

# SESSION 7. EVOLUTION

What makes us human?



# Genetic differences b/w humans and chimpanzees

Advanced Human Abilities: **Advances spoken and written language**, **abstract-level and creative thinking** (concept generation), and **reasoning**.

- How are significant the genetic differences b/w human and primates?
- The genetic elements are likely to affect brain function?

**Genome sequencing** were completed in multiple species including human, chimpanzee, dog, pig, cow, ... and can be compared in similarity and differences b/w genomes.

# Genetic differences b/w humans and chimpanzees

## Humans vs Chimpanzees

35million single nucleotide differences / 3 billions ~ 1% differences

→ point mutations could affect regulation of transcription and processes such as splicing as well as affect protein sequence.

5 million indels ~ 3% difference

→ regulation of gene expression

Structural variations ~ ? % i.g., duplicated paralogs with a new function

→ Such mutations may also be of interest when learning about functions that are specific to humans

# Identification of genes with a function specific to humans

- Genes or regulatory elements that are conserved during vertebrate or mammalian evolution, but where the human sequences have evolved at an unexpectedly high rate.
- **Noncoding RNA, HAR1F** (human accelerated region) is specifically expressed in the developing human neocortex (involved in function such as conscious thought and language).

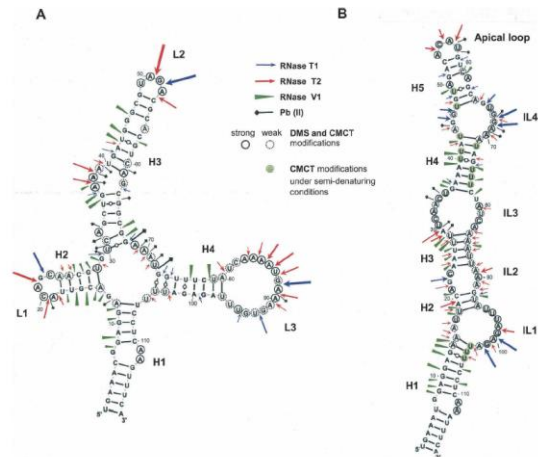
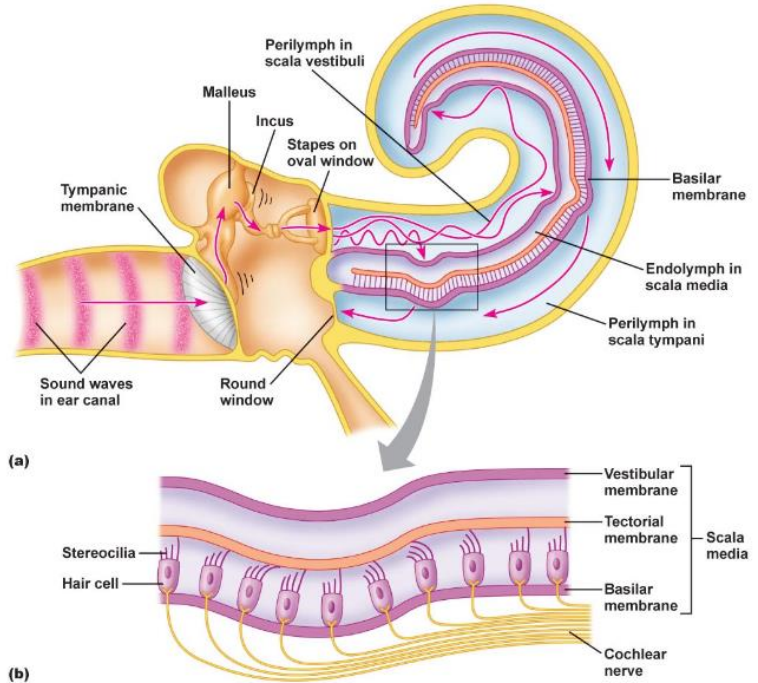


FIGURE 3. Two distinct experimentally supported secondary structure models for HAR1 RNAs. (A) The cloverleaf-like model of the human HAR1 RNA. (B) The chimpanzee HAR1 RNA adopts a hairpin structure. The length and thickness of the symbols represent the intensity of the cleavages. Bases reactive to DMS or CMCT under native conditions are circled; weak reactivities are depicted by dotted circles. Bases modified by CMCT under semi-denaturing conditions only are displayed with a green background. H, helix; IL, internal loop; L, loop.

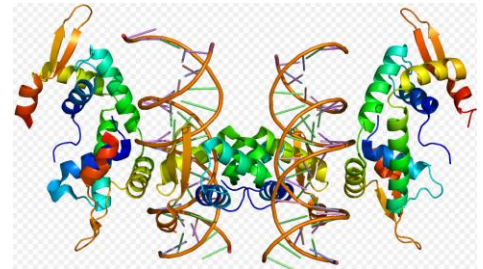
# Identification of genes with a function specific to humans

- **Alpha tectorin** is specifically developed in the human lineage and localized in the tectorial membrane of the inner ear.
- Specialized or adjusted to human-specific development of speech.



# A protein related to human speech: FOXP2

- A protein related to human speech: **FOXP2 gene**
- 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 vertebrates and mutations in the Foxp2 gene cause speech deficiency.
- **R553H** in the Foxp2 is aa change in the DNA binding domain and **R328X** is nonsense mutation.
- Both copies are required for normal function (**dominant phenotype**).



## FOXP2 in other animals

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



# Comparing FOXP2 in different animals

- We first have to know which mutations affect the function of FOXP2 in neural development, synaptic plasticity, and speech.
- Multiple sequence alignment identifies mutations across different vertebrates.

FOXP2\_HUMAN  
FOXP2\_GORGO (Gorilla)  
FOXP2\_MACMU (Rhesus macaque)  
FOXP2\_PANTR (Chimpanzee)  
FOXP2\_HYLLA (Common gibbon)  
FOXP2\_PONPY (Orangutan)  
FOXP2\_MOUSE (mouse)  
FOXP2\_XENLA (African frog)

foxp2.fa

```
FOXP2_GORGO      MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_MACMU     MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_PANTR     MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_HUMAN     MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_HYLLA     MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_PONPY     MMQESVITETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_MOUSE     MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
FOXP2_XENLA     MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTISSEVSTIVELLHLQQQQALQA
*****_*.....*
```



# Multiple sequence alignment

- Global alignment for more than two sequences
- **ClustalW**, Muscle, and T-coffee for MSA
- ClustalW: based on pairwise alignments of each pair – **Progressive approach**  
step-wise procedure
  - 1) all possible pairwise alignments
  - 2) guide-tree construction from pairwise alignments
  - 3) Single + Subalignment

```
[jwnam@biglab-master Session7]# clustalw2 foxp2.fa
```

```
CLUSTAL 2.1 Multiple Sequence Alignments
```

```
Sequence format is Pearson
Sequence 1: FOXF2_HUMAN      715 aa
Sequence 2: FOXF2_GORGO     713 aa
Sequence 3: FOXF2_MACMU     714 aa
Sequence 4: FOXF2_PANTR     716 aa
Sequence 5: FOXF2_HYLLA     713 aa
Sequence 6: FOXF2_PONPY     713 aa
Sequence 7: FOXF2_MOUSE     714 aa
Sequence 8: FOXF2_XENLA     706 aa
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]
```

```
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]
```

# MSA for FOXP2

## □ FOXP2.aln

LUSTAL 2.1 multiple sequence alignment

```
FOXP2_GORGO      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_MACMU      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_PANTR      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_HUMAN      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_HYLLA      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_PONPY      MMQESVTIETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_MOUSE      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSGDT SSEVSTVELLHLQQQQALQA
FOXP2_XENLA      MMQESATETISNSMNIQNGMSTLSSQLDAGSRDGRSSSDT SSEVSTVELLHLQQQQALQA
*****

FOXP2_GORGO      ARQLLQQQTSGLKSFKSSDKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_MACMU      ARQLLQQQTSGLKSFKSSDKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_PANTR      ARQLLQQQTSGLKSFKSSDKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_HUMAN      ARQLLQQQTSGLKSFKSSDKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_HYLLA      ARQLLQQQTSGLKSFKSSDKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_PONPY      ARQLLQQQTSGLKSFKSSDKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_MOUSE      ARQLLQQQTSGLKSFKSSEKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
FOXP2_XENLA      ARQLLQQQTSGLKSFKINNKQRLQVFPVSAMMT PQVITPQQMQQLLQQQVLSFQQQLQA
*****

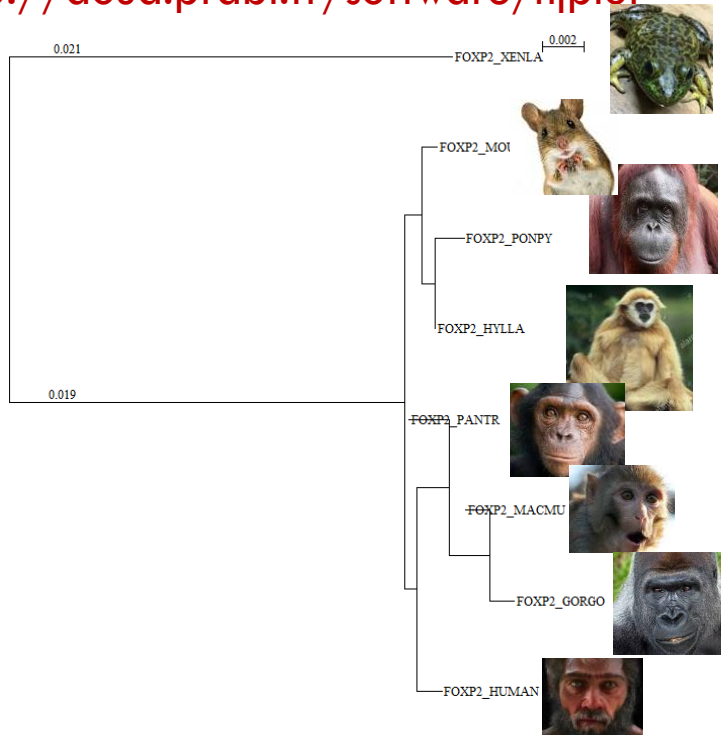
FOXP2_GORGO      LLQQQQAVMLQQQQQLQEFYKQKQEQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
FOXP2_MACMU      LLQQQQAVMLQQQQQLQEFYKQKQEQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
FOXP2_PANTR      LLQQQQAVMLQQQQQLQEFYKQKQEQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
FOXP2_HUMAN      LLQQQQAVMLQQQQQLQEFYKQKQEQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQQQ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

## □ FOXP2.dnd


```
(
(
(
FOXP2_HUMAN:0.00122,
(
(
FOXP2_GORGO:0.00117,
FOXP2_MACMU:-0.00117)
:0.00196,
FOXP2_PANTR:-0.00196)
:0.00157)
:0.00059,
FOXP2_XENLA:0.04051)
:0.00083,
(
FOXP2_HYLLA:0.00000,
FOXP2_PONPY:0.00140)
:0.00066,
FOXP2_MOUSE:0.00074);
```

# Phylogenetic tree from dendrogram

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



# foxp2.py: identification of mutation specific to human

clustalw2 foxp2.fa -output=fasta  foxp2.fasta

```
FOXFP2_GORGO
MMQESATETISNSMNNQGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELLHQQQALQA
ARQLLLQQQTSGLSPKSSDKQRLQVFPVSVAMMTQVITPQQMQQILQQQVLSFQQLQA
LLQQQAVMLQQQQLQEFYKQKQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQ
QQQQQQQQQQ---HFGKQAKEQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQHQHLLSLQRQ
GLISIPFGQAALFVQSLPQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTINSSST
TSSTTSKASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICED
FGQFLKHLNNEHALDDRSTAQCVRQMVVQVQLEIQLSKERERLQAMMTHLMRPFSEPKFS
PKPLMLVSVTMSKMLETSPQSLPQTPTTPTAFVTPITQGPVITPASVFNVAIRRRH
SDKYNIPMSSEIAFNVEFYKQADVRPPTVATLIRQAIMSSDRQLTLINEIYSWTRTFA
YFRNNAATWKNVRRHNSLHKCFVRVENVKGAVMTVDEVEYQKRRSQKITGSPTLVKNIP
TSLGYGAALNASLQAALAESSLPELLSNPGLINASSGLLQAVHEDLNGSLDHIDSNGNSS
PGCSFQPHIHSIHVKEEPVIAEDEDPCMSLVTTANHSPELEDDREIEEPLSEDL
>FOXFP2_MACMU
MMQESATETISNSMNNQGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELLHQQQALQA
ARQLLLQQQTSGLSPKSSDKQRLQVFPVSVAMMTQVITPQQMQQILQQQVLSFQQLQA
LLQQQAVMLQQQQLQEFYKQKQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQ
QQQQQQQQQQ---HFGKQAKEQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQHQHLLSLQRQ
GLISIPFGQAALFVQSLPQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTINSSST
TSSTTSKASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICED
FGQFLKHLNNEHALDDRSTAQCVRQMVVQVQLEIQLSKERERLQAMMTHLMRPFSEPKFS
PKPLMLVSVTMSKMLETSPQSLPQTPTTPTAFVTPITQGPVITPASVFNVAIRRRH
SDKYNIPMSSEIAFNVEFYKQADVRPPTVATLIRQAIMSSDRQLTLINEIYSWTRTFA
YFRNNAATWKNVRRHNSLHKCFVRVENVKGAVMTVDEVEYQKRRSQKITGSPTLVKNIP
TSLGYGAALNASLQAALAESSLPELLSNPGLINASSGLLQAVHEDLNGSLDHIDSNGNSS
PGCSFQPHIHSIHVKEEPVIAEDEDPCMSLVTTANHSPELEDDREIEEPLSEDL
>FOXFP2_PANTR
MMQESATETISNSMNNQGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELLHQQQALQA
ARQLLLQQQTSGLSPKSSDKQRLQVFPVSVAMMTQVITPQQMQQILQQQVLSFQQLQA
LLQQQAVMLQQQQLQEFYKQKQQLHLQLLQQQQQQQQQQQQQQQQQQQQQQQQQQ
QQQQQQQQQQHFGKQAKEQQQQQQQQQLAAQQLVFQQQLLQMQQLQQQHQHLLSLQRQ
GLISIPFGQAALFVQSLPQAGLSPAEIQQLWKEVTGVHSMEDNGIKHGGLDLTINSSST
TSSTTSKASPPITHHSIVNGQSSVLNARRDSSSHEETGASHTLYGHGVCKWPGCESICED
FGQFLKHLNNEHALDDRSTAQCVRQMVVQVQLEIQLSKERERLQAMMTHLMRPFSEPKFS
PKPLMLVSVTMSKMLETSPQSLPQTPTTPTAFVTPITQGPVITPASVFNVAIRRRH
SDKYNIPMSSEIAFNVEFYKQADVRPPTVATLIRQAIMSSDRQLTLINEIYSWTRTFA
YFRNNAATWKNVRRHNSLHKCFVRVENVKGAVMTVDEVEYQKRRSQKITGSPTLVKNIP
TSLGYGAALNASLQAALAESSLPELLSNPGLINASSGLLQAVHEDLNGSLDHIDSNGNSS
PGCSFQPHIHSIHVKEEPVIAEDEDPCMSLVTTANHSPELEDDREIEEPLSEDL
```

# foxp2.py: identification of mutal

At position	304
FOXP2_HUMAN	N
FOXP2_PANTR	T
FOXP2_MOUSE	T
FOXP2_GORGO	T
FOXP2_MACMU	T
FOXP2_PONPY	T
FOXP2_HYLLA	T
FOXP2_XENLA	T

## → Human-specific T303N

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

```
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
            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 = 1
    for id in nonhuman.keys():
        if seq_human[i] == nonhuman[id][i]:
            unique = 0

    if unique:
        pos = i + 1
        print "At position ", pos
        aa = seq_human[i]
        print id_human, '\t', aa
        for ID in nonhuman.keys():
            print ID, '\t', nonhuman[ID][i]
```