

SESSION 7. PRACTICE

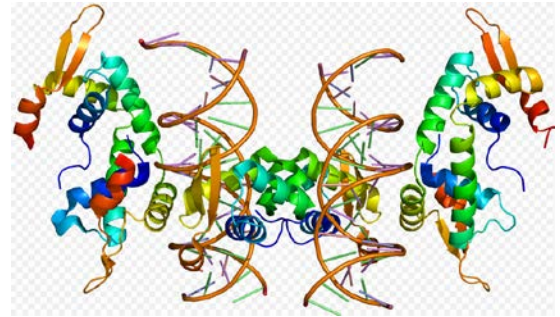
Evolution: What makes us human?

A protein related to human speech: FOXP2

- **A protein related to human speech: Foxp2**
- Foxp2 contains a forkhead box DNA binding domain, is expressed in CNS during development, and is likely to regulate a number of genes of importance for brain function.
- Foxp2 is fairly well conserved in all vertebrates and mutations in the Foxp2 gene cause speech deficiency.

R553H (arginine -> histidine) in the Foxp2 is a.a. change in the DNA binding domain and **R328X** (arginine -> any amino acid) is nonsense mutation.

- Both copies are required for normal function (**dominant phenotype**).



FOXP2 in other animals

- FOXP2 seems to be related in synaptic plasticity (for motor-skill learning).
- Bird song has similarities and differences to human speech.
- Expression of FOXP2 in zebra finch was reduced to 50% of its normal level and the bird is not efficient for sound-imitation.

Basic Shell Commands

```
$ cd [User_Folder]  
$ mkdir week7_7  
$ cd week7_7
```

Comparing FOXP2 in different animals

FOXP2_HUMAN	<i>Homo sapiens</i>
FOXP2_GORGO	<i>Gorilla gorilla</i>
FOXP2_MACMU	<i>Macaca mulatta</i> (Rhesus macaque)
FOXP2_PANTR	<i>Pan troglodytes</i> (Chimpanzee)
FOXP2_HYLLA	<i>Hylobates lar</i> (Common gibbon)
FOXP2_PONPY	<i>Pongo pygmaeus</i> (Bornean orangutan)
FOXP2_MOUSE	<i>Mus musculus</i> (Mouse)
FOXP2_XENLA	<i>Xenopus laevis</i> (African clawed frog)



```
$ ln -s /home/biguser/tutor/Week6/swissprot.* .
```

```
[biguser@biglab-master session_7]$ ls  
swissprot.phr  swissprot.pnd  swissprot.psd  swissprot.psq  
swissprot.pin  swissprot.pni  swissprot.psi
```

Comparing FOXP2 in different animals

```
$ vi seqids.txt
```

```
FOXP2_HUMAN  
FOXP2_GORGO  
FOXP2_MACMU  
FOXP2_PANTR  
FOXP2_HYLLA  
FOXP2_PONPY  
FOXP2_MOUSE  
FOXP2_XENLA
```

Comparing FOXP2 in different animals

```
$ blastdbcmd -entry FOXP2_HYLLA -db swissprot -long_seqids > FOXP2_HYLLA.fa
```

```
$ blastdbcmd -entry_batch seqids.txt -db swissprot -long_seqids > foxp2.fa
```

```
[biguser@biglab-master session_7]$ ls  
FOXP2_HYLLA.fa  seqids.txt  swissprot.pin  swissprot.pni  swissprot.psi  
foxp2.fa       swissprot.phr  swissprot.pnd  swissprot.psd  swissprot.psq
```

Comparing FOXP2 in different animals

```
$ clustalw2 foxp2.fa
```

CLUSTAL 2.1 Multiple Sequence Alignments

Sequence format is Pearson

Sequence 1:	gi 17432967 sp O15409.2 FOXP2_HUMAN	715	aa
Sequence 2:	gi 51701430 sp Q8MJ99.1 FOXP2_GORGO	713	aa
Sequence 3:	gi 51701429 sp Q8MJ97.1 FOXP2_MACMU	714	aa
Sequence 4:	gi 38503046 sp Q8MJA0.1 FOXP2_PANTR	716	aa
Sequence 5:	gi 62286912 sp Q5QL03.1 FOXP2_HYLLA	713	aa
Sequence 6:	gi 146345420 sp Q8MJ98.3 FOXP2_PONPY	713	aa
Sequence 7:	gi 51704301 sp P58463.2 FOXP2_MOUSE	714	aa
Sequence 8:	gi 82227296 sp Q4VYS1.1 FOXP2_XENLA	706	aa

There are 7 groups
Start of Multiple Alignment

Aligning..

Group 1:	Sequences: 2	Score:15411
Group 2:	Sequences: 3	Score:15416
Group 3:	Sequences: 4	Score:15408
Group 4:	Sequences: 2	Score:15421
Group 5:	Sequences: 3	Score:15400
Group 6:	Sequences: 7	Score:15379
Group 7:	Sequences: 8	Score:14978

Alignment Score 118874

CLUSTAL-Alignment file created [foxp2.aln]

Start of Pairwise alignments
Aligning..

Sequences (1:2)	Aligned.	Score: 99
Sequences (1:3)	Aligned.	Score: 99
Sequences (1:4)	Aligned.	Score: 99
Sequences (1:5)	Aligned.	Score: 99
Sequences (1:6)	Aligned.	Score: 99
Sequences (1:7)	Aligned.	Score: 99
Sequences (1:8)	Aligned.	Score: 95
Sequences (2:3)	Aligned.	Score: 100
Sequences (2:4)	Aligned.	Score: 100
Sequences (2:5)	Aligned.	Score: 98
Sequences (2:6)	Aligned.	Score: 98
Sequences (2:7)	Aligned.	Score: 99
Sequences (2:8)	Aligned.	Score: 95
Sequences (3:4)	Aligned.	Score: 100
Sequences (3:5)	Aligned.	Score: 100
Sequences (3:6)	Aligned.	Score: 99
Sequences (3:7)	Aligned.	Score: 98
Sequences (3:8)	Aligned.	Score: 95
Sequences (4:5)	Aligned.	Score: 100
Sequences (4:6)	Aligned.	Score: 99
Sequences (4:7)	Aligned.	Score: 99
Sequences (4:8)	Aligned.	Score: 95
Sequences (5:6)	Aligned.	Score: 99
Sequences (5:7)	Aligned.	Score: 99
Sequences (5:8)	Aligned.	Score: 95
Sequences (6:7)	Aligned.	Score: 99
Sequences (6:8)	Aligned.	Score: 95
Sequences (7:8)	Aligned.	Score: 95

Guide tree file created: [foxp2.dnd]

Comparing FOXP2 in different animals

foxp2.aln

CLUSTAL 2.1 multiple sequence alignment

```
gi|51701430|sp|Q8MJ99.1|FOXP2_      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|51701429|sp|Q8MJ97.1|FOXP2_      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|38503046|sp|Q8MJA0.1|FOXP2_      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|17432967|sp|O15409.2|FOXP2_      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|62286912|sp|Q5Q0.1|FOXP2_        MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|146345420|sp|Q8MJ98.3|FOXP2_      MMQESVTEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|51704301|sp|P58463.2|FOXP2_      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|82227296|sp|Q4VYS1.1|FOXP2_      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
**** *
gi|51701430|sp|Q8MJ99.1|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|51701429|sp|Q8MJ97.1|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|38503046|sp|Q8MJA0.1|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|17432967|sp|O15409.2|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|62286912|sp|Q5Q0.1|FOXP2_        HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|146345420|sp|Q8MJ98.3|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|51704301|sp|P58463.2|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|82227296|sp|Q4VYS1.1|FOXP2_      HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
*****
gi|51701430|sp|Q8MJ99.1|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|51701429|sp|Q8MJ97.1|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|38503046|sp|Q8MJA0.1|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|17432967|sp|O15409.2|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|62286912|sp|Q5Q0.1|FOXP2_        PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|146345420|sp|Q8MJ98.3|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|51704301|sp|P58463.2|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
gi|82227296|sp|Q4VYS1.1|FOXP2_      PQMQQILQQQVLSPOQLQALLQQQAVMLQQQQLQEFYKKQEQQLHLQL
*****
```

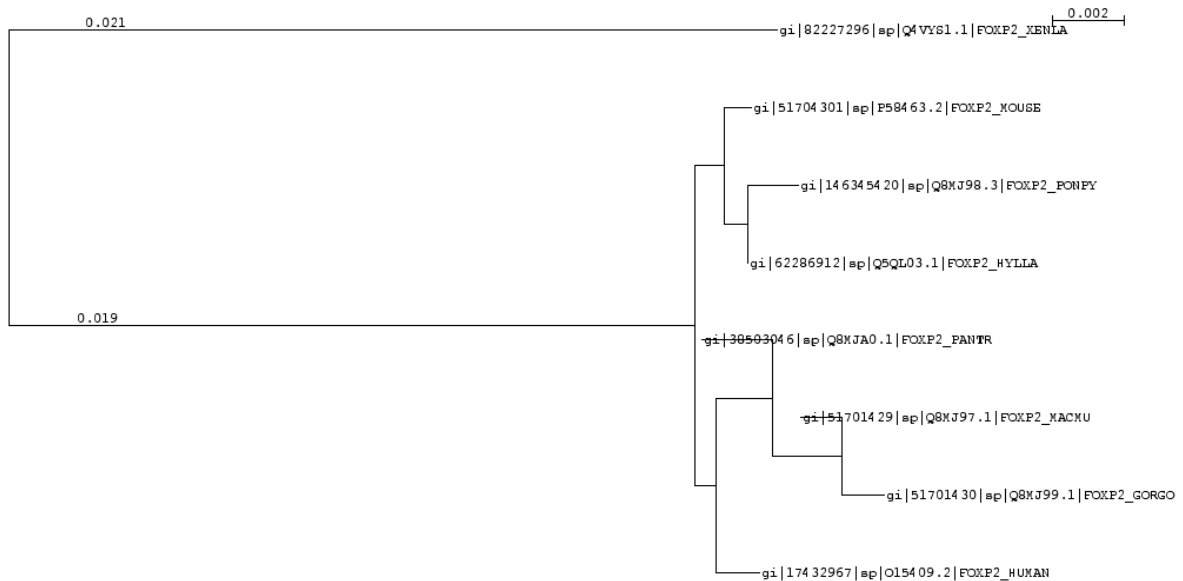
foxp2.dnd

```
(
(
(
gi|17432967|sp|O15409.2|FOXP2_HUMAN:0.00122,
(
(
gi|51701430|sp|Q8MJ99.1|FOXP2_GORGO:0.00117,
gi|51701429|sp|Q8MJ97.1|FOXP2_MACMU:-0.00117)
:0.00196,
gi|38503046|sp|Q8MJA0.1|FOXP2_PANTR:-0.00196)
:0.00157)
:0.00059,
gi|82227296|sp|Q4VYS1.1|FOXP2_XENLA:0.04051)
:0.00083,
(
gi|62286912|sp|Q5Q0.1|FOXP2_HYLLA:0.00000,
gi|146345420|sp|Q8MJ98.3|FOXP2_PONPY:0.00140)
:0.00066,
gi|51704301|sp|P58463.2|FOXP2_MOUSE:0.00074);
```

- * -- all residues or nucleotides in the
- : -- conserved substitutions have been
- . -- semi-conserved substitutions have
- no match.

Phylogenetic tree from dendrogram

- NJPlot: <http://doua.prabi.fr/software/njplot>



Identification of mutation specific to human

```
$ clustalw2 foxp2.fa -output=fasta  
$ less foxp2.fasta
```

```
>gi|51701430|sp|Q8MJ99.1|FOXP2_GORGO  
MMQESATETISNSSMNQNGMSTLSSQLDAGSRDRSSSGDT  
SSEVSTVELLHLQQQALQAARQLLLQQQTSGLKSPKSSD  
KQRPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPOQLQA  
LLQQQAVMLQQQQLQEFYKKQQEQHLQLLQQQQQQQQQ  
QQQQQQQQQQQQQQQQQQQQQQQQQQQQ--HPGKQAKE  
QQQQQQQQQLAAQQLVFQQQLLQMQLQQQHLLSLQRQ  
GLISIPPGQAALPVQSLPQAGLSPAETQQLWKEVTGVHSM  
EDNGIKHGGLDLTTNNSSTTSSTTSKASPPITHHSIVNG  
QSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICED  
FGQFLKHLNNEHALDDRSTAQCRVQMVVQQLEIQLSKER  
ERLQAMMTHLHMRPSEPKPSKPLNLVSSVTMSKNMLET  
PQSLPQTPPTAPVTPITQGPSVITPASVPNVGAIARRRH  
SDKYNIPMSSEIAPNYEFYKNADVRPPFTYATLIRQAIME  
SSDRQLTLNEYISWFTRTFAYFRNAAWKNAVRHNLSLH  
KCFVRVENVKGAWTVDEVEYQKRRSQKITGSPTLVKNIP  
TSLGYGAALNASLQAALAESSLPLLSNPGLINNASSGLLQ  
AVHEDLNGSLDHIDSNGNSSPGCSPQPHIHSIHVKEEPI  
AEDEDCPMSLVTTANHSPELEDDREIEEPLSEDL  
>gi|51701429|sp|Q8MJ97.1|FOXP2_MACMU
```

만약에 linux 상에서 작업할 수 없거나 결과를 만들지 못했으면

In `~$ /home/biguser/tutor/Week7/foxp2.fasta .`

를 실행시키고 foxp2.fasta 파일 링크가 형성됐는지 확인하세요!

Identification of mutation specific to human

□ foxp2.py

	POSITION 1	POSITION 2
HUMAN	A	T
GORGO	A	A
MACMU	A	C
PANTR	A	C
HYLLA	A	A
PONPY	A	G

```
import re

inFile = open('foxp2.fasta','r')

nonhuman = dict()

id = ''
seq = ''

for line in inFile.readlines():
    line = line.strip()
    #print line
    if re.search('^>', line):
        if id != '':
            if not "HUMAN" in id:
                nonhuman[id[1:]] = seq
            else:
                id_human = id[1:]
                seq_human = seq
            id = line
            seq = ''
        else:
            id = line
    else:
        seq += line

if not "HUMAN" in id:
    nonhuman[id[1:]] = seq
else:
    id_human = id[1:]
    seq_human = seq

inFile.close()

for i in xrange(0, len(seq_human)):
    unique = True
    for id in nonhuman.keys():
        human_seq_posi = seq_human[i]
        nonhuman_seq_posi = nonhuman[id][i]
        if human_seq_posi == nonhuman_seq_posi:
            unique = False
    if unique:
        pos = i + 1
        print "At position ", pos
        aa = human_seq_posi
        print id_human, '#t', aa
        for ID in nonhuman.keys():
            print ID, '#t', nonhuman[ID][i]
```

Identification of mutation specific to human

□ results

At position 304

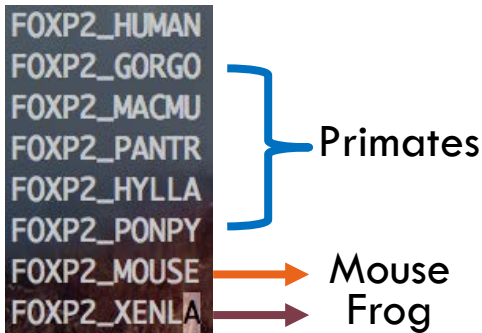
sp O15409 FOXP2_HUMAN	N
sp P58463 FOXP2_MOUSE	T
sp Q8MJ98 FOXP2_PONPY	T
sp Q4VYS1 FOXP2_XENLA	T
sp Q8MJ97 FOXP2_MACMU	T
sp Q8MJAO FOXP2_PANTR	T
sp Q5QLQ3 FOXP2_HVLLA	T
sp Q8MJ99 FOXP2_GORGO	T

Human-specific T303N

- T303N and N325S are positively selected for language and speech
- No N325S in our res. due to *Xenopus*

Exercise

- Modify `foxp2.py` such that you may identify
 1. positions where the human sequence is different from all other primates and mouse;
 2. positions where the human and primate sequences are identical but different to both mouse and frog.



Exercise - 1

```
At position 304
sp|O15409|FOXP2_HUMAN   N
sp|P58463|FOXP2_MOUSE  T
sp|Q8MJ98|FOXP2_PONPY  T
sp|Q8MJ97|FOXP2_MACMU  T
sp|Q8MJA0|FOXP2_PANTR  T
sp|Q5QL03|FOXP2_HYLLA  T
sp|Q8MJ99|FOXP2_GORGO  T
At position 326
sp|O15409|FOXP2_HUMAN   S
sp|P58463|FOXP2_MOUSE  N
sp|Q8MJ98|FOXP2_PONPY  N
sp|Q8MJ97|FOXP2_MACMU  N
sp|Q8MJA0|FOXP2_PANTR  N
sp|Q5QL03|FOXP2_HYLLA  N
sp|Q8MJ99|FOXP2_GORGO  N
```



```
import re

inFile = open('foxp2.fasta','r')

nonhuman = dict()

id = ''
seq = ''

for line in inFile.readlines():
    line = line.strip()
    #print line
    if re.search('<?>', line):
        if id != '':
            if not "HUMAN" in id:
                nonhuman[id[1:]] = seq
            else:
                id_human = id[1:]
                seq_human = seq
                id = line
                seq = ''
        else:
            id = line
    else:
        seq += line

if not "HUMAN" in id:
    nonhuman[id[1:]] = seq
else:
    id_human = id[1:]
    seq_human = seq

inFile.close()

for i in xrange(0, len(seq_human)):
    unique = True
    for id in nonhuman.keys():
        if not "XENLA" in id:
            human_seq_posi = seq_human[i]
            nonhuman_seq_posi = nonhuman[id][i]
            if human_seq_posi == nonhuman_seq_posi:
                unique = False

    if unique:
        pos = i + 1
        print "At position ", pos
        aa = human_seq_posi
        print id_human, "#t", aa
        for ID in nonhuman.keys():
            if not "XENLA" in ID:
                print ID, "#t", nonhuman[ID][i]
```

Exercise-2

```
At position 80
sp|O15409|FOXP2_HUMAN   D
sp|P58463|FOXP2_MOUSE  E
sp|Q8MJ98|FOXP2_PONPY   D
sp|Q4VYS1|FOXP2_XENLA   E
sp|Q8MJ97|FOXP2_MACMU   D
sp|Q8MJAD|FOXP2_PANTR   D
sp|Q5QL03|FOXP2_HYLLA   D
sp|Q8MJ99|FOXP2_GORGO   D
```



```
import re

inFile = open('foxp2.fasta', 'r')

nonhuman = dict()

id = ''
seq = ''

for line in inFile.readlines():
    line = line.strip()
    #print line
    if re.search('^>', line):
        if id != '':
            if not "HUMAN" in id:
                nonhuman[id[1:]] = seq
            else:
                id_human = id[1:]
                seq_human = seq
            id = line
            seq = ''
        else:
            id = line
    else:
        seq += line

if not "HUMAN" in id:
    nonhuman[id[1:]] = seq
else:
    id_human = id[1:]
    seq_human = seq

inFile.close()

for i in xrange(0, len(seq_human)):
    unique = True
    for id in nonhuman.keys():
        human_seq_posi = seq_human[i]
        if "XENLA" in id or "MOUSE" in id:
            if human_seq_posi == nonhuman[id][i]:
                unique=False
        else:
            if human_seq_posi != nonhuman[id][i]:
                unique=False
    if unique:
        pos = i + 1
        print "At position ", pos
        aa = human_seq_posi
        print id_human, '\t', aa
        for ID in nonhuman.keys():
            print ID, '\t', nonhuman[ID][i]
```


Assignment

- In the default output from the ClustalW program (a file named as foxp2.aln) there are asterisks (*) that indicate positions where the sequence is same in all sequences. Make a Python script to **count the total length and number of such positions** in all alignment.
- Output

```
Total length : 716 Number of asterisk (match) : 674
```
- Submit a word file w/ your python script and result.
- 10/24 월요일 오후 6시까지 제출

Assignment

In `-s/home/biguser/tutor/Session7/foxp2.aln`.

linux 환경에서 여러분의 directory 에서 실행시키고 해당 파일 link 가 형성 됐는지 확인. 해당 파일을 일어서 작업할 것

```
CLUSTAL 2.1 multiple sequence alignment
```

```
gi|51701430|sp|Q8MJ99.1|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|51701429|sp|Q8MJ97.1|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|38503046|sp|Q8MJA0.1|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|17432967|sp|O15409.2|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|62286912|sp|Q5QL03.1|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|146345420|sp|Q8MJ98.3|FOXP2 MMQESVTEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|51704301|sp|P58463.2|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELL
gi|82227296|sp|Q4VYS1.1|FOXP2_ MMQESATEITISNSSMNQNGMSTLSSQLDAGSRDGRSSSDTSSSEVSTVELL
*****

gi|51701430|sp|Q8MJ99.1|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|51701429|sp|Q8MJ97.1|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|38503046|sp|Q8MJA0.1|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|17432967|sp|O15409.2|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|62286912|sp|Q5QL03.1|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|146345420|sp|Q8MJ98.3|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|51704301|sp|P58463.2|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVIT
gi|82227296|sp|Q4VYS1.1|FOXP2_ HLQQQQALQAARQLLLQQQTSGLKSPKNEKQRPLQVPVSMAMMTPQVIT
*****
```