Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle

Chen, Songlin; Zhang, Guojie; Shao, Changwei; Huang, Quanfei; Liu, Geng; Zhang, Pei; Song, Wentao; An, Na; Chalopin, Domitille; Volff, Jean-Nicolas; Hong, Yunhan; Li, Qiye; Sha, Zhenxia; Zhou, Heling; Xie, Mingshu; Yu, Qiulin; Liu, Yang; Xiang, Hui; Wang, Na; Wu, Kui; Yang, Changgeng; Zhou, Qian; Liao, Xiaolin; Yang, Linfeng; Hu, Qiaomu; Zhang, Jilin; Meng, Liang; Jin, Lijun; Tian, Yongsheng; Lian, Jinmin; Yang, Jingfeng; Miao, Guidong; Liu, Shanshan; Liang, Zhou; Yan, Fang; Li, Yangzhen; Sun, Bin; Zhang, Hong; Zhang, Jing; Zhu, Ying; Du, Min; Zhao, Yongwei; Schartl, Manfred; Tang, Qisheng; Wang, Jun

Published in:
Nature Genetics

DOI:
10.1038/ng.2890

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle

Songlin Chen1,10,11, Guojie Zhang2,10, Changwei Shao1,10, Quanfei Huang2,10, Geng Liu2,10, Pei Zhang2,10, Wentao Song1, Na An2, Domitille Chalopin3, Jean-Nicolas Volff3, Yunhan Hong4, Qiye Li2, Zhenxia Sha1, Heling Zhou2, Mingshu Xie1, Qiulin Yu2, Yang Liu5, Hui Xiang6, Na Wang1, Kui Wu2, Changgeng Yang1, Qian Zhou2, Xiaolin Liao1, Linfeng Yang2, Qiaomu Hu1, Jilin Zhang2, Liang Meng1, Lijun Jin2, Yongsheng Tian1, Jinmin Lian2, Jingfeng Yang1, Guidong Miao1, Shanshan Liu1, Zhuo Liang1, Fang Yan1, Yangzhen Li1, Bin Sun1, Hong Zhang1, Jing Zhang1, Ying Zhu1, Min Du1, Yongwei Zhao1, Manfred Schartl7,11, Qisheng Tang1,11 & Jun Wang2,8,9,11

Genetic sex determination by W and Z chromosomes has developed independently in different groups of organisms. To better understand the evolution of sex chromosomes and the plasticity of sex-determination mechanisms, we sequenced the whole genomes of a male (ZZ) and a female (ZW) half-smooth tongue sole (Cynoglossus semilaevis). In addition to insights into adaptation to a benthic lifestyle, we find that the sex chromosomes of these fish are derived from the same ancestral vertebrate protochromosome as the avian W and Z chromosomes. Notably, the same gene on the Z chromosome, dmrt1, which is the male-determining gene in birds, showed convergent evolution of features that are compatible with a similar function in tongue sole. Comparison of the relatively young tongue sole sex chromosomes with those of mammals and birds identified events that occurred during the early phase of sex-chromosome evolution. Pertinent to the current debate about heterogametic sex-chromosome decay, we find that massive gene loss occurred in the wake of sex-chromosome ‘birth’.

The existence of males and females, which are often strikingly different in morphology, reproductive strategies and behavior, is one of the most widespread phenomena in biology. However, the genetic mechanisms that generate this ubiquitous pattern are surprisingly diverse and do not follow a phylogenetic pattern. Sex-determination mechanisms can differ between even closely related species and arise frequently and independently. Fish provide a paradigmatic example, as their sex-determination mechanisms range from environmental to different modes of genetic determination. The evolutionary meaning of this remarkable plasticity is unknown.

For genetic sex determination, where the trigger for female or male development comes from the genetic constitution of the individual, the evolution of sex-determination mechanisms is connected to a very peculiar phenomenon, namely the formation of sex chromosomes1–4. To improve understanding of the function and evolution of sex chromosomes, their genetic organization must be deciphered. However, owing to their degenerate nature and high repetitive DNA content, sex chromosomes pose almost insurmountable problems in deciphering their gene content and organization. So far, only the human5, chimpanzee6 and rhesus macaque Y chromosomes7 and the male-specific region on the Y chromosome of one fish, the medaka8, have been sequenced. These analyses have nevertheless provided important insights into the evolution of Y chromosomes, their genomic organization and their degeneration processes, as well as predictions as to their likely evolutionary fate9–12.

Much less genomic information exists on W chromosomes because, as with Y chromosomes, they are predominantly highly repetitive in nature. The prevailing theory of the evolution of sex chromosomes predicts that degeneration of the heterogametic sex chromosome is a stepwise process that occurs over an extended period of time. We therefore reasoned that an evolutionarily young W chromosome...
## RESULTS

### Genome assembly and annotation

We sequenced DNA from one male (ZZ) and one female (ZW) C. semilaevis separately with the Illumina sequencing platform using a whole-genome shotgun approach with 212-fold coverage (Table 1, Supplementary Fig. 1 and Supplementary Tables 1 and 2). We first assembled the male and female genomes separately (Supplementary Tables 2 and 3 and Supplementary Note). We considered scaffolds with double coverage in males as putative Z scaffolds (a total of 23.3 Mb) (Supplementary Note). 94% of these scaffolds could be anchored in one linkage group of the male genetic map by microsatellite markers (Supplementary Table 4 and Supplementary Note). We further confirmed Z-chromosome assignment by quantitative PCR in ZZ, ZW and ZW individuals (Supplementary Table 5 and Supplementary Note).

We retrieved W sequences by subtracting the male assembly from the female assembly, yielding 16.4 Mb of scaffolds (Table 1 and Supplementary Table 4), which is consistent with the different sizes of the female and male haploid genomes as estimated by k-mer calculation (Supplementary Figs. 2 and 3). Owing to the lack of marker resolution, we ordered the W scaffolds in a pseudo-W chromosome on the basis of their synteny with the Z chromosome. We derived autosome sequences from the female whole-genome sequence after subtraction of the putative W and Z chromosomes. As a result, the final genome assembly was 477 Mb with a scaffold N50 size of 867 kb (Table 1 and Supplementary Table 3). We next constructed a high-resolution genetic map with 12,142 SNPs by restriction site–associated DNA sequencing (RAD-seq), which we assigned to 22 linkage groups corresponding to the 20 autosomes and both sex chromosomes (Supplementary Table 4 and Supplementary Note). Together with the former genetic map (942 simple sequence repeat (SSR) markers), we consistently anchored 445 Mb (93.3%) of the assembly on 20 autosomes, Z and W harboring 92% of the predicted genes (Supplementary Table 4).

Transposable elements (TEs) can constitute a large portion of a eukaryotic genome; for example, the human genome is 45% TEs\(^{23}\). The genome of the tongue sole, however, is only 5.85% TEs (Supplementary Tables 6 and 7 and Supplementary Figs. 4–7), which is in the range of the compact pufferfish genomes (365 Mb, 2.7% TEs)\(^{24}\). Clearly, small genomes will be poor in TEs. We found a low diversity of DNA transposons (2.45%, mainly hAT and Tc-Mariner) and retrotransposons (1.34%, mainly RTE from LINE) (Supplementary Fig. 4 and Supplementary Table 7). More than two-thirds of the copies are almost complete, suggesting recent activity (Supplementary Fig. 4 and Supplementary Tables 8 and 9). Penelope retroelements are highly repeated, but only a few copies are complete, suggesting an ancient invasion that is now extinct (Supplementary Table 8). Moreover, the genome is almost devoid of long terminal repeat retrotransposons (Supplementary Fig. 5 and Supplementary Table 8).

We generated a total of 12.38 Gb of transcriptome sequences to aid in gene annotation (Supplementary Table 10). Using homologous protein alignment, ab initio prediction and expression data prediction, we annotated 21,516 protein-coding genes (Supplementary Figs. 8 and Supplementary Tables 11 and 12). 99% of the predicted genes were supported by homologs in other organisms or in the transcriptome (Supplementary Figs. 8 and 9 and Supplementary Table 11). We also identified 1,439 expanded gene families and 2,743 contracted gene families by comparing the family size between the tongue sole and the teleost common ancestor (Supplementary Figs. 10–13 and Supplementary Note). In addition, we annotated 674 tRNA genes, 104 rRNA genes, 285 microRNAs and 434 small nuclear RNAs in the assembly (Supplementary Table 13).

### Adaptation to a benthic lifestyle

Flatfish are characterized by a transition from pelagic to benthic habitats, which is accompanied by a morphological transition from a symmetric to an asymmetric body shape\(^{22,25}\). As such, it is a natural paradigm of an adaptive response to a changed environment with concomitant changes in development and behavior. The availability of the genome sequence of the tongue sole allowed us to study, at a genome-wide level, differential gene expression patterns between pelagic and benthic stages (i.e., fish before and after metamorphosis). Gene Ontology (GO) categories for differentially expressed genes between fish before and after metamorphosis are related to general metabolism and immunity but also include those categories potentially associated with a metamorphic transition (for example, response to steroid hormones or organ regeneration) and environmental cues (for example, response to light stimulus and gravity) that complement adaptation to a benthic lifestyle (Supplementary Tables 14–18). When we performed likelihood ratio tests based on branch-site models\(^{26}\), 15 of these differentially expressed genes showed a highly significant signature for positive selection (\(P < 0.01\)) (Supplementary Table 19 and Supplementary Note). This group included genes (\(cp\), \(mep1b\), \(hnf4a\), \(ace2\) and \(tmem67\)) with a putative

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Statistics of sequencing and assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insert size</td>
<td>Library number</td>
</tr>
<tr>
<td></td>
<td>Female</td>
</tr>
<tr>
<td>170–800 bp</td>
<td>7</td>
</tr>
<tr>
<td>2–10 kb</td>
<td>5</td>
</tr>
<tr>
<td>20–40 kb</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
</tr>
</tbody>
</table>

Assembly | Length (Mb) | N50 (kb) | Anchored on chromosome (%) | Anchored on autosome (Mb) | Anchored on Z (Mb) | Anchored on W (Mb) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig</td>
<td>452</td>
<td>26.5</td>
<td>93.5</td>
<td>389</td>
<td>20.8</td>
<td>13</td>
</tr>
<tr>
<td>Scaffold</td>
<td>477</td>
<td>867</td>
<td>93.3</td>
<td>407</td>
<td>21.9</td>
<td>16.4</td>
</tr>
</tbody>
</table>
role in the development and function of internal organs, genes (fbrn1, cdhr2, pepd and ith2) that function in skin and cartilage remodeling, which have undergone intensive alteration during metamorphosis, and genes (mgam and cbp1) that may be related to changes in diet that accompany metamorphosis (Supplementary Fig. 14).

The most conspicuous feature of the transition from symmetric to asymmetric body shape is lateralization of the eyes to the same side21. We noted that the expression of opsin genes displays significant differences between the pelagic and benthic stages. The expression levels of rod pigment (rh1) and cone pigment (lws1), also called opn1lw1, genes, which are responsible for scotopic vision and long-wavelength sensitivity27, respectively, are significantly higher in the benthic stages (P < 0.01), whereas the expression of the middle wavelength–sensitive pigment gene (rh2) was significantly higher in the pelagic stages (P < 0.01) (Supplementary Fig. 14 and Supplementary Table 20).

These observations may suggest an enhanced sensitivity to longer wavelengths in the benthic phase, reflecting an adaptation of the visual system accompanying the shift to the benthic environment with weak light. Notably, by screening the genomes of six teleosts for vision-related genes, we found that two crystallin genes (cryybb2 and cryybb3), which encode proteins that maintain the transparency and refractive index of the lens28, have been lost specifically in tongue sole (Supplementary Table 21). We also observed that several crystallin genes (cryybb, crygs1 and crygs3) have evolved into pseudogenes, suggesting decay in the visual system (Supplementary Table 21)29. This suggestion may be explicable in terms of the emergence of dependence of the adult lateralized tongue sole on the lateral-line sense organ for prey capture, as well as papillae that are specific sense organs distributed from the tip of the head to just inside the scaly region30. The development of the strong lateral-line sense organ and species-specific sensory papillae, serving as a substitute for their decayed visual system, are likely adaptations to the benthic habitat.

Genome evolution

The tongue sole is the first flatfish to be sequenced31. To ascertain the evolutionary trajectory of the tongue sole genome, we addressed the phylogenetic branching and divergence time by comparing sets of 1:1 orthologous genes in tongue sole and other teleost fishes using human and chicken as out-groups (Supplementary Figs. 15 and 16). The phylogenetic tree based on 2,628 1:1 high-quality orthologs dates the divergence of flatfish from other bony fish to about 197 million years ago, which is consistent with previous estimates (Supplementary Fig. 16)32.

The conservation of synteny between the vertebrate genomes allowed the reconstruction of the karyotype of their last common ancestor23. We determined the orthology and paralogy relationships of the tongue sole genes and established conserved synteny with other fish genomes (Tetraodon nigroviridis, medaka and zebrafish) using human as the outgroup (Supplementary Tables 22–25 and Supplementary Note). We then clustered conserved syntenic blocks, detected rearrangements in comparison to other genomes and inferred the occurrence of such rearrangements in relation to the phylogenetic tree of the genomes being compared34. This method allowed the reconstruction of the evolutionary history of the tongue sole genome at the chromosomal level. Whole-genome duplications (WGDs) and their subsequent genomic rearrangements have important roles in genome evolution35.

Two rounds of WGD have occurred at the base of vertebrate evolution. The so-called modern fish (Teleostei) underwent an additional genome duplication event (TGD) in a common ancestral lineage from which all teleost fishes evolved36. The TGD is proposed to have facilitated teleost diversification and species radiation. Previous analyses of other fish genomes have documented the TGD34,37 and indicated eight major subsequent interchromosomal rearrangements in the ancestral teleost lineage within a short time period after the TGD34. 2,733 paralogous genes in the tongue sole genome were clustered into paralogous chromosomal regions distributed over 21 tongue sole chromosomes as result of the TGD (Fig. 1, Supplementary Fig. 17 and Supplementary Table 22). The genome-wide comparison of medaka, T. nigroviridis and zebrafish with the tongue sole genome identified the orthology relationships with the protochromosomes and three major lineage-specific chromosomal fusion events, leading to the reduced chromosome number of 21 in extant tongue sole after separation from the medaka lineage, which has 24 chromosomes (Supplementary Fig. 18 and Supplementary Tables 23–26).

The C. semilaevis Z and W chromosomes are derived from one duplicate descendant of protochromosome A for the vertebrate ancestor and A0 for the gnathostome ancestor, which gave rise to a pair of autosomes in other teleosts (Fig. 2a and Supplementary Figs. 17 and 18). Thus, the sex chromosomes of tongue sole evolved from a pair of autosomes after separation of these lineages. Notably, these
sex chromosomes share large portions a common ancestry with the chicken Z and W chromosomes (Fig. 2a and Supplementary Figs. 17 and 19). We therefore posed the question of whether this sharing is coincidence or reflects a relevant evolutionary process.

Genomic organization and evolution of Z and W

To find out how far beyond the recognized conservation of synteny the similarity of tongue sole and chicken sex chromosomes goes, we analyzed the genomic organization of C. semilaevis Z and C. semilaevis W. Although W is the largest chromosome in metaphases, fewer sequences are anchored on W than on Z, possibly because of its higher content of repetitive DNA and TEs (Supplementary Fig. 20 and Supplementary Table 27). Both telocentric sex chromosomes share a small region (~640 kb) at their distal telomeric ends containing 22 protein-coding genes and 1 pseudogene with no relevant sequence divergence (Fig. 2b and Supplementary Table 28). No indication of dosage compensation for these genes was apparent. We conclude that this region represents the pseudoautosomal region (PAR), where both sex chromosomes still pair in female meiosis and crossing over occurs at a normal rate.

The remaining region of the sex chromosomes contains 297 genes, which are also present on Z and W but show some sequence divergence (Supplementary Table 29). A uniformly distributed Ks (number of synonymous substitutions per synonymous site) of around 0.15 for this region compared to 0.0115 for the autosomes and 0.0188 for the PAR (Supplementary Fig. 21) can be taken as an indication of reduced or even absent recombination. Not surprisingly, the W chromosome displays a high content of TEs (29.94% as compared to 13.13% on Z and 4.33% on autosomes) and pseudogenes (19.74% as compared to 3.54% on Z and 2.48% on autosomes) (Supplementary Tables 27 and 30). Together with genes that are not shared with the Z chromosome, the non-PAR of the W chromosome has 317 predicted functional protein-coding genes, which is about one-third of the 904 functional genes on the Z chromosome (Supplementary Table 29). On the chimpanzee and human Y chromosomes, there are only 40–80 intact genes, as compared to 1,098 protein-coding genes on the X chromosome. 26 expressed genes have been attributed to the chicken W chromosome, whereas the Z chromosome has about 1,000 genes, as in tongue sole. This observation points to a relatively recent evolutionary origin for the tongue sole sex chromosome pair, where degeneration on the W chromosome has not progressed to a stage at which almost its entire original gene content has disappeared. We determined the age of the tongue sole sex chromosomes to be about 30 million years (Supplementary Table 31), which is consistent with the general view that fish have young sex chromosomes and avian W chromosomes and could serve to explain why there are still many intact genes in the non-recombining region of the tongue sole W chromosome.

The non-PAR of the Z chromosome exhibits only one evolutionary stratum, which is evident from the almost uniform distribution of Ks values for the Z-W gene pairs (Supplementary Fig. 21). We did not find evidence for recent transposition of genes or chromosome segments from autosomes or between the sex chromosomes.

![Figure 2](image_url)
We determined whether there is dosage compensation for Z-linked genes that are present in a single dosage in females (Supplementary Tables 32–39 and Supplementary Note). For those 763 Z-linked genes (reads per kilobase of gene per million mapped reads (RPKM) >1), male expression in whole-body (without gonad) transcriptomes was on average 1.32 times higher than female expression (95% confidence interval (CI) 1.291–1.362). This result is markedly less than the expected twofold difference for a complete absence of dosage compensation and indicates incomplete gene dosage compensation (Supplementary Fig. 22).

To elucidate the underlying mechanism, we analyzed the global gene expression pattern for sex bias of autosomal as compared to Z-chromosome genes. The Z-to-A ratios in males were close to 1 (95% CI 0.962–0.967 by Wilcoxon rank-sum test). In females, however, the ratios were 0.727–0.732 (95% CI by Wilcoxon rank-sum test), which is clearly above the 0.5 Z-to-A ratio that would be expected for no compensation. This result suggests a compensatory mechanism working through upregulation of female genes (Supplementary Fig. 23), as has also been noted previously in birds and silkworm42,43.
Sex determination by the Z chromosome
We next posed the question of whether sex determination in the
tongue sole operates through a Z-linked male-determining gene, a
W-linked female-determining gene or a combination of both. This
question is still unclear in birds but can be addressed in tongue
sole because of the possibility of producing sex reversals by high-
temperature treatment (Fig. 3a and Supplementary Note). Compared
to a low spontaneous sex-reversal rate (14%) of ZW individuals raised
at normal ambient temperature (22 °C), a rate of ~73% sex-reversed ZW
males (fertile pseudomales) appeared after high-temperature (28 °C)
treatment during the sensitive developmental period (Supplementary
Tables 40 and 41). We then crossed these pseudomales to normal
ZW females and raised the offspring under non–sex-reverting condi-
tions (22 °C). Surprisingly, there was a huge male bias, with all ZZ
fish being male but also 94% of the ZW fish developing as pseudoma-
les (Fig. 3b and Supplementary Table 40). Microsatellite markers
that distinguish paternal and maternal Z chromosomes revealed that
all second-generation pseudomales had inherited the Z chromo-
some from their sex-reversed fathers (Fig. 3c and Supplementary
Table 42). Moreover, Z-chromosomal genes in the offspring of
sex-reverted fish retained the paternal methylation pattern, imply-
ing that transgenerational inheritance of DNA methylation status
at certain loci on the Z chromosome is particularly important for
the inheritance of sex reversal (Supplementary Fig. 24). In sum-
mary, these data show that in tongue sole, sex determination operates
through a Z-encoded mechanism that determines male development.
However, we cannot exclude the possibility that temperature-induced
sex determination may operate through a different mechanism than
generic sex determination and might even involve different genes on
the Z chromosome.
For the purpose of identifying a putative primary sex-determining
gene, we could thus exclude all W-encoded female-determining fac-
tors. On the tongue sole Z chromosome, we identified four genes that
are known to be involved in sexual development in other vertebrates
(Supplementary Table 43). For sf-1, ptc1 and fst, gene expression patterns
at the sex-determination stage and gene methylation in normal males and females or in sex-reverted animals were incom-
patible with a role as a male sex-determining gene (Supplementary
Figs. 25–28). However, dmrt1 displayed many features that make it
an outstanding candidate for a master sex-determining gene (Fig. 4
and Supplementary Figs. 26–32). Only the Z chromosome contains a functional copy of dmrt1, whereas we found a heavily
corrupted pseudogenized copy on the W chromosome (Fig. 4a and
Supplementary Fig. 29). This gene is highly expressed specifically in
male germ cells and preosomatic cells of the undifferentiated gonad at
the sex-determination stage and persists at high levels during testis
development (Fig. 4b,d and Supplementary Fig. 32). These pat-
terns are paralleled by demethylation of the dmrt1 promoter region
(Fig. 4c). The expression of dmrt1 in sex-reversed ZW males was
dosage compensated by upregulation to a level that is observed in
normally developing ZZ males (Supplementary Fig. 28). All of these
observations are defining features for a dosage-dependent male sex-
determining gene, although functional proof cannot be obtained
because tongue sole is refractory to transgenic technologies, as are
all marine fish.
Notably, we found that an E3 ubiquitin ligase gene, neur13, is located
only on the Z chromosome and is absent from W. Several E3 ubiquiti-

suppression of recombination is a primary driving force of sex-
chromosome evolution immediately after a genetic determinant
on this chromosome has begun to operate. Notably, suppression
of recombination has spread over most of the chromosome in tongue
sole. The PAR of tongue sole is already very small and is comparable
to the much more advanced sex chromosomes of birds and mammals.
Accumulation of TEs also appears as an early event. The tongue sole Z
chromosome, which is in common with the mammalian X and chicken
Z chromosomes, has a higher content of TEs. The even higher TE
content of the W chromosome might explain why it, unlike the fully
differentiated W chromosome, is even larger than the Z chromosome.
The accumulation of sex-specific genes appears as another early event
of sex-chromosome evolution. A ubiquitin E3 ligase gene, neur13,
which is clearly involved in spermatogenesis, is only present
on the Z chromosome. It will be interesting to determine whether
this chromosomal assignment is conserved in other flatfishes
with WZ sex determination and hence might have contributed to
the stabilization of the female heterogametic system in this group of
fishes. Contrary to the mammalian X and chicken Z chromosomes,
the tongue sole Z chromosome does not display a lower gene den-
sity as compared to the autosomes, and we found no indication for
pronounced gene traffic between autosomes and sex chromosomes.
The necessity for gene dosage compensation appears to be an early
requirement during sex-chromosome evolution. Dosage compensa-
tion in tongue sole has already proceeded to the level of complexity
seen in birds but differs from mammals in terms of the mechanism by
which the mammalian X chromosome is compensated. The question
as to whether dosage compensation in birds represents a primordial
form that is common to other types of sex chromosome has not been
answered because avian sex chromosomes are not at an early stage of
evolution. Our finding of the same compensation mechanisms for
the relatively young sex chromosomes of tongue sole provides sup-
port for this hypothesis. However, although the dosage sensitivity of
specific sets of Z-chromosome genes is evolutionarily conserved
between zebrafish and chicken, the compensated Z-chromosome
genes in tongue sole are not the same as those in birds (Supplementary
Tables 32–39).
Sex-chromosomal gene loss and degeneration has sparked specula-
tion and discussion about ongoing decay and even the looming extinc-
tion of the human Y chromosome. However, it was recently shown
that the process of gene loss has ceased during the past 25 million
years. Thus the majority of the Y chromosome’s original gene com-
plement must have been lost much earlier. During its 30 million years
of existence, the tongue sole W chromosome has already lost about
two-thirds of its original protein-coding information. Extrapolation
of this rate of decay would mean that the W chromosome will arrive
at the gene content that is characteristic of the mammalian and avian
sex chromosomes, which are several million years old, within less
than a few tens of millions of years. This observation demonstrates
that gene loss on sex chromosomes is a very early and rapid feature

of their evolution. If the situation in tongue sole is generalizable, then from the various theoretically possible kinetics of W or Y chromosome degradation, an exponential decay that slows down in its final stages appears to be a reasonable scenario.

Parallel evolution of dmrt1 toward an important sex-determining gene in chicken, and probably also in tongue sole, provides support for the hypothesis that some ancestral chromosomes are more prone to evolve as sex chromosomes because of their particular gene content. Genes such as dmrt1 can take over a master sex-determining role more easily than other genes, as indicated by the fact that this gene evolves to be the master male sex-determining gene independently in birds, Xenopus laevis, medaka, and probably also tongue sole. It is noteworthy that of the 58 sex-related genes that we analyzed in tongue sole, we found more than twice the average number on the sex chromosome pair.

In summary, our sequencing of the genome of the half-smooth tongue sole provided insights into adaptation to its benthic life and has allowed the first comprehensive view of the structure and evolution of W chromosomes. The first full sequence of the Z chromosome outside of birds showed an unexpected parallel evolution of the same ancestral autosomal gene, dmrt1, which has been shown to be the primary male sex-determining gene in chicken. In addition to the evidence presented here, future studies in tongue sole will be necessary to provide further clues for such a function of dmrt1. Studies on the W chromosomes of other species will be informative with respect to the question of whether W chromosome evolution is just a female version of the much better studied Y chromosome evolution.


METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. The tongue sole whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the project accession PRJNA73987 (with the associated accession code AGRG00000000). RNA-Seq and miRNA sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA047922 and SRA122228), and BAC sequences have been deposited in the GSS division of GenBank (JQ003878–JQ003881).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
We thank H. Lin, F. Zhao, J. Lei, Z. Zhu and Q. Wang for their support of the tongue sole genome project. We thank Z. Liu, Z. Zhuang, R. Li, Z. Jiang, P. Ni, X. Qian and T. Xu for discussion. We thank J. Zhai, S. Liu and Z. Li for their help in the culture of tongue sole families. We thank M. Yi and L. Huang for their help with chromosome FISH. We are grateful to the faculty and staff of Yellow Sea Fisheries Research Institute, CAFS and BGI-Shenzhen, whose names were not included in the author list but who nevertheless contributed to this project. This work was supported by grants from the Special Fund for Agro-Scientific Research in the Public Interest (200903046) of China to S.C., the State 863 High Technology R&D Project of China (2012AA020203) to S.C., the State 863 High Technology R&D Project of China (2012AA10A403-2) to C.S., the National Nature Science Foundation of China (31103057 and 31072202) to S.C., the National Nature Science Foundation for Young Scholar (41006107) to C.S., the Taishan Scholar Project Fund of Shandong of China to S.C., the JSPS RONPAKU (Dissertation PhD) program to C.S., the Basic Research Program Supported by Shenzhen City (JCYJ2012060119) to G.Z., the Key Laboratory Project Supported by Shenzhen City (CXR20090310066A) to J.W., the Shenzhen Key Laboratory of Transomics Biotechnology (CXR201108250096A) to J.W. and the Shenzhen Key Laboratory of Bank Gene for National Life Science.

AUTHOR CONTRIBUTIONS

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/.

ONLINE METHODS

Genome sequence and assembly. High-quality genomic DNA was extracted from one adult female and one adult male tongue sole. Then, according to the Illumina standard operating procedure, we constructed 15 paired-end libraries for the female and 11 paired-end libraries for the male (170 bp to ~40 kb) (Supplementary Tables 1 and 2). Paired-end sequencing was performed on an Illumina HiSeq2000 for each library to produce the raw data. We then filtered artificial and low-quality reads to obtain a set of usable reads that contained 857.5 million and 730.0 million reads, representing 63.86 Gb and 46.67 Gb of data for the female and male individuals, respectively, covering in total 212 times the genome (Supplementary Fig. 1 and Supplementary Tables 1 and 2).

In addition, we corrected sequencing errors for the 17-mers with a frequency of 857.5 million and 730.0 million reads, representing 63.86 Gb and 46.67 Gb of artificial and low-quality reads to obtain a set of usable reads that contained RepeatScout 60, to construct a

Online methods set containing 21,516 functional genes (Supplementary Table 12). For gene annotation, we detected the motifs and domains by InterProScan74 against publicly available databases, including Pfam, PRINTS, PROSITE, ProDom, SMART and PANTHER, and then retrieved GO annotation from the results of InterProScan. From the reference gene set, 17,890 and 14,935 genes could be annotated by InterPro and GO annotation, respectively. We also annotated 20,265 genes by searching the Swiss-Prot database using BLASTP at E value < 1 × 10−5.

Sex-chromosome identification and chromosomal assignment of scaffolds. With the same sequencing coverage, the depth of Z-linked scaffolds in the non-PAR region in the female is expected to be half that in the male Z chromosome, in female autosomes and in male autosomes (Zf = 1/2Zm = 1/2Af = 1/2Am), where f is female, m is male, Z is the Z chromosome, and A is the autosome. We thus identified 26 and 126 Z-linked scaffolds in the male and female assemblies, respectively. For the W-linked scaffolds that displayed the female assembly only, we expected that those scaffolds should not be covered by reads from the male genome, and the sequencing depth should be about half of the average value of autosomes in the female genome reads. Using this method, we identified 306 W-linked scaffolds in the non-PAR region representing a 16.4 Mb length with a scaffold N50 of 128 kb. Considering the interference of W reads and the high-quality assembly of the male genome relative to the female genome for the Z assembly (scaffold N50 of 1,305 kb as compared to 357 kb, respectively), we chose scaffolds from the male assembly as the Z-linked scaffolds in the final version. For other scaffolds representing autosomes and the W chromosome and other undetected Z-linked scaffolds (if any), we used the female version. We then used BLASTX (E value < 1 × 10−5, identity ≥ 29%) and aligning rate > 50%) to map molecular markers to scaffolds. We linked scaffolds onto chromosomes with a string of 100 ‘N’s representing the gap between two adjacent scaffolds based on the high-resolution genetic map. In total, 944 scaffolds with 445 Mb length (93.3% of scaffolds in length) were anchored to 22 chromosomes, representing the 20 autosomes, Z and W (Supplementary Table 4).

Reconstruction of ancestral vertebrate chromosomes. We first performed an all-against-all comparison between tongue sole protein sequences and human protein sequences using BLASTP (E value < 1 × 10−10) to identify the paralogs in the tongue sole genome (Supplementary Table 22). Then we identified reciprocal best-match orthologous genes between tongue sole and each other fish using BLASTP at E value < 1 × 10−10, 14,231, 14,310 and 13,084 orthologous genes were identified for T. nigroviridis, medaka and zebrafish, respectively (Supplementary Tables 23–25). We further identified deeply conserved synteny by synteny analysis and then deduced the ancestral teleost karyotype by considering the results from the human genome as an outgroup, as was performed in a previous study (Supplementary Table 26)17. The ancestral vertebrate ancestor consisting of ten protochromosomes was reconstructed, and an evolutionary hierarchy from the ancestral vertebrate ancestor to the genomes of human, chicken and medaka was assigned as reported previously (Supplementary Figs. 17 and 18)13.

Dosage compensation. We used RNA-seq data from the whole bodies (without gonad) of male and female fish to test for dosage compensation of the Z chromosome. The male-to-female (M:F) gene expression ratio was used to measure the dosage compensation level for every gene in the female relative to the normal male calculated as the RPKM ratio of each gene between the two samples. Only genes with RPKM > 1 in both the normal male and female were considered. The Z-to-autosomes (Z:A) expression ratio for every gene in the Z chromosome was calculated by dividing the RPKM of the gene by the median RPKM of all autosomal genes. After filtering out genes with RPKM < 1, we calculated the M:F expression ratio for all Z-linked genes (Supplementary Note).

Production and genetic analyses of sex reversals. Treatment of tongue sole with high temperature (28 °C) during the critical developmental stage directly affects the sex ratio of the progeny75. Briefly, about 3,000 larvae were collected at 25 d after hatching (dph) and allocated evenly into three tanks (3 m3) at 23 °C. The sea water was heated up gradually and maintained 28 °C until 100 dph. The genetic sex identification of larvae was performed by PCR analysis using the sex-linked SSR marker, Csef-SSR1, which produced one band of 206 bp for ZZ and two bands of 206 bp and 218 bp for ZW76. The phenotypic sex was identified by routine histology of gonads for the presence of oocytes or spermatoocytes (Supplementary Note)77. Under treatment with high temperature, about 70% of ZW individuals developed as male, whereas under normal conditions (22 °C), the spontaneous sex-reversal rate was about 14%. The sex-reversed fish produced by high temperature could cross with normal females. Surprisingly, there was an extremely male-skewed sex ratio (>94%) in offspring of pseudo-male families raised under normal conditions (22 °C) (Supplementary Table 40). This was caused by an extremely high sex-reversal rate of genetic females to phenotypic males (~94% compared to the sex-reversal rate of 14% under normal conditions). A similar phenomenon was also detected in the offspring of pseudo-male families crossbreeding of spontaneous sex-reverted males with normal females.

FISH mapping. Metaphase chromosome spreads were obtained from the head kidney of female and male fish and were then passed through an ethanol series...
and air dried. The dmrt1 BAC DNA was extracted and labeled by nick translation using a Nick Translation System (DIG-Nick Translation Mix, Roche). The BAC FISH probe contained 1 µg of labeled dmrt1 BAC DNA, 50 µg of sonicated salmon sperm DNA and 10 µg of Cot-1 DNA. After hybridization in a moist chamber at 37 °C for 24 h, chromosome slides were subjected to a series of washing steps (2× SSC (TaKaRa) for 5 min, 50% formamide for 5 min and 1× SSC for 5 min). Signal detection and amplification were performed using sheep anti-digoxigenin and FITC donkey anti-sheep (Dig-Nick Translation System using Hotstart Taq polymerase (Qiagen), and the

RT-PCR and bisulfite PCR (BS-PCR) analysis. Total RNA was isolated and purified from all the samples using a traditional phenol method78, and the RNA concentration was measured using Nanodrop technology. Primers for quantitative RT-PCR (qRT-PCR) analysis were designed using the Primer

Supplementary Note). Additional results are shown in Supplementary Figures 36–38 and Supplementary Tables 44–55.