Molecular cloning and characterization of the mouse Acdp gene family

Cong-Yi Wang*1, Ping Yang1, Jing-Da Shi2, Sharad Purohit1, Dehuang Guo1, Haiqian An1, Jian-Guo Gu3, Jennifer Ling3, Zheng Dong4 and Jin-Xiong She1

Address: 1Center for Biotechnology and Genomic Medicine, Medical College of Georgia, 1120 15th Street, PV6B108, Augusta, GA 30912, USA, 2Center for Mammalian Genetics, College of Medicine, University of Florida, Gainesville, FL 32610, USA, 3McKnight Brain Institute of the University of Florida & Division of Neuroscience, Department of Oral Surgery, University of Florida, Gainesville, Florida, 32610, USA and 4Department of Cellular Biology and Anatomy, Medical College of Georgia, 1120 15th Street, CB2917, Augusta, GA 30912, USA

Email: Cong-Yi Wang* - cwang@mail.mcg.edu; Ping Yang - pyang@mail.mcg.edu; Jing-Da Shi - Shi@pathology.ufl.edu; Sharad Purohit - spurohit@mail.mcg.edu; Dehuang Guo - dguo@mail.mcg.edu; Haiqian An - han@mail.mcg.edu; Jian-Guo Gu - jgu@dental.ufl.edu; Jennifer Ling - jling@dental.ufl.edu; Zheng Dong - zdong@mail.mcg.edu; Jin-Xiong She - jshe@mail.mcg.edu

* Corresponding author

Abstract

Background: We have recently cloned and characterized a novel gene family named ancient conserved domain protein (ACDP) in humans. To facilitate the functional study of this novel gene family, we have cloned and characterized Acdp, the mouse homologue of the human ACDP gene family.

Results: The four Acdp genes (Acdp1, Acdp2, Acdp3 and Acdp4) contain 3,631 bp, 3,244 bp, 2,684 bp and 2,743 bp of cDNA sequences, and encode deduced proteins of 951, 874, 713 and 771 amino acids, respectively. The mouse Acdp genes showed very strong homologies (>90%) in both nucleotide and amino acid sequences to their human counterparts. In addition, both nucleotide and amino acid sequences within the Ancient Conserved Domain (ACD) are highly conserved in many different taxonomic species. Particularly, Acdp proteins showed very strong AA homologies to the bacteria CorC protein (35% AA identity with 55% homology), which is involved in magnesium and cobalt efflux. The Acdp genes are widely expressed in all tissues tested except for Acdp1, which is only highly expressed in the brain with low levels of expression in kidney and testis. Immunostaining of Acdp1 in hippocampus neurons revealed a predominant localization on the plasma membrane.

Conclusion: The Acdp genes are evolutionarily conserved in diverse species and ubiquitously expressed throughout development and adult tissues suggesting that Acdp may be an essential gene. Acdp showed strong homology to bacteria CorC protein and predominantly localized on the plasma membrane. These results suggest that Acdp is probably a family of proteins involved in ion transport in mammalian cells

Background

We have recently cloned and characterized a novel gene family named ancient conserved domain protein (ACDP) which encodes four protein members in humans [1]. We found that this gene family is evolutionarily conserved in diverse species ranging from bacteria, yeast, C. elegans, and D. melanogaster to mammals. The sequence conservation and the presence of multiple members within a species...
may imply functional importance associated with the genes. To facilitate the functional analysis of the ACDP gene family, we cloned and characterized Acdp, the mouse homologue of the human ACDP gene family.

**Results**

**Molecular cloning of the Acdp gene family**

To clone the mouse Acdp genes, the human ACDP cDNA and predicted protein sequences were used to search the mouse EST database with the blastn and tblastn programs. Mouse EST markers corresponding to each Acdp member were then identified. For example, EST H3086H12-5 corresponds to Acdp1, W98010 for Acdp2, 603299135F1 for Acdp3 and BG083791 for Acdp4. A modified oligo-dT with a M13 tail was used for the RT reaction. A forward primer from each EST marker and the M13 primer (olig-dT tail) were used to amplify the 3' UTR sequence for each corresponding Acdp gene from the RT products. To obtain 5'-end coding sequences for the Acdp genes, we conducted a series nested PCR with combinations of mouse and human primers. The 5' UTR sequences were identified by directly sequencing BAC DNA containing the corresponding Acdp genes. The BAC clones were identified by screening a CIBR mouse BAC DNA library (Research Genetics). The 5' UTR sequences obtained from above were further confirmed by RT-PCR. The Acdp1 gene contains 3,631 bp of nucleotide sequence and encodes a predicted protein with 951 amino acids (AA). The other three Acdp genes (Acdp2, 3 and 4) contain 3,244 bp, 2,684 bp and 2,743 bp of cDNA sequences, and encode deduced proteins of 874 amino acids, 713 amino acids and 771 amino acids, respectively.

**Tissue distribution**

Northern blot analyses of the Acdp gene family were carried out using membranes purchased from Origene. A total of 12 mouse tissues were included in the study (Fig. 1). Due to sequence homologies between each Acdp member within the conserved domain, probes for Northern bolts were PCR fragments from the last exon and the 3' untranslated region sequences. The mouse Acdp messages showed almost the same tissue distributions as the human ACDP genes. Acdp1 message is highly expressed in the brain, while kidney and testis also showed low levels of expression. Acdp2 showed higher expressions in the brain, kidney and liver. However, the Acdp2 transcript was not present in the skeletal muscle and skin, and it showed very low levels of expression in the rest of tissues. Acdp3 and Acdp4 showed different levels of expression in all tissues tested; the highest expressions for Acdp3 were observed in the brain, kidney, liver and heart, and the highest expressions for Acdp4 were observed in the kidney, small intestine and testis. The expression levels for Acdp3 and 4 in skeletal muscle were barely detectable; however, β-actin showed normal expression suggesting that the results were not a consequence of bad RNA quality (data not shown). The ubiquitous expression pattern may be taken as another indication of the functional importance of Acdp proteins in fundamental biological processes in addition to the sequence conservation in evolutionarily divergent species.

**Chromosomal location**

Radiation hybrid mapping indicated that the Acdp1 gene maps to chromosome 19 between markers D19Mit119 (34.3 cR proximal) and D19Mit112 (13.6 cR distal). The Acdp2 gene maps slightly more distal to the Acdp1 on chromosome 19 between D19Mit9 (2.4 cR proximal) and D19Mit38 (15.1 cR distal). The Acdp3 and Acdp4 genes map to chromosome 1 within one BAC clone (RP23-294I17), proximal to marker D1Mit171 (17.4 cR). These regions are syntenic to the human counterparts.

**Sequence homology and molecular characteristics**

The mouse Acdp genes showed very strong homologies of both nucleotide and AA sequences to the human ACDP genes (Table 1). The highest homologies were observed between the human ACDP2 and the mouse Acdp2 gene (91% of nucleotide identity, 97% of AA identity and 99.4% of AA homology). In addition, the 5' UTR nucleotide sequences (20 bp of nucleotides before start codon) also showed high homologies to the human homologs, for example, the Acdp2 5' UTR sequence showed 95% identities to its human homolog. However, the homologies in the 3' UTR sequences (20 bp of nucleotides after stop codon) were much lower (40–55%) for all Acdp genes except Acdp4 (90% identity to its human homolog). The ancient conserved domain (ACD) has 55.3% of AA identity and 83.3% of homology between all mouse and human ACDPs (Fig. 2). The ACD domain is evolutionarily conserved in divergent species ranging from bacteria, yeast, C. elegans, D. melanogaster, mouse to human (Fig. 3). Particularly, as shown in Fig. 3, Acdp proteins showed very strong AA homology to bacteria CorC protein (35% AA identity with 55% homology), which is involved in magnesium and cobalt efflux [7]. High AA homology was also observed between the Acdp proteins and the yeast Amip3 protein (35% AA identity with 56% homology). The Amip3 is likely to be a homologous to the bacteria CorC protein. The Amip3 mutants confer resistance to copper toxicity (Personal communication with Dr. V.C. Culotte, John Hopkins Bloomberg, School of Public Health). The evolutionary relationships among those proteins are illustrated by a phylogenetic tree constructed based on the AA homology of proteins (Fig. 4).

We found that all mouse Acdp members contain four distinct transmembrane domains (Fig. 5), two CBS domains and a DUF21 domain that are found in bacteria CorC and yeast Amip3 proteins. CBS domains are small intracellular
modules that are mostly found in 2 or four copies within a protein. Pairs of CBS domains dimerise to form a stable globular domain [8]. DUF21 (CD: pfam01959.9) is a newly defined domain with unknown function. This domain is a transmembrane region and found to be located in the N-terminus of the proteins adjacent to two intracellular CBS domains http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=pfam01595. A cNMP-binding domain (cyclic nucleotide-monophosphate-binding domain) was found in all Acdp members.

In addition, Acdp1 contains an Alanine-rich region (2–10: AAAAAAAA), a Leucine-rich region (204–257: LLRVL-PRLYGPGGDLPPAWLRLAGALLLALSALF SGLRSLLLSDLPLVERVL), a Proline-rich region (78–130: PGPVPAAPVPAASLA PGENGTGWAPRLVFEIPEPPGAG- GAAPAVPVRPGP), and two amidation sites (917–920: MGKK; 926–929: SGRK). Acdp2 has a glycine-rich region (201–222: GAGGSGASGTVGKGAGVAG). Acdp3 possesses a large alanine-rich region (2–261) and a large leucine-rich region (201–299). Acdp4 contains a leucine zipper pattern (185–206: LVMVLVLSGIFSGLNGLMAL) and an amidation site (7–10: GRRR).

Table 1: Nucleotide and amino acid homologies (%) between human ACDP and mouse Acdp members.

| Acdp  | Acdp1 | Acdp2 | Acdp3 | Acdp4 |
|-------|-------|-------|-------|-------|
| DNA in coding region | 89    | 91    | 86    | 87    |
| AA identity | 94    | 97    | 86    | 89    |
| AA homology | 98.8  | 99.4  | 94    | 96    |
| 5' UTR  | 85    | 95    | 80    | 80    |
| 3' UTR  | 40    | 55    | 50    | 90    |
Antibody production, Western results and subcellular localization

Peptides from Acdp1 N- (TSFLLRVYFQPGPPATAAPVP-SPT) and C- (TQQLTLSPAAVPTR) terminuses, conserved peptide from ACD domain of Acdp1 (HNIVDILFVKDLLFVDPDDCPTLLTVTRF) were commercially synthesized (Sigma Genosys). These antigenic sites were predicted by software from Sigma Genosys and polyclonal antibodies for each peptide were produced by immunizing rabbits. To test the specificity of the antibodies, we conducted Western-blot analysis of mouse brain tissue extracts. As shown in Fig. 6A, the antibody produced by C-terminal peptide specifically detected Acdp1 (lane 3). The antibody generated by N-terminal peptide recognized Acdp4 in addition to Acdp1, although the reactivity to latter was significantly higher (Fig. 6A, lane 2). As expected, the antibody produced by the conserved sequence peptide detected all Acdp proteins (Fig. 6A, lane 1). To further determine the specificity of the antibody against the Acdp1 C-terminus, we analyzed extracts of HEK293, 3T3 and PC12 cells. The results are shown in Fig. 6B. Apparently, this antibody detected a specific band of Acdp1 in all cell lysates. Of note, shown in Fig. 6B are signals of 10 µg extracts of HEK293 cell lysates, 100 µg extracts of 3T3 and PC12 cell lysates. Thus, the expression levels of Acdp1 in these cell types vary a lot, with the highest expression in HEK293 cells. Nevertheless, these immunoblot results support our analysis of brain tissue extracts that the antibody against Acdp1 C-terminus specifically recognizes Acdp-1.

The specificity of the Acdp1 C-terminus antibody suggests the possibility of using it to localize Acdp-1 within cells. Since Northern blot revealed almost exclusive expression of Acdp1 in the brain, we examined its subcellular localization in hippocampus neurons isolated from mouse embryos. The neurons were cultured on glass coverslips coated with a confluent monolayer of mouse cortical astrocytes in dishes. Immunostaining was using the Acdp1 C-terminus specific antibody. Confocal imaging revealed that Acdp1 is predominantly localized on the plasma membrane. A series of sections of a cell at the thickness of 0.5 micrometer clearly showed membrane location of Acdp1-immunoreactivity (Fig. 7), which is consistent with the observation of transmembrane domains within the Acdp proteins.

Discussion

Although the exact functions of the ACDP gene family are not yet elucidated, several lines of evidence suggest that ACDP genes may play an important role in biological processes. First, these genes are evolutionarily conserved in many phylogenetically divergent species; Second, multiple genes are present in a species; Third, these genes are
Figure 3
Amino acid sequence alignment showing the conservation of ACD domain in various species. Amip3 is a protein from *Saccharomyces cerevisiae* (NP_014581). CanG is a protein from *Candida glabrata* (AAF33142). NeuC (EAA31204) is a hypothetical protein from *Neurospora crassa*. DroM is a gene product from *D. melanogaster*. The accession number for this gene is CG40084 in BDGP (Berkeley Drosophila Genome Project). AnoG represents a protein from the *anopheles gambiae* str. (EAA01004). CaeE (AAK77203) is a hypothetical protein from the *Caenohabditis elegans*. CorC represents bacteria magnesium and cobalt efflux protein from the *Shewanella oneidensis*. XyFD is a hypothetical protein from the *Xylella fastidiosa* Dixon (ZP_00038107).

Figure 4
Phylogenetic tree showing relationships among proteins containing the ACD domain from figure 2 and 3. The phylogenetic tree was constructed according to the calculation of the best match for the selected sequences. Abbreviations for each protein are the same as presented in figure 3.
ubiquitously expressed throughout development and adult tissues (unpublished data). The cloning and characterization of the mouse Acdp gene family are a very important step towards the elucidation of the functions of this multigene family.

Sequence homology analyses revealed that Acdp proteins shared very strong AA homologies to the bacteria CorC protein and yeast Amip3 protein. CorA, B, C and D belong to a protein family involved in both influx and efflux of magnesium and cobalt. It has been shown that CorA mutants confer resistance to cobalt toxicity [9]. Amip3 appears to be a homologous to CorC protein which is involved in efflux of magnesium and cobalt in bacteria. Amip3 mutants confer resistant to copper toxicity. Acdp proteins also possess the domains that are found in bacteria CorC and yeast Amip3 proteins, such as the CBS domains, DUF21 domain and transmembrane domains. In addition, a cNMP-binding domain was found in all Acdp proteins, which is usually present in ion channels and cNMP-dependent kinases [10-13]. Using antibody produced by Acdp1 C-terminal peptide, we have shown that Acdp1 is predominantly localized on the plasma membrane in hippocampus neurons. In our previous study, we found human ACDP proteins are predominantly localized in the nucleus in HeLa cells [1]. The discrepancy for Acdp localization in neuronal cells could be caused by non-specificity for previous antibody which was produced by the recombinant ACD domain, or different cells used. In current study, the Acdp1 C-terminus antibody only recognizes Acdp1 in brain tissue extracts as well as in HEK293, 3T3 and PC12 cell lysates, suggesting the specificity of the antibody. Our observations suggested that Acdp might be involved in ion transport in mammalian cells. However, more detailed functional studies are needed to demonstrate the real functions of these proteins.

**Conclusions**

Our previous work has described human ACDP genes [1]. The new information of the current work includes: 1). It is the first time to report the existence of multiple Acdp
genes in other mammals in addition to human, while Acdp appears to be a single copy gene in lower organisms such as in *C. elegans*, yeasts and bacteria. We have also suggested the evolutionary relationships of Acdp genes in different species (phylogenetic tree); 2). Molecular cloning and characterization of murine Acdp gene family are essential for study of this novel gene family in animal model, e.g. for generation of knockout or transgenic models; 3). We have demonstrated both DNA and amino acid conservations between mouse and human for each Acdp gene and the whole Acdp gene family, which provide important information for the possibility of functional redundancy or overlap between Acdp members; 4). We have generated antibodies specific for Acdp1 and all Acdp proteins. The Acdp1 C-terminus antibody appears to specifically recognize Acdp1. Using this antibody, we have demonstrated that Acdp1 is predominantly localized on the plasma membrane in hippocampus neurons. These results represent an important step towards the characterization of functions for the Acdp gene family.

**Methods**

**cDNA cloning**

cDNA cloning of the Acdp gene family was performed based on human homologue sequences as previously reported [2]. Briefly, the human *ACDP* cDNAs and predicted AA sequences were used to search the mouse EST database for EST markers corresponding to each Acdp member. A forward primer within the human *ACDP* 5' cDNA coding region (after start codon) and a mouse reverse primer from the mouse EST marker were used to amplify the homologue sequence from mouse cDNA at very low annealing temperature (45–50°C). A nested PCR using an inside reverse primer from the mouse EST sequence and the same human forward primer was then carried out to amplify the specific mouse gene from the first round PCR products at high annealing temperature (62°C). The expected PCR products were directly excised from agarose gel and sequenced by an ABI377 automatic sequencer. The sequence was further confirmed using a forward primer from newly identified sequence and a
reverse primer from known sequence. Once most of the coding sequences were identified, partial sequence of exon 1 and the 5' UTR sequences were obtained by BAC DNA sequencing.

**Northern blot analyses**

Multiple Choice Northern Blot filters containing 12 different mouse tissues were purchased from Origene. The filters were probed for each Acdp gene with a PCR fragment (around 350 bp) from last exon and the 3' untranslated region labeled with α-32P dCTP using the random primer extension system (Life Technologies). Hybridizations were carried out overnight. The filters were washed twice with washing buffer I (2×SSC, 0.1% SDS) at 42°C for 15-min, and then washed twice with washing buffer II (0.25×SSC, 0.1% SDS) at 65°C for 15-min. Washed filters were exposed to X-ray films for overnight or longer [3,4].

**Antibody production and Western blot analyses**

Peptides linked to KLH (keyhole limpet hemacyanin) from Acdp1 N- and C-terminals and the conserved domain (ACD) were used for generation of antibodies specifically for Acdp1 and all Acdp members as reported, respectively [5]. Western blots were carried out Using ECL (PIERCE) as described previously [1]. The membranes were washed extensively after incubation with primary and secondary antibodies and were then developed with X-ray films with optimal exposure time.

**Chromosome localization**

The T31 mouse radiation hybrid panel from Research Genetics was used to map the chromosome location of each Acdp member. Primers from 3' UTR of each Acdp member were used to amplify the 100 radiation hybrid clones representing the mouse genome. The data were submitted to the Jackson Laboratory Mouse Radiation Hybrid Database for analysis.

**Sequence analyses**

Sequence assembly was performed with program Sequencher (Gene Codes Corp). Protein and DNA homology searches were carried out with tblastn, tblastx, blastp and blastn programs [6]. Multiple sequence alignments were performed with GeneDoc and pairwise sequence alignment [7]. The data were submitted to the GenBank database [8]. Multiple programs including BCM Search Launcher [9], ProfileScan [10], ExPASy [11] and 3Dpssm [12] were used for searching sequence features of known protein. Phylogenetic tree was constructed by Clustalw program (version 1.81) using UPGMA (Unweighted Pair Group Method using Arithmetic averages) algorithm [13].

**Figure 7**

Subcellular localization of Acdp1 in hippocampus neurons. A series of confocal images from a cultured neuron stained with an anti-Acdp1 antibody. The step of each imaging section is 0.5 µm, from the surface of the neuron (0 µm) to the middle plan (4.5 µm).
Neuronal cell preparation and immunostaining

Hippocampal neuron cultures were prepared as previously reported [6]. In brief, the hippocampi were dissected out from mouse embryos at 16 days in utero. The tissues were then incubated for 20 min at 37°C in MEM (minimum essential medium) modified for suspension culture (Life Technologies) plus 0.25% trypsin (Life Technologies). The dissociated hippocampal neurons were plated on glass coverslips coated with a confluent monolayer of mouse cortical astrocytes obtained as described below. The neurons were maintained at 37°C in a humidified atmosphere with 5% CO₂. Cortical astrocytes dissociated from newborn mouse cortices were grown in culture flasks at 37°C in a humidified atmosphere with 5% CO₂ until confluent. The cells were exposed to 10⁻⁵ M cytosine arabinoside (Sigma) and cultured for additional 12–24 hrs at 37°C. After remove of the media with cellular debris, the cells were used for coating coverslips.

For immunostaining, neuronal cells on the coverslips were first fixed in PBS containing 4% paraformaldehyde (PFA) for 12 hrs at 4°C and then incubated in a solution containing 4% PFA and 0.4% Triton X-100 at 4°C for 1 hr. After washing with PBS three times, the cells were incubated with a blocking solution containing 1:30 normal goat serum, and subsequently incubated with a rabbit polyclonal anti-ACDP antibody (1:3000) overnight at 4°C. After extensive washing with 1% goat serum PBS solution, the cells were incubated with an Alexa 488 conjugated secondary antibody (1:100 in 1% goat serum PBS solution, Molecular Probes) for 3 hrs at room temperature. Following final washes with 1% goat serum PBS solution, the neuronal cells on the coverslips were covered with a glycerol-based anti-photobleach medium. The cells were viewed under a confocal microscope (Carl Zeiss). Images were captured with a CCD camera and acquired by the Scion Image software (Scion Corporation, Frederick, MD).

Gene bank accession number

The cDNA sequences for the Acdp gene family have already been deposited in gene bank with accession numbers AF202994 (Acdp1), AF216961 (Acdp2), AF216964 (Acdp3) and AF216963 (Acdp4).

Authors’ contributions

This study was designed by CYW and JXS. The study was performed as follows: CYW, PY, JDS and HA contributed to the molecular cloning part of the manuscript. CYW and SP were responsible for the Acdp antibody production. JGG and JLG did immunostaining for Acdp1 localization in neuronal cells. CYW, DG and ZD carried out the Western analyses.

Acknowledgements

This work was partly supported by a CIGP grant to C.Y.W.

References

1. Wang CY, Shi JD, Yang P, Kumar PG, Li QZ, Run QG, Su YC, Scott HS, Kao KJ, She JX: Molecular cloning and characterization of a novel gene family of four ancient conserved proteins (ACDP). Gene 2003, 306:37-44.
2. Wang CY, Shi JD, Davoodi-Semiromi A, She JX: Cloning of Aire, the mouse homologue of the autoimmune regulator (AIRE) gene responsible for autoimmune polyglandular syndrome type 1 (APS1). Genomics 1999, 55:322-326.
3. Li QZ, Wang CY, Shi JD, Run QG, Eckersley S, Davoodi-Semiromi A, Kumar T, Gu Y, Lian W, Wu D, She JX: Molecular cloning and characterization of the mouse and human TUSP gene, a novel member of the tubby superfamily. Gene 2001, 273:275-284.
4. Ruan QQ, Wang CY, Shi JD, She JX: Expression and alternative splicing of mouse autoimmune regulator gene (Aire). J Autoimmun 1999, 13:307-313.
5. Kumar PG, Laloraya M, Wang CY, Ruan QG, Davoodi-Semiromi A, Kao KJ, She JX: The autoimmune regulator (AIRE) is a DNA-binding protein. J Biol Chem 2001, 276:41357-41364.
6. Gu JG, Albuquerque C, Lee CJ, MacDermott AB: Synaptic strengthening through activation of Ca2+-permeable AMPA receptors. Nature 1996, 381:793-796.
7. Kalhre DG, Lawyer CH, Maguire ME: The CorA magnesium transporter gene family. Microb Comp Genomics 1998, 3:151-169.
8. Bateman A: The structure of a domain common to archaeabacteria and the homocystinuria disease protein. Trends Biochem Sci 1997, 22:12-13.
9. Gibson MM, Bago DA, Miller CG, Maguire ME: Magnesium transport in Salmonella typhimurium: the influence of new mutations conferring Co2+ resistance on the CorA Mg2+ transport system. Mol Microbiol 1991, 5:2753-62.
10. Cochet C, Souvignet C, Keramidas M, Chambaz EM: Altered catalytic properties of protein kinase C in phorbol ester treated cells. Biochem Biophys Res Commun 1986, 134:1031-1037.
11. Shabb JB, Corbin JD: Cyclic nucleotide-binding domains in proteins having diverse functions. J Biol Chem 1992, 267:5723-5726.
12. Kaupp UB: The cyclic nucleotide-gated channels of vertebrate photoreceptors and olfactory epithelium. Trends Neurosci 1991, 14:150-157.
13. Weber IT, Shabb JB, Corbin JD: Predicted structures of the cGMP binding domains of the cGMP-dependent protein kinase: a key alanine/threonine difference in evolutionary divergence of cAMP and cGMP binding sites. Biochimica 1989, 28:6122-6127.
14. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 1994, 22:4673-4680.