Pakistan J. Zool., vol. 50(2), pp 417-428, 2018. DOI: http://dx.doi.org/10.17582/journal.pjz/2018.50.2.417.428

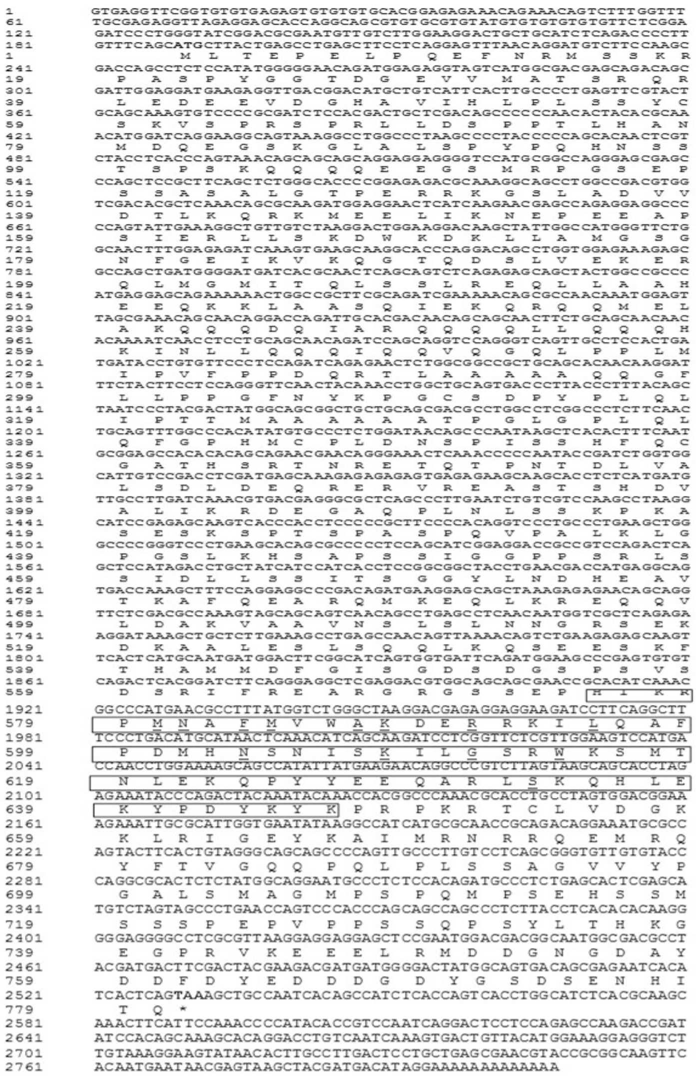
Supplementary Material

Identification, Molecular Characterization and Expression Pattern Analysis of *SoxD* Subgroup Genes in Yellow River Carp (*Cyprinus carpio*)

RuiHua Zhang, YongFang Jia, TingTing Liang, QianWen Yang, QiYan Du and ZhongJie Chang*
Molecular and Genetic Laboratory, College of Life Science, Henan Normal University, 46# East of Construction Road, Xinxiang, Henan 453007, P.R. China

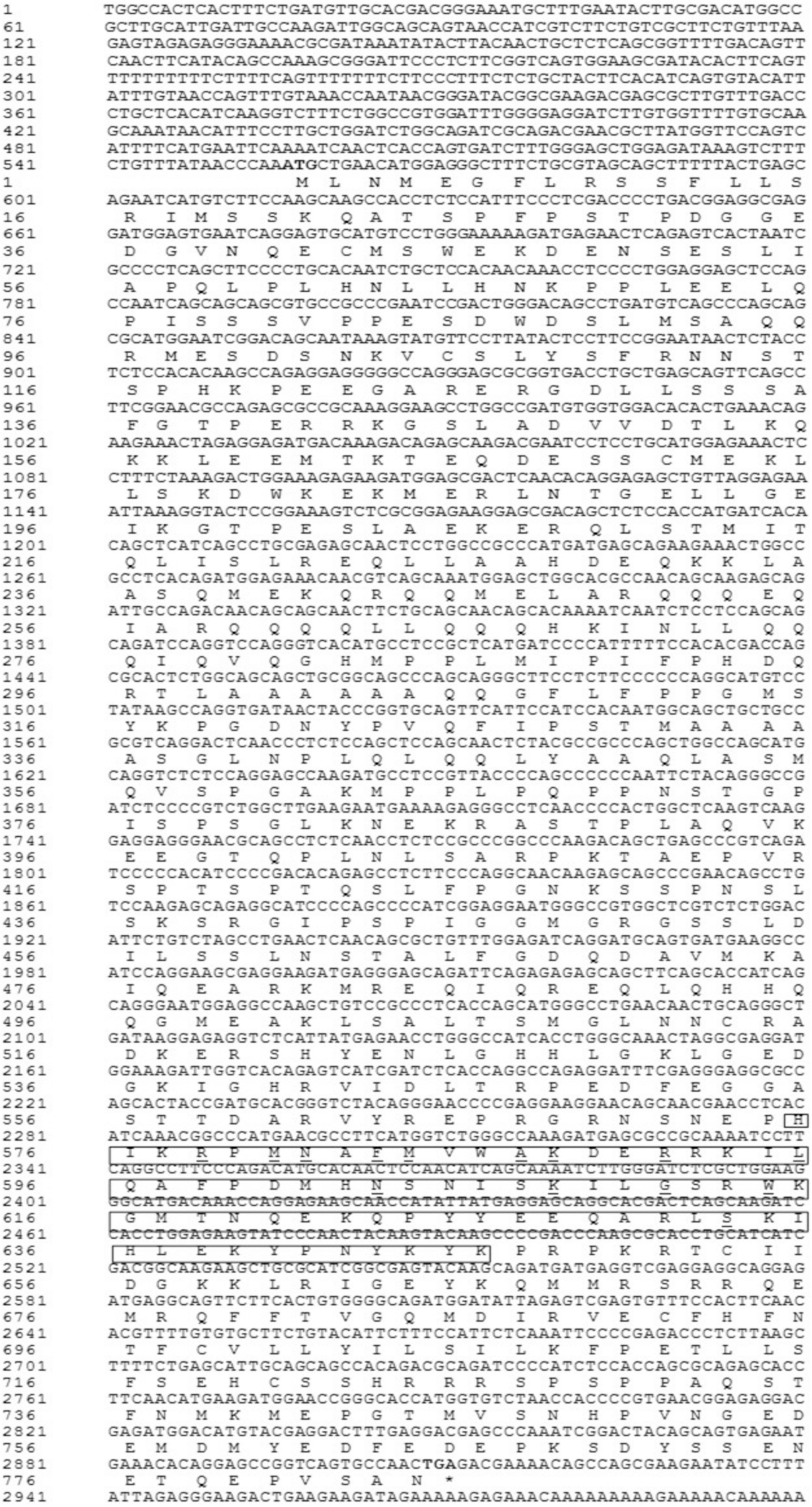
* Corresponding author: 13837331530@163.com

0030-9923/2018/0002-0417 $ 9.00/0
Copyright 2018 Zoological Society of Pakistan



```
1       GTGAGGTTCGGTGTGTGAGAGTGTGTGTGCACGGAGAGAAACAGAAACAGTCTTTGGTTT
61      TGCGAGAGGTTAGAGGAGCACCAGGCAGCGTGTGCGTGTATGTGTGTGTGTGTTCTCGGA
121     GATCCCTGGGTATCGGACGCGAATGTTGTCTTGGAAGGACTGCTGCATCTCAGACCCCTT
181     GTTTCAGCATGCTTACTGAGCCTGAGCTTCCTCAGGAGTTTAACAGGATGTCTTCCAAGC
1                M  L  T  E  P  E  L  P  Q  E  F  N  R  M  S  S  K  R
241     GACCAGCCTCTCCATATGGGGGAACAGATGGAGAGGTAGTCATGGCGACGAGCAGACAGC
19        P  A  S  P  Y  G  G  T  D  G  E  V  V  M  A  T  S  R  Q  R
301     GATTGGAGGATGAAGAGGTTGACGGACATGCTGTCATTCACTTGCCCCTGAGTTCGTACT
39        L  E  D  E  E  V  D  G  H  A  V  I  H  L  P  L  S  S  Y  C
361     GCAGCAAAGTGTCCCCGCGATCTCCACGACTGCTCGACAGCCCCCCAACACTACACGCAA
59        S  K  V  S  P  R  S  P  R  L  L  D  S  P  P  T  L  H  A  N
421     ACATGGATCAGGAAGGCAGTAAAGGCCTGGCCCTAAGCCCCTACCCCCAGCACAACTCGT
79        M  D  Q  E  G  S  K  G  L  A  L  S  P  Y  P  Q  H  N  S  S
481     CTACCTCACCCAGTAAACAGCAGCAGCAGGAGGAGGGGTCCATGCGGCCAGGGAGCGAGC
99        T  S  P  S  K  Q  Q  Q  Q  E  E  G  S  M  R  P  G  S  E  P
541     CCAGCTCCGCTTCAGCTCTGGGCACCCCGGAGAGACGCAAAGGCAGCCTGGCCGACGTGG
119       S  S  A  S  A  L  G  T  P  E  R  R  K  G  S  L  A  D  V  V
601     TCGACACGCTCAAACAGCGCAAGATGGAGGAACTCATCAAGAACGAGCCAGAGGAGGCCC
139       D  T  L  K  Q  R  K  M  E  E  L  I  K  N  E  P  E  E  A  P
661     CCAGTATTGAAAGGCTGTTGTCTAAGGACTGGAAGGACAAGCTATTGGCCATGGGTTCTG
159       S  I  E  R  L  L  S  K  D  W  K  D  K  L  L  A  M  G  S  G
721     GCAACTTTGGAGAGATCAAAGTGAAGCAAGGCACCCAGGACAGCCTGGTGGAGAAAGAGC
179       N  F  G  E  I  K  V  K  Q  G  T  Q  D  S  L  V  E  K  E  R
781     GCCAGCTGATGGGGATGATCACGCAACTCAGCAGTCTCAGAGAGCAGCTACTGGCCGCCC
199       Q  L  M  G  M  I  T  Q  L  S  S  L  R  E  Q  L  L  A  A  H
841     ATGAGGAGCAGAAAAAACTGGCCGCTTCGCAGATCGAAAAACAGCGCCAACAAATGGAGT
219       E  E  Q  K  K  L  A  A  S  Q  I  E  K  Q  R  Q  Q  M  E  L
901     TAGCGAAACAGCAACAGGACCAGATTGCACGACAACAGCAGCAACTTCTGCAGCAACAAC
239       A  K  Q  Q  Q  D  Q  I  A  R  Q  Q  Q  Q  L  L  Q  Q  Q  H
961     ACAAAATCAACCTCCTGCAGCAACAGATCCAGCAGGTCCAGGGTCAGTTGCCTCCACTGA
259       K  I  N  L  L  Q  Q  Q  I  Q  Q  V  Q  G  Q  L  P  P  L  M
1021    TGATACCTGTGTTCCCTCCAGATCAGAGAACTCTGGCGGCCGCTGCAGCACAACAAGGAT
279       I  P  V  F  P  P  D  Q  R  T  L  A  A  A  A  A  Q  Q  G  F
1081    TTCTACTTCCTCCAGGGTTCAACTACAAACCTGGCTGCAGTGACCCTTACCCTTTACAGC
299       L  L  P  P  G  F  N  Y  K  P  G  C  S  D  P  Y  P  L  Q  L
1141    TAATCCCTACGACTATGGCAGCGGCTGCTGCAGCGACGCCTGGCCTCGGCCCTCTTCAAC
319       I  P  T  T  M  A  A  A  A  A  A  T  P  G  L  G  P  L  Q  L
1201    TGCAGTTTGGCCCACATATGTGCCCTCTGGATAACAGCCCAATAAGCTCACACTTTCAAT
339       Q  F  G  P  H  M  C  P  L  D  N  S  P  I  S  S  H  F  Q  C
1261    GCGGAGCCACACACAGCAGAACGAACAGGGAAACTCAAACCCCCAATACCGATCTGGTGG
359       G  A  T  H  S  R  T  N  R  E  T  Q  T  P  N  T  D  L  V  A
1321    CATTGTCCGACCTCGATGAGCAAAGAGAGAGAGAGTGAGAGAAGCAAGCACCTCTCATGATG
379       L  S  D  L  D  E  Q  R  E  R  V  R  E  A  S  T  S  H  D  V
1381    TTGCCTTGATCAAACGTGACGAGGGCGCTCAGCCCTTGAATCTGTCGTCCAAGCCTAAGG
399       A  L  I  K  R  D  E  G  A  Q  P  L  N  L  S  S  K  P  K  A
1441    CATCCGAGAGCAAGTCACCCACCTCCCCCGCTTCCCCACAGGTCCCTGCCCTGAAGCTGG
419       S  E  S  K  S  P  T  S  P  A  S  P  Q  V  P  A  L  K  L  G
1501    GCCCCGGGTCCCTGAAGCACAGCGCCCCCTCCAGCATCGGAGGACCGCCGTCCAGACTCA
439       P  G  S  L  K  H  S  A  P  S  S  I  G  G  P  P  S  R  L  S
1561    GCTCCATAGACCTGCTATCATCCATCACCTCCGGCGGCTACCTGAACGACCATGAGGCAG
459       S  I  D  L  L  S  S  I  T  S  G  G  Y  L  N  D  H  E  A  V
1621    TGACCAAAGCTTTCCAGGAGGCCCGACAGATGAAGGAGCAGCTAAAGAGAGAACAGCAGG
479       T  K  A  F  Q  E  A  R  Q  M  K  E  Q  L  K  R  E  Q  Q  V
1681    TTCTCGACGCCAAAGTAGCAGCAGTCAACAGCCTGAGCCTCAACAATGGTCGCTCAGAGA
499       L  D  A  K  V  A  A  V  N  S  L  S  L  N  N  G  R  S  E  K
1741    AGGATAAAGCTGCTCTTGAAAGCCTGAGCCAACAGTTAAAACAGTCTGAAGAGAGCAAGT
519       D  K  A  A  L  E  S  L  S  Q  Q  L  K  Q  S  E  E  S  K  F
1801    TCACTCATGCAATGATGGACTTCGGCATCAGTGGTGATTCAGATGGAAGCCCGAGTGTGT
539       T  H  A  M  M  D  F  G  I  S  G  D  S  D  G  S  P  S  V  S
1861    CAGACTCACGGATCTTCAGGGAGGCTCGAGGACGTGGCAGCAGCGAACCGCACATCAAAC
559       D  S  R  I  F  R  E  A  R  G  R  G  S  S  E  P  H  I  K  R
1921    GGCCCATGAACGCCTTTATGGTCTGGGCTAAGGACGAGAGGAGGAAGATCCTTCAGGCTT
579       P  M  N  A  F  M  V  W  A  K  D  E  R  R  K  I  L  Q  A  F
1981    TCCCTGACATGCATAACTCAAACATCAGCAAGATCCTCGGTTCTCGTTGGAAGTCCATGA
599       P  D  M  H  N  S  N  I  S  K  I  L  G  S  R  W  K  S  M  T
2041    CCAACCTGGAAAAGCAGCCATATTATGAAGAACAGGCCCGTCTTAGTAAGCAGCACCTAG
619       N  L  E  K  Q  P  Y  Y  E  E  Q  A  R  L  S  K  Q  H  L  E
2101    AGAAATACCCAGACTACAAATACAAACCACGGCCCAAACGCACCTGCCTAGTGGACGGAA
639       K  Y  P  D  Y  K  Y  K  P  R  P  K  R  T  C  L  V  D  G  K
2161    AGAAATTGCGCATTGGTGAATATAAGGCCATCATGCGCAACCGCAGACAGGAAATGCGCC
659       K  L  R  I  G  E  Y  K  A  I  M  R  N  R  R  Q  E  M  R  Q
2221    AGTACTTCACTGTAGGGCAGCAGCCCCAGTTGCCCTTGTCCTCAGCGGGTGTTGTGTACC
679       Y  F  T  V  G  Q  Q  P  Q  L  P  L  S  S  A  G  V  V  Y  P
2281    CAGGCGCACTCTCTATGGCAGGAATGCCCTCTCCACAGATGCCCTCTGAGCACTCGAGCA
699       G  A  L  S  M  A  G  M  P  S  P  Q  M  P  S  E  H  S  S  M
2341    TGTCTAGTAGCCCTGAACCAGTCCCACCCAGCAGCCAGCCCTCTTACCTCACACACAAGG
719       S  S  S  P  E  P  V  P  P  S  S  Q  P  S  Y  L  T  H  K  G
2401    GGGAGGGGCCTCGCGTTAAGGAGGAGGAGCTCCGAATGGACGACGGCAATGGCGACGCCT
739       E  G  P  R  V  K  E  E  E  L  R  M  D  D  G  N  G  D  A  Y
2461    ACGATGACTTCGACTACGAAGACGATGATGGGGACTATGGCAGTGACAGCGAGAATCACA
759       D  D  F  D  Y  E  D  D  D  G  D  Y  G  S  D  S  E  N  H  I
2521    TCACTCAGTAAAGCTGCCAATCACAGCCATCTCACCAGTCACCTGGCATCTCACGCAAGC
779       T  Q  *
2581    AAACTTCATTCCAAACCCCATACACCGTCCAATCAGGACTCCTCCAGAGCCAAGACCGAT
2641    ATCCACAGCAAAGCACAGGACCTGTCAATCAAAGTGACTGTTACATGGAAAGGAGGGTCT
2701    TGTAAAGGAAGTATAACACTTGCCTTGACTCCTGCTGAGCGAACGTACCGCGGCAAGTTC
2761    ACAATGAATAACGAGTAAGCTACGATGACATAGGAAAAAAAAAAAAAA
```

Supplementary Fig. 1A. Full-length nucleotide cDNA sequence of the *CcSox5* gene.



```
1         TGGCCACTCACTTTCTGATGTTGCACGACGGGAAATGCTTTGAATACTTGCGACATGGCC
61        GCTTGCATTGATTGCCAAGATTGGCAGCAGTAACCATCGTCTTCTGTCGCTTCTGTTTAA
121       GAGTAGAGAGGGAAAACGCGATAAATATACTTACAACTGCTCTCAGCGGTTTTGACAGTT
181       CAACTTCATACAGCCAAAGCGGGATTCCCTCTTCGGTCAGTGGAAGCGATACACTTCAGT
241       TTTTTTTTTCTTTTCAGTTTTTTTCTTCCCTTTCTCTGCTACTTCACATCAGTGTACATT
301       ATTTGTAACCAGTTTGTAAACCAATAACGGGATACGGCGAAGACGAGCGCTTGTTTGACC
361       CTGCTCACATCAAGGTCTTTCTGGCCGTGGATTTGGGGAGGATCTTGTGGTTTTGTGCAA
421       GCAAATAACATTTCCTTGCTGGATCTGGCAGATCGCAGACGAACGCTTATGGTTCCAGTC
481       ATTTTCATGAATTCAAAATCAACTCACCAGTGATCTTTGGGAGCTGGAGATAAAGTCTTT
541       CTGTTTATAACCCAAATGCTGAACATGGAGGGCTTTCTGCGTAGCAGCTTTTTACTGAGC
1                         M  L  N  M  E  G  F  L  R  S  S  F  L  L  S
601       AGAATCATGTCTTCCAAGCAAGCCACCTCTCCATTTCCCTCGACCCCTGACGGAGGCGAG
16         R  I  M  S  S  K  Q  A  T  S  P  F  P  S  T  P  D  G  G  E
661       GATGGAGTGAATCAGGAGTGCATGTCCTGGGAAAAAGATGAGAACTCAGAGTCACTAATC
36         D  G  V  N  Q  E  C  M  S  W  E  K  D  E  N  S  E  S  L  I
721       GCCCCTCAGCTTCCCCTGCACAATCTGCTCCACAACAAACCTCCCCTGGAGGAGCTCCAG
56         A  P  Q  L  P  L  H  N  L  L  H  N  K  P  P  L  E  E  L  Q
781       CCAATCAGCAGCAGCGTGCCGCCCGAATCCGACTGGGACAGCCTGATGTCAGCCCAGCAG
76         P  I  S  S  S  V  P  P  E  S  D  W  D  S  L  M  S  A  Q  Q
841       CGCATGGAATCGGACAGCAATAAAGTATGTTCCTTATACTCCTTCCGGAATAACTCTACC
96         R  M  E  S  D  S  N  K  V  C  S  L  Y  S  F  R  N  N  S  T
901       TCTCCACACAAGCCAGAGGAGGGGGCCAGGGAGCGCGGTGACCTGCTGAGCAGTTCAGCC
116        S  P  H  K  P  E  E  G  A  R  E  R  G  D  L  L  S  S  S  A
961       TTCGGAACGCCAGAGCGCCGCAAAGGAAGCCTGGCCGATGTGGTGGACACACTGAAACAG
136        F  G  T  P  E  R  R  K  G  S  L  A  D  V  V  D  T  L  K  Q
1021      AAGAAACTAGAGGAGATGACAAAGACAGAGCAAGACGAATCCTCCTGCATGGAGAAACTC
156        K  K  L  E  E  M  T  K  T  E  Q  D  E  S  S  C  M  E  K  L
1081      CTTTCTAAAGACTGGAAAGAGAAGATGGAGCGACTCAACACAGGAGAGCTGTTAGGAGAA
176        L  S  K  D  W  K  E  K  M  E  R  L  N  T  G  E  L  L  G  E
1141      ATTAAAGGTACTCCGGAAAGTCTCGCGGAGAAGGAGCGACAGCTCTCCACCATGATCACA
196        I  K  G  T  P  E  S  L  A  E  K  E  R  Q  L  S  T  M  I  T
1201      CAGCTCATCAGCCTGCGAGAGCAACTCCTGGCCGCCCATGATGAGCAGAAGAAACTGGCC
216        Q  L  I  S  L  R  E  Q  L  L  A  A  H  D  E  Q  K  K  L  A
1261      GCCTCACAGATGGAGAAACAACGTCAGCAAATGGAGCTGGCACGCCAACAGCAAGAGCAG
236        A  S  Q  M  E  K  Q  R  Q  Q  M  E  L  A  R  Q  Q  Q  E  Q
1321      ATTGCCAGACAACAGCAGCAACTTCTGCAGCAACAGCACAAAATCAATCTCCTCCAGCAG
256        I  A  R  Q  Q  Q  Q  L  L  Q  Q  Q  H  K  I  N  L  L  Q  Q
1381      CAGATCCAGGTCCAGGGTCACATGCCTCCGCTCATGATCCCCATTTTTCCACACGACCAG
276        Q  I  Q  V  Q  G  H  M  P  P  L  M  I  P  I  F  P  H  D  Q
1441      CGCACTCTGGCAGCAGCTGCGGCAGCCCAGCAGGGCTTCCTCTTCCCCCCAGGCATGTCC
296        R  T  L  A  A  A  A  A  A  Q  Q  G  F  L  F  P  P  G  M  S
1501      TATAAGCCAGGTGATAACTACCCGGTGCAGTTCATTCCATCCACAATGGCAGCTGCTGCC
316        Y  K  P  G  D  N  Y  P  V  Q  F  I  P  S  T  M  A  A  A  A
1561      GCGTCAGGACTCAACCCTCTCCAGCTCCAGCAACTCTACGCCGCCCAGCTGGCCAGCATG
336        A  S  G  L  N  P  L  Q  L  Q  Q  L  Y  A  A  Q  L  A  S  M
1621      CAGGTCTCTCCAGGAGCCAAGATGCCTCCGTTACCCCAGCCCCCCAATTCTACAGGGCCG
356        Q  V  S  P  G  A  K  M  P  P  L  P  Q  P  P  N  S  T  G  P
1681      ATCTCCCCGTCTGGCTTGAAGAATGAAAAGAGGGCCTCAACCCCACTGGCTCAAGTCAAG
376        I  S  P  S  G  L  K  N  E  K  R  A  S  T  P  L  A  Q  V  K
1741      GAGGAGGGAACGCAGCCTCTCAACCTCTCCGCCCGGCCCAAGACAGCTGAGCCCGTCAGA
396        E  E  G  T  Q  P  L  N  L  S  A  R  P  K  T  A  E  P  V  R
1801      TCCCCCACATCCCCGACACAGAGCCTCTTCCCAGGCAACAAGAGCAGCCCGAACAGCCTG
416        S  P  T  S  P  T  Q  S  L  F  P  G  N  K  S  S  P  N  S  L
1861      TCCAAGAGCAGAGGCATCCCCAGCCCCATCGGAGGAATGGGCCGTGGCTCGTCTCTGGAC
436        S  K  S  R  G  I  P  S  P  I  G  G  M  G  R  G  S  S  L  D
1921      ATTCTGTCTAGCCTGAACTCAACAGCGCTGTTTGGAGATCAGGATGCAGTGATGAAGGCC
456        I  L  S  S  L  N  S  T  A  L  F  G  D  Q  D  A  V  M  K  A
1981      ATCCAGGAAGCGAGGAAGATGAGGGAGCAGATTCAGAGAGAGCAGCTTCAGCACCATCAG
476        I  Q  E  A  R  K  M  R  E  Q  I  Q  R  E  Q  L  Q  H  H  Q
2041      CAGGGAATGGAGGCCAAGCTGTCCGCCCTCACCAGCATGGGCCTGAACAACTGCAGGGCT
496        Q  G  M  E  A  K  L  S  A  L  T  S  M  G  L  N  N  C  R  A
2101      GATAAGGAGAGGTCTCATTATGAGAACCTGGGCCATCACCTGGGCAAACTAGGCGAGGAT
516        D  K  E  R  S  H  Y  E  N  L  G  H  H  L  G  K  L  G  E  D
2161      GGAAAGATTGGTCACAGAGTCATCGATCTCACCAGGCCAGAGGATTTCGAGGGAGGCGCC
536        G  K  I  G  H  R  V  I  D  L  T  R  P  E  D  F  E  G  G  A
2221      AGCACTACCGATGCACGGGTCTACAGGGAACCCCGAGGAAGGAACAGCAACGAACCTCAC
556        S  T  T  D  A  R  V  Y  R  E  P  R  G  R  N  S  N  E  P  H
2281      ATCAAACGGCCCATGAACGCCTTCATGGTCTGGGCCAAAGATGAGCGCCGCAAAATCCTT
576        I  K  R  P  M  N  A  F  M  V  W  A  K  D  E  R  R  K  I  L
2341      CAGGCCTTCCCAGACATGCACAACTCCAACATCAGCAAAATCTTGGGATCTCGCTGGAAG
596        Q  A  F  P  D  M  H  N  S  N  I  S  K  I  L  G  S  R  W  K
2401      GGCATGACAAACCAGGAGAAGCAACCATATTATGAGGAGCAGGCACGACTCAGCAAGATC
616        G  M  T  N  Q  E  K  Q  P  Y  Y  E  E  Q  A  R  L  S  K  I
2461      CACCTGGAGAAGTATCCCAACTACAAGTACAAGCCCCGACCCAAGCGCACCTGCATCATC
636        H  L  E  K  Y  P  N  Y  K  Y  K  P  R  P  K  R  T  C  I  I
2521      GACGGCAAGAAGCTGCGCATCGGCGAGTACAAGCAGATGATGAGGTCGAGGAGGCAGGAG
656        D  G  K  K  L  R  I  G  E  Y  K  Q  M  M  R  S  R  R  Q  E
2581      ATGAGGCAGTTCTTCACTGTGGGGCAGATGGATATTAGAGTCGAGTGTTTCCACTTCAAC
676        M  R  Q  F  F  T  V  G  Q  M  D  I  R  V  E  C  F  H  F  N
2641      ACGTTTTGTGTGCTTCTGTACATTCTTTCCATTCTCAAATTCCCCGAGACCCTCTTAAGC
696        T  F  C  V  L  L  Y  I  L  S  I  L  K  F  P  E  T  L  L  S
2701      TTTTCTGAGCATTGCAGCAGCCACAGACGCAGATCCCCATCTCCACCAGCGCAGAGCACC
716        F  S  E  H  C  S  S  H  R  R  R  S  P  S  P  P  A  Q  S  T
2761      TTCAACATGAAGATGGAACCGGGCACCATGGTGTCTAACCACCCCGTGAACGGAGAGGAC
736        F  N  M  K  M  E  P  G  T  M  V  S  N  H  P  V  N  G  E  D
2821      GAGATGGACATGTACGAGGACTTTGAGGACGAGCCCAAATCGGACTACAGCAGTGAGAAT
756        E  M  D  M  Y  E  D  F  E  D  E  P  K  S  D  Y  S  S  E  N
2881      GAAACACAGGAGCCGGTCAGTGCCAACTGAGACGAAAACAGCCAGCGAAGAATATCCTTT
776        E  T  Q  E  P  V  S  A  N  *
2941      ATTAGAGGGAAGACTGAAGAAGATAGAAAAAGAGAAACAAAAAAAAAGAAAAACAAAAAA
```

Supplementary Fig. 1B. Full-length nucleotide cDNA sequence of the *CcSox6* gene. The predicted amino acid sequences are shown underneath the CDS (coding sequence). The start and stop codons are bold-typed. The HMG (high mobility group) box domain is boxed. The DNA binding site is underlined.

```
1      GGAAACTGAACACCAGAAAGTCACTCTGGAGTAGTGAAGGGAGGCTGGATCGGCGAGAGT
61     GCGACGCATTCATCAGCTGTGCGCTCGGAGTCTCGGGTTGTGCTGTAAGTCGTTGATAGA
121    CAGAGGCGTATTATTCAATGACACGAGAGGAAACCACATTTTGCATGCGGCGTTATCGAT
181    TCTGCGGCCCAGAGACGACGGAAAGACTGAAGGAATAAAGTTTAGGGAGCGCAAAGCGCA
241    GCGCACAGTTTTTTTGGAGTTGAATGTGTGAGCCCAGTCCACCTGCTCTTCATCCCCAAG
1                             M  C  E  P  S  P  P  A  L  H  P  Q
301    CCAAAGGCAATGTGACCATGGTGGGTGTTGGTATCAAGCTTGAAGACGACGAATCAGGAG
13     A  K  G  N  V  T  M  V  G  V  G  I  K  L  E  D  D  E  S  G
361    GTAGAGTCATCAACCGAGAATCCCAGCACTCCTCTCCAGTCACTGATTGGTCAGTTCGGC
33     G  R  V  I  N  R  E  S  Q  H  S  S  P  V  T  D  W  S  V  R
421    AGCTCTCAGACAGAGACGACTCTGAGAAGTCAGCGCCCCCTAATGCACAATCTCCTTTGG
53     Q  L  S  D  R  D  D  S  E  K  S  A  P  P  N  A  Q  S  P  L
481    CGAGCTCTCCCAGTAAAGACCCAGTACGATCATACACACCTGCCAGTCAAGATGGAAGCA
73     A  S  S  P  S  K  D  P  V  R  S  Y  T  P  A  S  Q  D  G  S
541    GTAGAGCTCATGATAAGACTGACAGCAAATCTAATGTCAGTGAGGTGATGCCCACTATTG
93     S  R  A  H  D  K  T  D  S  K  S  N  V  S  E  V  M  P  T  I
601    AAAAGCTTCTGAATACGGACTGGAAAGAAAAGTTCCTTGGGAAGGGTATTCAGGGCAGTG
113    E  K  L  L  N  T  D  W  K  E  K  F  L  G  K  G  I  Q  G  S
661    TCAACCTAAAAGGAACCCCAGAGAGTCTTGCAGAGAAGGAGCTTCAGTTGTTGTTGATGA
133    V  N  L  K  G  T  P  E  S  L  A  E  K  E  L  Q  L  L  L  M
721    TAAACCAGCTCTCTGGTCTGCGAGAGCAGCTGCTAGGAGCTCATTCTGAGCAGAGGAATA
153    I  N  Q  L  S  G  L  R  E  Q  L  L  G  A  H  S  E  Q  R  N
781    TGGCCGCTCTGCTGCTGGAGAAACAGCAGCAACAGATGGAACTGGCAAGGCAGCAGCAAG
173    M  A  A  L  L  L  E  K  Q  Q  Q  Q  M  E  L  A  R  Q  Q  Q
841    AACAGATCGCCAAACAGCAGCAGCAGTTGATACAGCAGCAACATAAAATCAACTTACTGC
193    E  Q  I  A  K  Q  Q  Q  Q  L  I  Q  Q  Q  H  K  I  N  L  L
901    AGCAGCAAATTCAGCAGGTAAACATGCCGTATGTTATGATCCCTGCCTTTCATCCCAACG
213    Q  Q  Q  I  Q  Q  V  N  M  P  Y  V  M  I  P  A  F  H  P  N
961    CCCAACCGTTGCCCGTCACCTCTGAGGCACAGATGGGATTGCCTCGGCATCCCATACACT
233    A  Q  P  L  P  V  T  S  E  A  Q  M  G  L  P  R  H  P  I  H
1021   GCAAACCAGTGGAATATCCCTTGCCGTTGCTGCCTAATCCACACTCCACCCCAGTCAAAA
253    C  K  P  V  E  Y  P  L  P  L  L  P  N  P  H  S  T  P  V  K
1081   GGTCCAGTGGTACAGTCTTTCGCCAGGATTCCAGTCAACCTCTCAACCTCACGGCTAAAC
273    R  S  S  G  T  V  F  R  Q  D  S  S  Q  P  L  N  L  T  A  K
1141   CCAAGACCCCAAGTCCCCAGGCACTGGAAATGGCCCACTTACAAGCAGGGTACCGAGCCA
293    P  K  T  P  S  P  Q  A  L  E  M  A  H  L  Q  A  G  Y  R  A
1201   GAGATCTGACCCATAGCCCTACTCGATCTGCACTCAGCCTGAGCTTCTTGGGAGAGGGTG
313    R  D  L  T  H  S  P  T  R  S  A  L  S  L  S  F  L  G  E  G
1261   ATGTTGTGACACAAGCCATTCATGATGCCCAGCAGCTGTTGCGTGGAGGACAAGGCCCGA
333    D  V  V  T  Q  A  I  H  D  A  Q  Q  L  L  R  G  G  Q  G  P
1321   CAGGCCGAGAGAGGGACCGTGAGAGGGACAACAACACAAGGCTGGAGGCATCTCGAGAGC
353    T  G  R  E  R  D  R  E  R  D  N  N  T  R  L  E  A  S  R  E
1381   GGATGGATGATGGACAGTCCCTGCGGATGAATGAGGACCGCCTCAGCTGTGACACTGAGG
373    R  M  D  D  G  Q  S  L  R  M  N  E  D  R  L  S  C  D  T  E
1441   GGCAGATGGCAGCCTCTGGTGTGGGAAGCTTTGGTGAATCCCGGACTCCCTCCTCTGGCC
393    G  Q  M  A  A  S  G  V  G  S  F  G  E  S  R  T  P  S  S  G
1501   ACATCAAACGTCCCATGAACGCCTTCATGGTGTGGGCCAAAGACGAAAGACGGCGTATCC
413    H  I  K  R  P  M  N  A  F  M  V  W  A  K  D  E  R  R  R  I
1561   TGCAGGCGTTCCCTGACATGCACAACTCCAGTATCAGCAAGATCCTCGGTTCCCGATGGA
433    L  Q  A  F  P  D  M  H  N  S  S  I  S  K  I  L  G  S  R  W
1621   AGACGATGTCCAATCAGGAGAAGCAGCCGTACTATGAAGAGCAGGCCAGATTGAGTCGGC
453    K  T  M  S  N  Q  E  K  Q  P  Y  Y  E  E  Q  A  R  L  S  R
1681   AGCACCTGGAACGTTATCCAGACTACAAATACAAACCCCGGCCAAAGCGGACCTGTATCG
473    Q  H  L  E  R  Y  P  D  Y  K  Y  K  P  R  P  K  R  T  C  I
1741   TAGAGGGCAGACGACTGCGGGTCGGAGAGTACAAGGCCATGATGAAGAGTCGACGACAGG
493    V  E  G  R  R  L  R  V  G  E  Y  K  A  M  M  K  S  R  R  Q
1801   AGCAACGAGTAACTTACCCTCCAAGCCAATCAGAGCAGCAGCTACCGTACCCTCCCAGTG
513    E  Q  R  V  T  Y  P  P  S  Q  S  E  Q  Q  L  P  Y  P  P  S
1861   AAGGACAGTACTCGACCACGCCGGTGTCACTGGCCCCATTGCCCTTGCACCCTGCCCTCC
533    E  G  Q  Y  S  T  T  P  V  S  L  A  P  L  P  L  H  P  A  L
1921   TGGAGCACTACCTTCCCCGGGGTCTGGAGCCCATGAAAGGCCAAGGGAGTGAGATGCGGG
553    L  E  H  Y  L  P  R  G  L  E  P  M  K  G  Q  G  S  E  M  R
1981   AGGCCACGCGGCCCAGACAGCCCTACAGTGAAGGGGAGGAGAGTGATGTGGGTGAACGCA
573    E  A  T  R  P  R  Q  P  Y  S  E  G  E  E  S  D  V  G  E  R
2041   GTGAGGGAGAGCTGGTTGTCCTCACTGACTGAGACAGAAGGACGAAAGGATTGGAGGACA
593    S  E  G  E  L  V  V  L  T  D  *
2101   GAAAGGGCGGAAAACGTGAAAAAGGGGGAGTGAGATGGAGAAAATAACTGGGAGATTTTG
2161   CAGTCTCTGCGTACAACATGATGGTGTCCAATAAATGGTATAGGTCAGTGTTTTCCAGTC
2221   CAGTTGTGAGGCTGTTCTGGTTGCAAACTGGGTTGGTTAGGAAGGCAGCAGACATAAATG
2281   AAGCATAGATCTTCATTTTTCTGATTTTTCCAGTGGTGTAAGCCTAGACAGGATAGGTCT
2341   TACATACTGACCACAGCTTGGAAAAAAAAAAAAAA
```

Supplementary Fig. 1C. Full-length nucleotide cDNA sequence of the *CcSox13* gene. The predicted amino acid sequences are shown underneath the CDS (coding sequence). The start and stop codons are bold-typed. The HMG (high mobility group) box domain is boxed. The DNA binding site is underlined.