) of *E. coli*, respectively. Authors thanks Seong-A Kang, Su-Mi Son, Min-Ju Kim for their technical assistance in experiments.

References

- Acaster MA, Scott-White S, Weitzman PDJ (1989) Carbamoyltransferase reactions in plants. A survey for enzymatic diversity and the potential for herbicidal activity for transition state analogue inhibitors. J Exp Bot 40:121-1125.
- Danielle RN, Nira P, Suely SK, Virmondes R, Antonio C-N (2008) Identification of *Mycobacterium tuberculosis* Ornithine carbamoyltransferase in urine as a possible molecular marker of active pulmonary tuberculosis. Clinic Vaccine Immunol 15: 638-643.
- deRuiter H, Kollöffel C (1985) Properties of ornithine carbamoyltransferase from *Pisum sativum* L. Plant Physiol 77:695-699.
- Eid S, Waly Y, Abdelal AT (1974) Separation and properties of two ornithine carbamoyltransferases from *Pisum sativum* seedlings. Phytochem 13:99-102.
- Eisenstein E, Osborne JC Jr, Chaiken IM, Hensley P (1984) Purification and characterization of ornithine transcarbamoylase from *Saccharomyces cerevisiae*. J Biol Chem 259:5139-5145.

- Glenn E, Maretzki A (1977) Properties and subcellular distribution of two partially purified ornithine transcarbamoylases in cell suspensions of sugarcane. Plant Physiol 60:22-126.
- Ha Y, McCann MT, Tuchman M, Allewell NM (1997) Substrateinduced conformational change in a trimeric ornithine transcarbamoylase. Proc Natl Acad Sci USA 94:9550-9555.
- Huygen R, Crabeel M, Glansdor. N (1987) Nucleotide sequence of the ARG3 gene of the yeast Saccharomyces cerevisiae encoding ornithine carbamoyltransferase. Comparison with other carbamoyltransferases. Eur J Biochem 166:371-377
- Jin L, Seaton BA, Head JF (1997) Crystal structure at 2.8 A resolution of anabolic ornithine transcarbamoylase from *Escherichia coli*. Nat Struct Biol 4:622-625.
- Kim J, Leustek T (1996) Cloning and analysis of the gene for cystathionine γ-synthase from Arabidopsis thaliana. Plant Mol Biol 32:1117-1124.
- Kumar S, Nei M, Dudley J, Tamura K (2008). MEGA: A biologistcentric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9(4):299-306.
- Langley DB, Templeton MD, Fields BA, Mitchell RE, Collyer CA (2000) Mechanism of inactivation of ornithine transcarbamoylase by Ndelta-(N'-sulfodiaminophosphinyl)-L-ornithine, a true transition state analogue? Crystal structure and implications for catalytic mechanism. J Biol Chem 275:20012-20019.
- Lee Y, Chin BL, Sang-Gu K, Kwon YM (1997) Purification and characterization of ornithine carbamoyltransferase from chloroplasts of *Canavalia lineate*. Plant Sci 122:217-224.
- Massant J, Wouters M, Glansdorff M (2003) Refined structure of *Pyrococcus furiosus* ornithine carbamoyltransferase at 1.87 A. Biol Crystall Ogr 59:2140-2149.
- Morris S M (2006) Arginine: beyond protein. American J Clinic Nutrit 83(2):508-512.
- Mosqueda G, Van den Broek G, Saucedo O, Bailey AM, Alvarez-Morales A (1990) Herrera-Estrella L., Isolation and characterization of the gene from *Pseudomonmas syringae* pv. *phaseolicola* encoding the phaseolotoxin-insensitive ornithine carbamoyltransferase. Mol Gen Genet 222:461-466.
- Osato N, Itoh M, Konno H, Kondo S, Shibata K, Carninci P, Shiraki T, Shinagawa A, Kikuchi S, Sato K, Kawai J, Hayashizaki Y (2002) A computer-based method of selecting clones for a full-length cDNA project: simultaneous collection of negligibly redundant and variant cDNAs. Genome Res 12:1127-1134.
- Sambrook J, Fritsch EJ, Maniatis T (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory press Cold Spring Harbor New York.
- Sankaranarayanan R, Maia MC, Leonid TC, Craig RG, Fatemeh M, Michae NGJ (2008) The crystal structures of ornithine carbamoyltransferase from *Mycobacterium tuberculosis* and its ternary complex with carbamoyl phosphate and L-norvaline reveal the enzyme's catalytic mechanism. J Mol Biol 375: 1052-1063.
- Shi D, Morizono H, Ha Y, Aoyagi M, Tuchman M, Allewell NM (1998) 1.85 Å Resolution crystal structure of human ornithine transcarbamoylase complexed with N-phosphonacetyl-L-ornithine. Catalytic mechanism and correlation with inherited de-

ficiency. J Biol Chem 273:34247-34254.

- Slocum RD (2005) Genes, enzymes and regulation of arginine biosynthesis in plants. Plant Physiol Biochem 43:729-745.
- Slocum RD, Nichols HF, Williamson CL (2000) Purification and characterization of Arabidopsis ornithine transcarbamoylase (OTCase), a member of a distinct and evolutionarily-conserved group of plant OTCases. Plant Physiol Biochem 38(4):279-288.
- Slocum RD, Richardson DP (1991) Purification and characterization of ornithine transcarbamoylase from pea (*Pisum sativum* L.). Plant Physiol 96:262-268.
- Spencer PW, Titus JS (1974)The occurrence and nature of ornithine carbamoyltransferase in senescing apple leaf tissue, Plant Physiol 54:382-385.
- Taylor AA, Stewart GR (1981) Tissue and subcellular localization of enzymes or arginine metabolism in *Pisum sativum*. Biochem Biophys Res Commun 101:1281-1289.
- Theze JD, Margarita GN, Cohen F, Borne J, Patte C (1974) Mapping of the Structural genes of the three aspartokinases and of the two homoserine dehydrogenases of *Escherichia coli*K-12. J Bacteriol 117:133-140.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W:improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.
- Thompson JF (1980) Arginine synthesis, proline synthesis, and

related processes. In: Mi⁻in BJ (ed) The biochemistry of plants, vol 5, Amino acids and derivatives. Academic Press New York 375-402.

- Tuchman M., Morizono H, Rajagopal BS, Plante RJ, Allewell NM (1998) The biochemical and molecular spectrum of ornithine transcarbamoylase deficiency. J Inherited Metab Dis 21:40-58.
- Vanetten CH, Miller RW, Wolff IA, Jones Q (1963) Amino acid composition of seeds from 200 angiospermous plant species Agric Food Chem. 5:399-410.
- Villeret V, Clantin B, Tricot C, Legrain C, Roovers M., Stalon V (1998) The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. Proc Natl Acad Sci USA 95:2801-2806.
- Villeret V, Tricot C, Stalon V, Dideberg O (1995) Crystal structure of *Pseudomonas aeruginosa* catabolic ornithine transcarbamoylase at 3.0-A resolution: a different oligomeric organization in the transcarbamoylase family. Proc Natl Acad Sci USA 92: 10762-10766.
- Williamson CL, Lake MR, Slocum RD (1996) Isolation and characterization of a cDNA encoding pea ornithine transcarbamoylase (*argF*) and comparison with other transcarbamoylases. Plant Mol Biol 31:1087-1092.
- Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L (2003) Database resources of the National Center for Biotechnology. Nucleic Acids Res 31(1):28-33.