Transcriptome-Wide Patterns of the Genetic and Expression Variations in Two Sympatric Schizothoracine Fishes in a Tibetan Plateau Glacier Lake

Juan Chen,[†,1,2] Liandong Yang,[†,1] Renyi Zhang,[3] Severin Uebbing,[4] Cunfang Zhang,[3] Haifeng Jiang,[1,2] Yi Lei,[1,2] Wenqi Lv,[1,2] Fei Tian,*[,3] Kai Zhao,*[,3] and Shunping He*[,1,5,6]

[1]The Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China

[4]Department of Genetics, Yale University School of Medicine, New Haven, CT

[5]Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, China

[6]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

†These authors contributed equally to this work.

*Corresponding authors: E-mails: tianfei@nwipb.cas.cn; zhaokai@nwipb.cas.cn; clad@ihb.ac.cn.

Abstract

Sympatric speciation remains a central focus of evolutionary biology. Although some evidence shows speciation occurring in this way, little is known about the gene expression evolution and the characteristics of population genetics as species diverge. Two closely related *Gymnocypris* fish (*Gymnocypris chui* and *Gymnocypris scleracanthus*), which come from a small glacier lake in the Tibetan Plateau, Lake Langcuo, exist a possible incipient sympatric adaptive ecological speciation. We generated large amounts of RNA-Seq data from multiple individuals and tissues from each of the two species and compared gene expression patterns and genetic polymorphisms between them. Ordination analysis separated samples by organ rather than by species. The degree of expression difference between organs within and between species was different. Phylogenetic analyses indicated that the two closely related taxa formed a monophyletic complex. Population structure analysis displayed two distinctly divergent clusters of *G. chui* and *G. scleracanthus* populations. By contrast, *G. scleracanthus* population genetic diversity is higher than that of *G. chui*. Considerable sites of the two populations were differentiated with a coefficient of $F_{ST} = 0.25$–0.50, implying that a small proportion of loci nevertheless exhibited deep divergence in two comparisons. Concomitantly, putatively selected genes during speciation revealed functional categories are enriched in bone morphogenesis, cell growth, neurogenetics, enzyme activity, and binding activity in *G. chui* population. In contrast, nutrition and localization were highlighted in *G. scleracanthus*. Collectively, morphological traits and dietary preference combine with genetic variation and expression variation, probably contributed to the incipient speciation of two sympatric populations.

Key words: sympatric speciation, transcriptomics, gene expression, population genetics.

Introduction

Speciation—the origin of new species—is the source of the diversity of life. Speciation is thought generally to involve the gradual accumulation of genetic differences between geographically isolated (allopatric) populations through selection or genetic drift (Friesen et al. 2007). The origin of species, including sympatric speciation (Via 2001), parapatric speciation (Gavrilets et al. 2000), peripatric speciation

(Odeen and Florin 2002), and allopatric speciation (Coyne and Orr 2004; April et al. 2013), has challenged biologists for over two centuries, since it was raised as the "mystery of mysteries" by Darwin (1859). So far, little work has focused on gene expression variation, although gene expression underlying functional variation has been shown to be important in speciation and diversification (Jeukens et al. 2010; Wolf et al. 2010). Understanding the causes and consequences of speciation, including genetic separation and phenotypic differentiation, requires a combination of genomic and morphological data to capture faint signals of both genetic and expression variations. Recent advances in sequencing and bioinformatics have led to the wide use of transcriptomics technologies to study the fundamental processes underlying genetic differentiation and adaptation of species. Sympatric speciation, the process by which populations can become genetically isolated without geographical separation, remains one of the most contentious concepts in evolutionary biology (Darwin 1859). In recent years, a growing body of theoretical studies as well as a proliferation of empirical examples have established that sympatric speciation is possible and may be more common than previously thought (Dieckmann and Doebeli 1999; Higashi et al. 1999; Tregenza and Butlin 1999; Via 2001; Bolnick and Fitzpatrick 2007). Until now, a myriad of taxa have been suggested to have diverged primarily by sympatric speciation, including phytophagous insects (Berlocher and Feder 2002), palms on Lord Howe Island (Savolainen et al. 2006), cichlid fish in a Nicaraguan crater lake (Barluenga et al. 2006; Kautt, Machado-Schiaffino, Meyer 2016; Kautt, Machado-Schiaffino, Torres-Dowdall, et al. 2016), sticklebacks (Rundle et al. 2000), spiny mice (Hadid et al. 2014; Li et al. 2016), blind mole rats (Hadid et al. 2013; Li et al. 2015), fruit flies (Hubner et al. 2013), beetles (Sharaf et al. 2013), wild barley (Nevo 2014), and *Gymnocypris* fish in Lake Sunmcuo (Zhao et al. 2009).

The schizothoracine fish (Teleostei: Cypriniformes: Cyprinidae), commonly known as "mountain carps," consists of 15 genera and more than 100 species all over the world (Mirza 1991). Based on specialization in scales, pharyngeal teeth and barbels (Cao et al. 1981) divided schizothoracine fishes into three grades: primitive grade, specialized grade, and highly specialized grade. They are widely distributed in the Qinghai–Tibetan Plateau (QTP) and its peripheral regions, representing the largest and most diverse taxon of the QTP ichthyo fauna. The elevation in this area ranges from 700 to 5,000 m, and the environment is characterized by hypoxia and low temperatures (Cao et al. 1981). Previous studies on the schizothoracine fish mainly phylogenetic analysis or high-latitude environments adaptation to cold or hypoxic conditions (Cao et al. 2008; Li et al. 2013; Guan et al. 2014; Xia et al. 2016; Qi et al. 2018). However, phylogenetic studies based on only mitochondrial data or incomplete taxon sampling revealed many potential misclassifications in the

schizothoracine fishes at the genus/species level (He and Chen 2006; Yonezawa et al. 2014; Zhang, Chen, et al. 2016; Tang et al. 2019). The most current speciation hypothesis is the divergence of species resulting from geographical isolation (Cao et al. 1981; Yun-Fei 1991); however, morphological analyses and population genetics studies have found sympatric speciation occurred in the same glacier lake in the Tibetan Plateau (Chen et al. 1982; Zheng et al. 2004; Zhao et al. 2009). Glacier Lake Langcuo, located on the northern of the Yarlung Zangbo River, at an altitude about 4,300 m (Cao et al. 1981), represents an ideal location to study sympatric speciation. It is small (area, ~12 km^2; max depth, ~33 m), has a homogenous habitat with few fish fauna habitat, and is completely isolated. Geological data show that Lake Langcuo is a small plateau glacial lake, which used to be an outflow lake. However, as the plateau gradually rose, throughout the Late Pleistocene, it become a stable closed inland lake, inside of which geographical isolation cannot realistically occur (Wang and Dou 1998; Li 2000).

Two sister species of naked carp, *Gymnocypris chui* Tchang, Yueh, and Hwang, 1964 and *Gymnocypris scleracanthus* (Cyprinidae: Schizothoracinae), are endemic fish that are sympatrically distributed in Lake Langcuo in Tibet, China (Wu YF and Wu CZ 1992; Zhang et al. 1995; Chen and Cao 2000). Morphological differences have been investigated (Wu YF and Wu CZ 1992) and showed that they are clearly morphologically distinct. *Gymnocypris chui* has a more oblique mouth and dense gill rakers. Previous cytogenetic analyses showed that *G. chui* and *G. scleracanthus* have the same chromosome number ($2n = 92$), but different karyotypes (Zhang, Tang, et al. 2016), which may have played a role in postzygotic reproductive isolation during the speciation process (Cardoso et al. 2013). In addition, an early phylogenetic study indicated that *G. chui* and *G. scleracanthus* (based on mitochondrial genes) grouped more closely than other schizothoracine fish, indicating their close relationship. Taken together, previous work has shown that *G. chui* and *G. scleracanthus* exist a possible incipient sympatric adaptive ecological speciation.

This study is a continuation of current studies of two closely species with possible sympatric speciation which fulfill the criteria that a firm corroboration for the origin of one or more species in sympatry. Here, we applied transcriptome sequencing of different organs in population samples of *G. chui* and *G. scleracanthus* to explore the character of gene expression evolution as species diverge and to capture faint signals of both genetic variation and expression variation. Meanwhile, we analyzed cDNA single nucleotide polymorphisms (SNPs), which used to examine population genetic structure, and explore putative selected genes related to species speciation. Our prime goals in this study were to expand and deepen our understanding of sympatric speciation in plateau lakes through the insights of the entire transcriptome analysis.

Materials and Methods

Fish Sampling and RNA Extraction

Experiments involving animals in this study were conducted in accordance with the Laboratory Animal Management Principles of China. All experimental protocols were approved by the Ethics Committee of the Institute of Hydrobiology, Chinese Academy of Sciences. We obtained 52 schizothoracine fish samples from three groups distributed on the QTP as listed in supplementary table S1, Supplementary Material online. Fishes were identified based on the taxonomic description by Chen and Cao (2000) and Wu YF and Wu CZ (1992). To maximize the sampling of genes, tissue samples from five major organs including heart, liver, kidney, muscle, and brain were collected immediately and placed in liquid nitrogen during field and transportation, subsequently transferred to the −80 °C freezer for long-term storage. For each tissues collected from a single fish, up to 100 mg of tissue was flash-frozen in liquid nitrogen and used for total RNA extraction with the TRIzol reagent (Invitrogen, Carlsbad, CA), following the manufacturer's protocol and performing the optional DNase digestion step. The quality and quantity of RNA extracts were assessed with an Agilent 2100 BioAnalyzer, and samples with low integrity values were re-extracted. RNA extracts with RNA concentration $\geq$200 ng/μl, total $\geq$10μg (5 μg for a single build); OD260/280 between 1.8 and 2.2 for each sample for transcriptome sequencing. In addition, six wild male schizothoracine fish (three *G. chui* and three *G. scleracanthus*) were collected form the small glacier lake Langcuo in Xigaze, Tibet.

cDNA Library Construction and Transcriptome Assembly

Five nonmixed cDNA libraries (heart, liver, brain, kidney, and muscle) were built and assessed, and mRNA was enriched using beads with oligo (dT) and fragmented using fragmentation buffer. The first cDNA chain was synthesized using random hexamers, and the second chain was synthesized with buffer, dNTPs, RNase H, and DNA polymerase I. Each of the libraries was amplified by polymerase chain reaction (PCR). The conditions for PCR amplification were an initial denaturation at 72 °C for 3 min, 98 °C for 30 s, ten cycles of denaturation at 98 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 5 min, and a final extension at 72 °C for 5 min. The libraries were quantified by RT-PCR, and transcriptome sequencing was performed on an Illumina HiSequation 2500 platform. Short sequence reads of paired end 150 bp were generated. The RNA-Seq data were filtered and trimmed to control the quality of raw reads using FastQC v0.11.8 (available from http://www.bioinformatics.babraham. ac.uk/projects/fastqc; last accessed November 15, 2018). We used CUTADAPT v2.1 (Martin 2011) and Trimgalore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; last accessed December 7, 2018) to remove illumine adaptors,

and trim reads. Only paired-end reads for which either read was longer than 25 bp after trimming were retained for subsequent analysis. After that, all clean reads were combined together to obtain an integrated transcriptome, which we used for variant calling. TRINITY de novo assembly was performed using default parameters (Grabherr et al. 2011). Then, the CD-HIT-EST software (Li and Godzik 2006) was used to remove the redundancy with an identity threshold of 0.95.

Phylogenetic Analyses

Transcriptome sequencing (RNA-Seq) provides massive transcript information from the genome. Phylogenetic reconstructions based on RNA-Seq are more efficient and cost-effective than traditional approaches. Here, all 52 schizothoracine fishes and other teleosts including *Danio rerio*, *Ctenopharyngodon idellus*, *Sinocyclocheilus grahami*, and *Percocypris pingi* RNA-Seq data set were applied in the phylogenomic analysis. A homologous gene tree was built with RAxML software, v 8.0.20, by implementing the maximum likelihood method (Stamatakis 2014). Branch support was evaluated using 1,000 bootstrap replicates. The phylogenetic tree structure was visualized by FigTree v1.4.3. The Bayesian relaxed-molecular clock method, implemented in the MCMCTree program in PAML v.4.8a (Yang 2007), was used to estimate the divergence time between different species. Two calibration time points based on records of *C. idellus–D. rerio* ($\sim$36–72 Ma), and *D. rerio–S. grahami* ($\sim$87.8–124.7 Ma) (http://timetree.org/; last accessed March 1, 2019), were used as constraints in the MCMCTree estimation. The MCMC process was run for 5,000,000 steps and sampled every 5,000 steps.

Expression Analyses

Consensus sequences of the contigs were used as a reference for the RNA-Seq analysis. The expression abundance of each gene was described in terms of the expected fragments per kilobase of transcript per million fragments mapped (FPKM) and was calculated using RSEM1.3.1 (Li and Dewey 2011). We extracted raw read counts for each gene from RSEM and normalized them to control for differences in sequencing depth across different samples using the TMM method. FPKM values were then further normalized using the procedure described in Hart et al. (2013), and genes were defined as being expressed using the cutoff of 0.125 zFPKM suggested in that study. We also extracted raw read counts for analyses of differential expression with the edgeR package (Robinson et al. 2010) using a minimal fold-change of 2 and an adjusted *P*-value cutoff of 0.001. Euclidean and correlation-based distances have been proposed for measuring expression divergence. Euclidean distances from zFPKM expression values were calculated by nonmetric multidimensional scaling (NMDS) as implemented in the R package MASS v. 7.3–51.1 (Venables and Ripley 2002) for ordination

plotting. The scaling procedure was iterated until convergence. We used analysis of variance (ANOVA) on NMDS axes values for samples with "organ," "species," and their interactions as fixed effect factors.

SNP Calling

SNPs were identified from these mapped reads as follows. First, we implemented the BWA-MEM algorithm in Burrows–Wheeler Alignment (BWA) tool (Li and Durbin 2009) to align clean reads from each individual against the transcript consensus, which we preconnected for all the sequences into five large fragments as five hypothetical chromosomes. The generated SAM format files were transformed into binary BAM files using SAMtools v 1.5 (Li et al. 2009), which was followed by BAM files sorting and head adding. Next, rmdup function of Samtools software was used to remove reads that were duplicated in library preparation or sequencing. In order to minimize the mapping error, we need to recalibrate the area where potential sequence insertion or sequence deletion exists. So, we used the Genome Analysis Toolkit (McKenna et al. 2010) to perform local realignment around insert or deletion (INDELs) to avoid alignment errors, which contains two steps: 1) the Realigner Target Creator to identify regions and 2) the Indel Realigner to realign the regions. The "mpileup" command was executed to identify raw SNPs with the parameters as "-q 20 -Q 20 -C 50 -m 2 -F 0.002." The generated Variant Calling Format (VCF) files were filtered through the command "varFilter" with parameters as "-d 20 -D 140 –w 5 –e 0.05" of the vcfutil.pl script in the BCFtools v1.5 (Li et al. 2009) package, removing coverage depth <20 or >140, or root mean square of mapping quality <10, or an SNP <5 bp from the adjacent INDELs. SNPs deviating from Hardy–Weinberg equilibrium ($P < 0.05$) were excluded from subsequent analysis. Then, we only retained SNPs without INDELs using VCFtools (Danecek et al. 2011). Through the above series of processes, 932,370 SNPs were retained. Finally, we excluded SNPs that could not meet the following criteria: 1) minor allele frequency >0.1 and 2) maximum per-SNP missing rate <0.1. After this step, there were 477,146 SNPs in the genetic diversity analysis data set. The nucleotide diversity ($\theta\pi$), population-differentiation statistic (F_{ST}), and Tajima's D population genetic statistics were calculated by VCFtools (Danecek et al. 2011) using a sliding window approach, with a window size of 10 kb and a step size of 2,500 bp. The significance of the difference between two populations was tested with the Mann–Whitney U test.

Linkage Disequilibrium and Population Structure Analysis

Linkage disequilibrium (LD) refers to the nonrandom associations of alleles at different loci. To estimate the LD patterns between the two populations, we used the program PopLDdecay v3.40 (Zhang et al. 2019) to calculate the correlation coefficient (r^2) between any two loci, with the "-InVCF" option in a 50-kb window. We used nonparametric principal component analysis (PCA) approach to investigate the genetic structure of the species based on SNP data. The PCA was performed for six individuals using GCTA v1.91.7 with the parameter "-pca 2." We also reconstructed a phylogenetic tree using the concatenation of all sequences using a neighbor-joining (NJ) method (Saitou and Nei 1987; Zhang and Sun 2008). The variant calling format was converted to binary ped format using PLINK v1.07 (pngu.mgh.harvard.edu/~purcell/plink/), then converted to FASTA format using a custom home Perl script. The unrooted NJ tree was constructed with TreeBeST (Vilella et al. 2009) using the matrix of pairwise genetic distances based on 1,000 bootstrap replicates. FigTree v1.4.3 and MEGA v7.0 (Kumar et al. 2016) were used to visualize the phylogenetic trees.

Putatively Selected Genes during Speciation

To identify genomic regions that may have been subject to selection during speciation, we computed some genetic statistics. The nucleotide differentiation F_{ST} (Weir and Cockerham 1984) between the populations and Tajima's D (Tajima 1989) and nucleotide diversity ($\theta\pi$) (Nei and Li 1979) for each population were calculated using a sliding window approach with a window size of 10 kb and a step size of 2.5 kb by Vcftools (Danecek et al. 2011). The $\theta\pi$ log-ratio was calculated as ln($\theta\pi$_pop1)/ln($\theta\pi$_pop2), where $\theta\pi$_pop1 and $\theta\pi$_pop2 are the nucleotide diversity values for *G. scleracanthus* and *G. chui*, respectively. Subsequently, F_{ST} values and $\theta\pi$ were sorted in descending order separately. The putative selected regions were screened from the overlap of the top 5% log-odds ratios of both $\theta\pi$ and F_{ST}. Finally, we captured the putatively selected genes on the basis of their representation for both putative selected regions and differentially expressed (DE) genes. Through these steps, we screened 212 putative selection candidate genes, used for subsequent analysis and discussion. The GORILLA Functional Annotation tool (Eden et al. 2009) was used to perform gene functional enrichment analyses.

Results

Monophyly of Sympatrically Distributed Sister Species *G. chui* and *G. scleracanthus*

To infer the monophyly of sister taxa *G. chui* and *G. scleracanthus* and the phylogenetic relationships of all members of the Schizothoracinae, we constructed a robust and well-resolved phylogenetic tree, using high-quality transcriptome sequencing data, which covered 52 schizothoracine fishes from three groups and four other teleosts, distributed in different river systems of the QTP (fig. 1B). All species of the Schizothoracinae clustered into two major
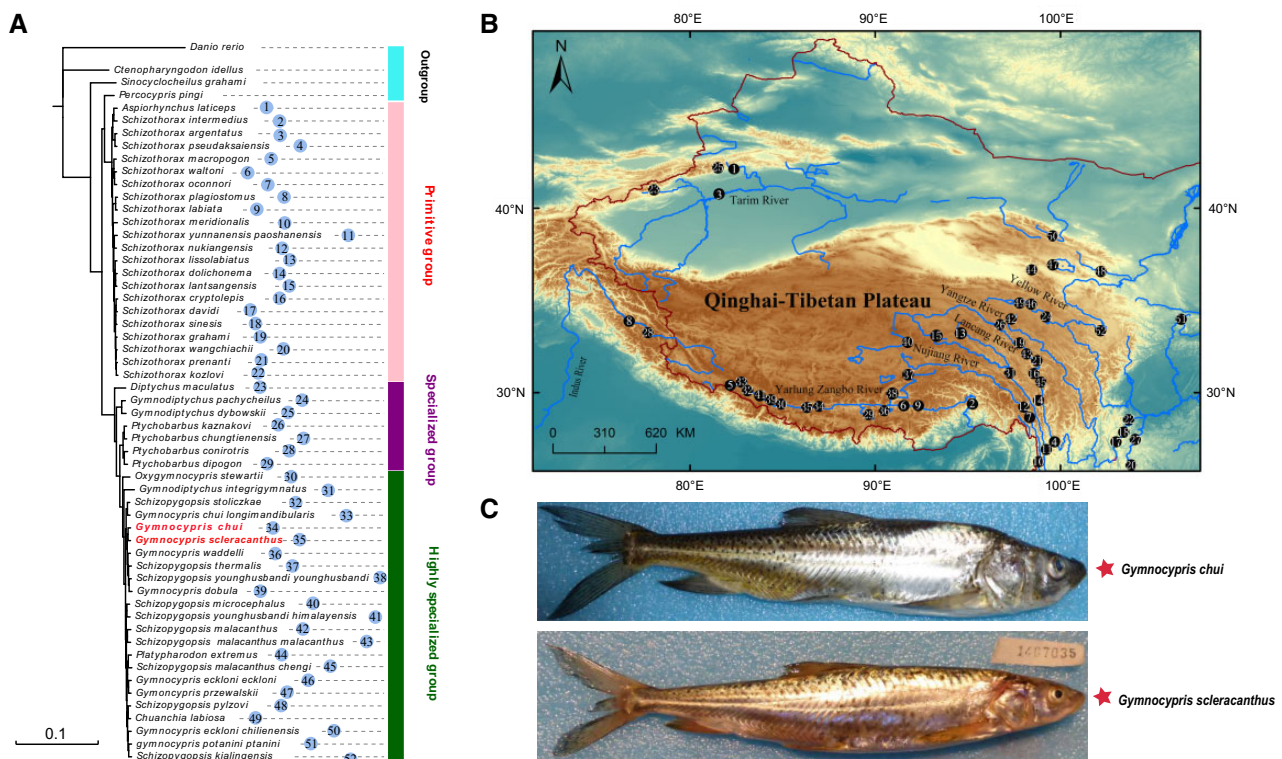
Fig. 1.—Phylogenetic tree and geographical distribution. (A) Phylogenetic tree of schizothoracine fish on the QTP using the maximum likelihood method. Different color blocks represent three groups, primitive, specialized, and highly specialized. (B) Geographical distribution of schizothoracine fish samples involved in this study (dots indicate sampling locations). The names of the main rivers in the Tibetan plateau are shown, including the Yangtze, Yellow, Yarlung Zangbo, and Indus rivers. Maps in this figure were generated using ArcGIS v10.5 and modified in Photoshop. (C) Pictures of *G. chui* and *G. scleracanthus*.

clades, one containing the primitive group and the other clade comprising the specialized and highly specialized groups (fig. 1A). A maximum likelihood phylogeny was constructed based on these data. *Gymnocypris chui* and *G. scleracanthus* were embedded in the lineages of the highly specialized grade schizothoracine fishes, and is placed a monophyletic group with 100% BP (supplementary fig. S1, Supplementary Material online). Moreover, these two species were included in a monophyletic group with *Gymnocypris waddelli* and *Schizopygopsis thermalis*, *Schizopygopsis younghusbandi younghusbandi*, and *Gymnocypris dobula*. The reliable phylogenetic relationship strongly indicates that *G. chui* and *G. scleracanthus* exist closest sister relationship of kinship, and genetically maintained monophyletic. The estimated time of divergence between *G. chui* and *G. scleracanthus* was 0.0286 Ma during the Pleistocene period by using the 95% credibility interval (supplementary fig. S2, Supplementary Material online). During this period, the QTP is undergoing the Great Lakes period (0.04–0.025 Ma), forming many closed lakes under the background of the plateau uplift (Li 2000). Thus, combined with previous evidence (Zhang, Tang, et al. 2016), we demonstrated that these two species have a sympatric distribution, are monophyletic sister species, and are in an ecological setting where allopatric speciation is unlikely.

Ordination Separates Samples According to Organ

Using an NMDS algorithm, we carried out ordination analysis based on Euclidean distances from zFPKM expression values between all organ/individual combinations. The NMDS reflected that samples clustered mainly by organs and not by species, with some amount of overlap between the two species, demonstrating unique and organ-specific transcriptome profiles (fig. 2B). In addition, the ANOVA also confirmed that organ identity had consistently strong effects on the data, whereas the effect of species was not significant on any NMDS axes. We further confirmed even their interactions as fixed effect factors, no significantly effects (supplementary table S2, Supplementary Material online). When we made separate ordination plots for individual organs, the muscle organ resolved fully according to species; however, the others could not resolve through this way (fig. 2A). This phenomenon seems to be consistent with the morphological differences in the two sympatric distribution species. In general, there was a very weak but detectable signal of interspecies difference in gene expression.
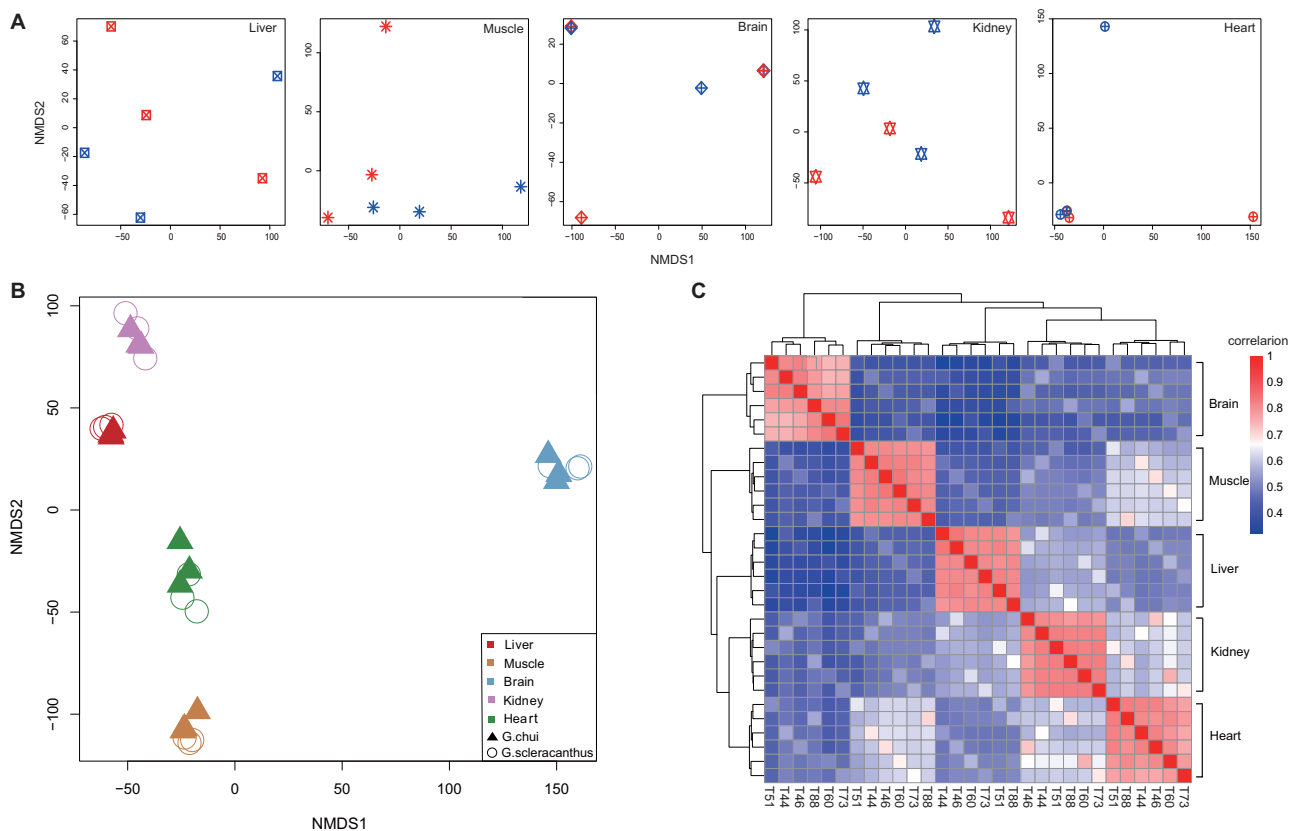
FIG. 2.—Expression analysis in two sympatric species. (A) Separate NMDS ordination plots of all analyzed tissues. The red color represents *G. scleracanthus*, and the blue color represents *G. chui*. (B) Ordination of gene expression data from *G. chui* and *G. scleracanthus* samples. (C) Tissue heat map and hierarchical clustering of gene expression data. Color key indicates the level of expression "relatedness" between tissue types, the red the color, the more similar the pattern of gene expression, with blue being the most distant (or least correlated). The tissues are clustered based on the Pearson correlation coefficients distance matrix.

Differences in Gene Expression Levels among Individuals and Species

To investigate whether there were gene expression differences during the differentiation and separation of populations and species, we analyzed gene expression differences in five organs between two closely related schizothoracine fish, which may exist a possible incipient sympatric adaptive ecological speciation. When tissues were clustered on their gene expression patterns, using the full normalized count matrix for all genes, tissues were fully grouped together into clusters reflecting their biological relationship (fig. 2C). The primary separation was between brain and other tissues, which may reflect the hyperactive transcriptional machinery in the brain. Of note, we found that heart and kidney tissues were consistently highly correlated, and the muscle and heart showed strong correlations; therefore, they are expected to share similar functionality in species speciation. Furthermore, the expression variance among individuals varied considerably between organs and was largest in the heart and smallest in the kidney (supplementary fig. S3A, Supplementary

Material online). Expression variance for individual genes was correlated among organs, with Spearman ρ ranging from 0.287 between brain and liver to 0.645 between muscle and heart (supplementary table S3, Supplementary Material online). Concomitantly, we found strong correlations between muscle and heart, which is consistent with the common understanding that a large proportion of heart is composed of cardiac muscle tissue. Within-species expression variance was correlated with the difference in log2 mean expression level between species for all organs. We found that within-species expression variance varied considerably between organs and was largest in muscle and smallest in the kidney (supplementary fig. S3B, Supplementary Material online).

Gene expression change has been considered a major driver of species differentiation and evolution (King and Wilson 1975). To identify organ-enriched DE genes during speciation, we used a t-test with a Bonferroni-corrected P-value ≤ 0.05 to generate a list of data. Taken together, we obtained 1,328 upregulated DE genes and 1,403 downregulated DE genes. Of these organ-enriched DE genes, 1,709

(62.58%) were detected largely in the brain, 362(13.18%) in the kidney, and 248(9.08%) in the liver (supplementary fig. S4 and table S4, Supplementary Material online). The numbers of organ-enriched DE genes identified were185 and 227 in muscle and heart, respectively. We conducted gene ontology (GO) enrichment analysis of organ-enriched DE genes in each organ type. In general, the GO enrichment and the selection and ranking of the pathways based on organ-enriched genes were highly consistent with the biological functional activities of the organ for which the genes were enriched. For example, brain-enriched DE genes were associated with neurophysiological processes, including neurohypophyseal hormone activity and axon guidance, whereas heart-enriched genes were associated with ATPase activity and galactose metabolic process. Of particular surprise is that we found some GO terms that participate in glucose metabolism, including galactose catabolism, galactose metabolic process, monosaccharide catabolism, and hexose catabolism. This is likely in part because the two taxa of *Gymnocypris* in lake Langcuo are clearly morphologically distinct, and *G. chui* is characterized by a more oblique and more deeply arched mouth cleft (supplementary table S6, Supplementary Material online), which results in dietary preferences linked to the DE genes.

To infer genes specifically expressed in one species during sympatric speciation, we extracted genes that were expressed in at least three individuals of one species but completely absent from all organs in all individuals of the other species (Uebbing et al. 2016). Through this method, we found 28 genes exclusively expressed in *G. chui* and 18 genes exclusively expressed in *G. scleracanthus* (supplementary table S7, Supplementary Material online). Interestingly, of 28 genes exclusively expressed in *G. chui*, tbx15, prg4b, *LBX1*, atp1a3a were related to skeletal development, ATP synthesis. *Gymnocypris scleracanthus* exclusively expressed genes involved in several metabolic pathways. Genes that are exclusively expressed in one of the two species could potentially represent genetic differences. It is therefore likely that the independent or combined effects of gene expression combined with morphological and genetics differences.

Genetic Diversity and Population Structure

After variation calling, 429,785 and 432,374 SNPs were identified from the *G. chui* and *G. scleracanthus* populations, respectively. We also calculated the SNPs that were unique to each, and there were 44,772 and 47,361 SNPs unique to *G. chui* and *G. scleracanthus* (fig. 3A). The SNPs unique to the *G. scleracanthus* population were located in 29,910 genes, and the SNPs unique to the *G. chui* population were located in 28,754 genes. To explore the relationship among samples, phylogenetic analysis and PCA were conducted. PCA showed that *G. chui* and *G. scleracanthus* were clearly separated along the first principal component, which explained 21.07% of

genetic variation among samples (fig. 3B). To examine the relationships and divergence between *G. chui* and *G. scleracanthus* populations, we visualized pairwise genetic distances in an NJ tree. When we extracted the genetic data set ($F_{ST} > 0.2$) to construct NJ tree, we found that individuals from the *G. chui* formed one clade and individuals from the *G. scleracanthus* formed the other clade (fig. 3C, consensus tree based on 1,000 bootstrap replicates). This revealed a clear split between *G. chui* and *G. scleracanthus* which is consistent with PCA. Another NJ tree was constructed based on all the SNPs detected above (supplementary fig. S5, Supplementary Material online); however, we found less obvious differentiation, implying that there was thus overall a very weak but detectable signal of possible incipient sympatric ecological speciation. We obtained genetic diversity ($\theta\pi$) values for each individual, which were significantly higher in *G. scleracanthus* (mean $\theta\pi = 0.4318022$) than in *G. chui* (0.4260470; $P < 2.2 \times 10^{-16}$, Mann–Whitney *U* test). Furthermore, Tajima's *D* values were higher in *G. scleracanthus* (0.4705076) than in *G. chui* (0.4311897). We compared patterns of LD among *G. chui* and *G. scleracanthus* populations. LD was measured by the correlation coefficient (r^2) corresponding to distances between paired SNPs. We documented that LD decay decreased rapidly within 2,000 bp in both populations. Irrespective of unknown factors influencing background LD, we observed that the slope of all curves flattened between a distance of 2 and 5 kb. However, the slope of these curves shows less difference in both populations (fig. 3D). It is therefore likely that *G. chui* and *G. scleracanthus* populations have been ongoing incipient sympatric speciation.

Genetic Divergence and Putatively Selected Genes

We reused the SNP data to characterize the distribution of genetic differentiation (F_{ST}) for both the *G. chui* and *G. scleracanthus* population comparisons. A large part of F_{ST} was below 0.25; however, there are considerable sites with differentiation coefficients of ~0.25–0.50. The distribution of F_{ST} values has a striking mode near zero, tapering off steeply into a thin tail of higher differentiation values (a strongly "L-shaped" distribution, fig. 4C), which showed that a small proportion of loci nevertheless exhibited deep divergence in the two comparisons. Tajima's *D* test was performed on both of the *G. chui* and *G. scleracanthus* populations to assess the selection. The distribution of Tajima's *D* for both *G. chui* and *G. scleracanthus* populations is shown in figure 4B. Most of the *D* values for both of the two populations were positive, which may have been produced by a recent bottleneck or population subdivision.

To systematically define the roles that unique SNPs play in these two closely species, we initially identified SNPs located in 29,910 and 28,754 genes for *G. scleracanthus* and *G. chui*, respectively. To detect genes that are under positive
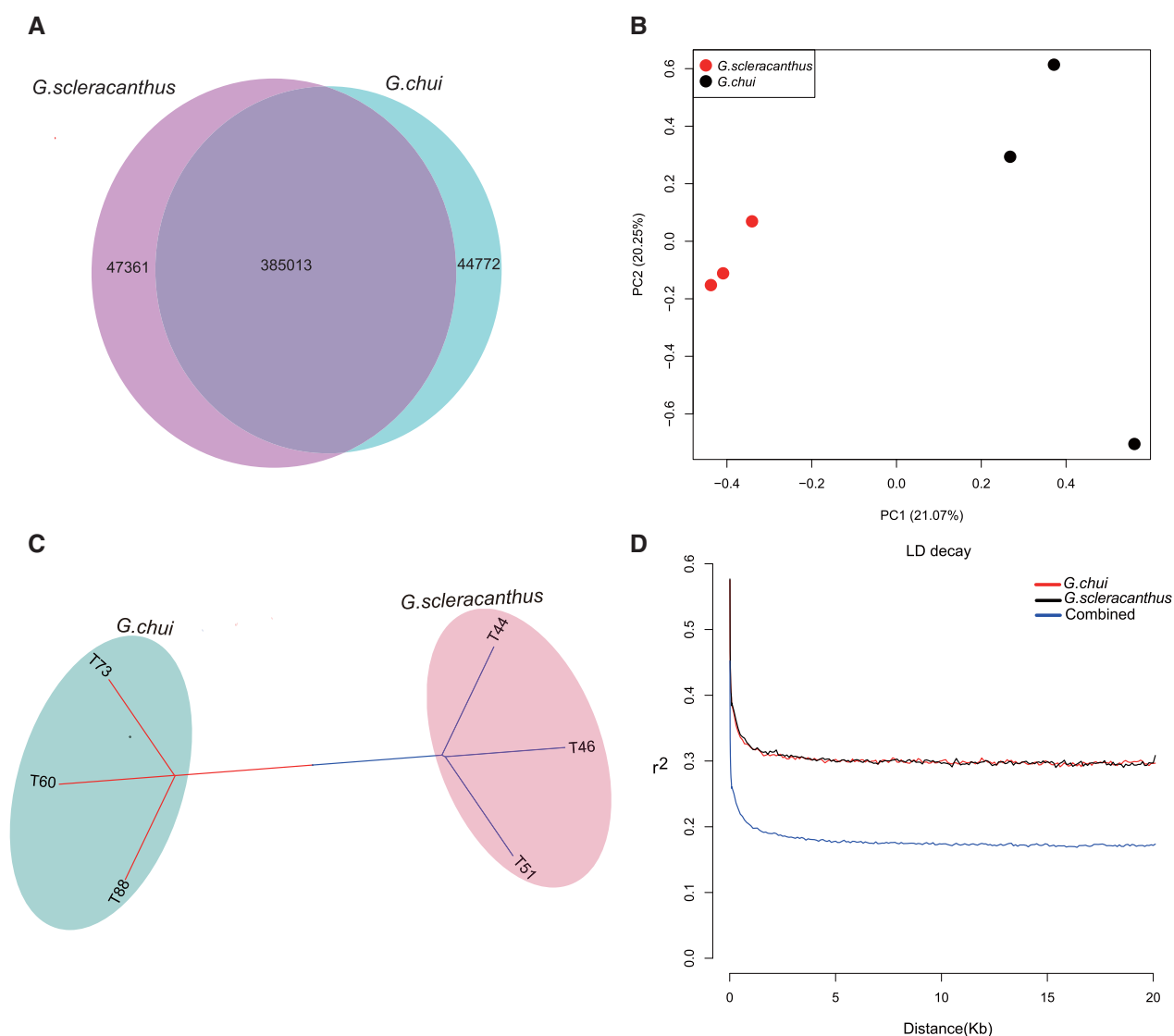
Fig. 3.—Population divergence of two fish taxa. (A) Venn diagram of SNPs unique to the *G. scleracanthus* and *G. chui* populations. (B) *Gymnocypris scleracanthus* and *G. chui* populations. Red dots, *G. scleracanthus* individuals; black dots, *G. chui* individuals. (C) A neighbor-joining phylogenetic tree constructed using SNP data ($F_{ST} > 0.2$). Red, *G. chui* individuals; blue, *G. scleracanthus* individuals. (D) LD decay by distance across the studied groups. *X*-axis stands for physical distances (kb), whereas *y*-axis stands for r^2.

selection and are important for adaptation, we screened for species-specific putatively selected genes from the overlap of the top 5% log-odds ratios of both $\theta\pi$ and F_{ST}. Transcripts under putative selection from *G. scleracanthus* and *G. chui* populations are marked in red and green dots, respectively, in figure 4A. There were 4,501 genes that were subjected to natural selection in the *G. scleracanthus* population and 4,323 putatively selected genes found in *G. chui*. To explore the function of candidate adaptive genes, we performed GO enrichment analysis. The divergence hotspots show significant ($P < 10^{-3}$) enrichment in 36 GO terms. The GO terms suggest that *G. chui* functional categories are enriched in bone morphogenesis, cell growth, neurogenetics, enzyme activity, and binding activity. Nutrition and localization

were highlighted in *G. scleracanthus* (fig. 5A and supplementary table S8, Supplementary Material online), which is consistent with the fact that they occupy different niches with different dietary preferences.

Furthermore, in order to mine the putatively selected genes during the process of sympatric speciation, we estimated these genes with the results from DE genes in edgeR. A total of 212 genes were identified with the strongest signature of positive selection (supplementary table S9, Supplementary Material online). Enriched GO terms showed that many of these genes clustered in membrane activity, neurogenetics, and ion transporters (fig. 5B). Functional enrichment networks of putative selection genes in biological process are shown in figure 5C. The results showed that
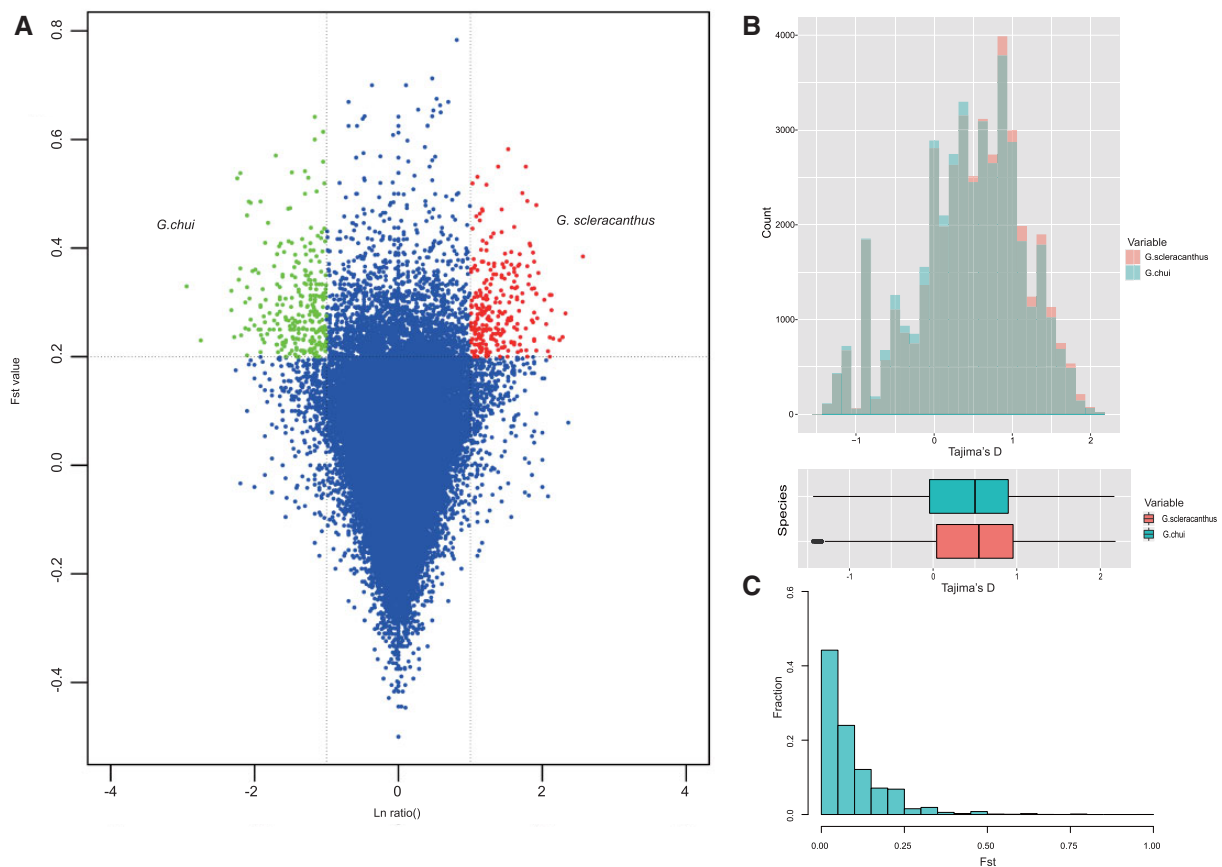
Fig. 4.—Natural selection in the two taxa of *Gymnocypris* in Lake Langcuo. (*A*) Distribution of ln ratio ($\theta\pi$_pop1/$\theta\pi$_pop2) and F_{ST} of each transcript. Red dots stand for *G. scleracanthus* (pop1), and green dots stand for *G. chui* (pop2), which represent transcript sunder putative selection (corresponding to $P < 0.05$, where $F_{ST} > 0.2$ and ln ratio > 1). (*B*) Tajima's *D* distribution of the *G. scleracanthus* and *G. chui* populations. Tajima's *D* for the *G. chui* population is smaller than that for the *G. scleracanthus* population. The *G. scleracanthus* and *G. chui* populations are marked in red and green, respective. (*C*) Distribution of the genetic differentiation (F_{ST}) at SNPs of *G. scleracanthus* and *G. chui* populations.

putatively selected genes were significantly enriched in neurogenetics cycle-related functions, including the synaptic signaling processes, the chemical synaptic transmission, and the regulation of postsynaptic membrane potential. Interestingly, of the 212 putatively selected genes during sympatric speciation, lgals9l3, pomt1, rxylt1, st3gal1, and ST6 were found to be related to carbohydrate metabolism which may be related to genetic adaptation to dietary specialization. Overall, the selected genes suggested that genes enriched in biological regulation processes, such as bone morphogenesis, neurogenetics, and vitamin transmembrane transport, could play an important role during sympatric speciation.

Discussion

The present study included comparisons of gene expression, phylogenetic analysis, population genetics (based on SNP), and ecological analysis. Here, we generated comprehensive transcriptional profiles, including multiple tissues. Our results confirmed that *G. chui* and *G. scleracanthus* form a monophyletic group with a sister species relationship based on a high-quality RNA-Seq data set covering 56 species, indicating their close relationship. This is the first step that we need to strengthen and fulfill the sympatric speciation criteria. Because fish from sympatric populations are related more closely to each other than to fish outside of their populations, this system provides an ideal situation for testing the hypothesis of ongoing sympatric speciation (Barluenga et al. 2006; Malinsky et al. 2015). Based on comparing multitissue transcriptome profiles, most organs showed unique transcriptome profiles within species. To obtain an initial overview of gene expression patterns, we performed an ordination analysis, which clearly separated the data according to tissue. This may be explained by the organs having evolved longer than the species. ANOVAs show consistently stronger effects for organs than for species (Brawand et al. 2011; Uebbing et al. 2016). Separate ordination plots were made for each organ; however, only muscle resolved fully according to species, whereas liver and kidney showed a tendency to do so. This
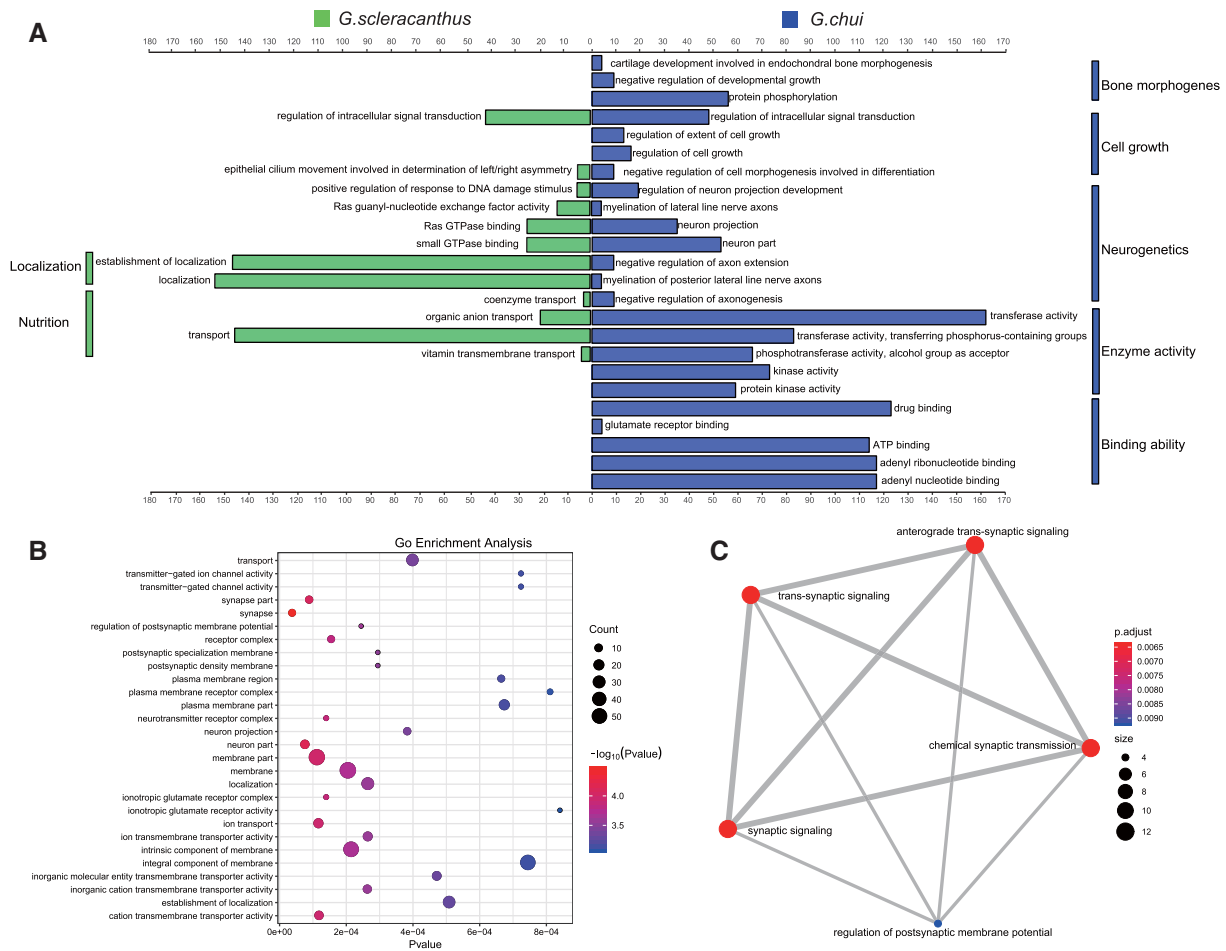
Fig. 5.—GO enrichment analysis of putatively selected genes. (A) GO enrichment analysis of putatively selected genes in *G. scleracanthus* and *G. chui* populations (based on unique SNPs). (B) GO enrichment analysis of putatively selected genes during sympatric speciation (based on the overlap of DE genes and all SNPs with $F_{ST} > 0.2$, 5% top ln ratio). (C) GO enrichment analysis of putatively selected genes focused on biological processes. Nodes represent enriched GO gene sets, whose size reflects the total number of genes in that gene.

may be explained by the morphological differences of *G. chui* and *G. scleracanthus*, especially the characteristics of the mouth cleft, which need muscle function. A similar phenomenon is seen in different soil types of the blind mole, where musculature plays a potentially broad importance in biological adaptation (Li et al. 2015). The gene expression level varied among individuals and species, and was largest in the heart. The divergence variance, however, was largest in muscle, implying that muscle expression and divergence may heavily effect on species speciation. The most divergent organs were predominantly involved in ecological adaptation, particularly the ability to process different food types, during which muscle function is needed to form more oblique and deeper mouth cleft. Previous studies in the blind mole rat with different soil types have also implicated muscle function (Hadid et al. 2013). Organs, like the brain and kidney, showed markedly lower expression divergence, and may be less affected by

the environment, similar as in birds (Uebbing et al. 2016) and primates (Khaitovich et al. 2005). Organs may differ for many other reasons than environmental impacts during species speciation, so this may be underestimated by the expression of genes (Uebbing et al. 2016).

In addition to the gene expression levels, we also explored potential mechanisms at the genetic level. We then investigated the population structure, according to the NJ tree as well as the PCA. The phylogenetic relationships were determined using the genetic distances calculated from the SNPs. The resulting NJ tree showed two divergent groups belonging to *G. scleracanthus* and *G. chui*, revealing the sister group of two sympatric schizothoracine fish (fig. 3C). PCA showed a strong population structure, with *G. chui* and *G. scleracanthus* individuals separated by the first eigenvector and forming separate clusters (fig. 3B). The genetic diversity was significantly higher in *G. scleracanthus* than in *G. chui*, which was

shown by θ_π. This finding is related to the feeding ecology and the characteristic of the mouth cleft. The densely rakered *G. chui* has longer, more densely packed rakers than the sparsely rakered *G. scleracanthus*, which leads to dietary differentiation. The densely rakered and oblique mouth cleft of *G. chui* makes more efficient for eating plankton, chironomid pupae, and insects at the water surface, whereas the diet of the sparsely rakered *G. scleracanthus* having a variety of food items includes zoobenthos such as benthic crustaceans and insect larvae, plankton or appendiculate algae. This feeding preference seems to facilitate the differentiation of niches within a single geographic area. *Gymnocypris chui* appears to be adapted to feeding mainly on the upper part of the water column, and on shoals where there is richer plankton, whereas *G. scleracanthus* has a wider niche that allows them to obtain food from different layers of the lake.

LD plots showed that LD decays with physical distance in *G. scleracanthus* on a scale similar to samples from *G. chui* populations, although somewhat less rapidly, possibly because of the more widely available food resource. The rate of LD decay appears higher in the combined than in the separate populations. It is therefore likely that LDs are diet specific for *G. chui* and *G. scleracanthus* separately rather than generally for both. The distribution of Tajima's *D* for both *G. chui* and *G. scleracanthus* populations is shown in figure 4*B*. Most of the *D* values for both of the two populations were positive, which may have been produced by a recent bottleneck or population subdivision (Schmidt and Pool 2002). Consistent with sympatric speciation, genetic divergence (F_{ST}) within the genome, followed an L-shaped distribution, with most loci showing low F_{ST} and only a small number of loci showing high levels of divergence. These loci are those most likely to be linked to genes subject to divergent selection during sympatric speciation (Savolainen et al. 2006).

Functional enrichment analysis of GO terms revealed remarkable divergence of biological differences between the sympatric schizothoracine fish populations, indicating genome-wide adaptive complexes for their morphological and dietary differentiation. GO-enriched terms were related to 1) morphogenesis (e.g., cartilage development involved in endochondral bone morphogenesis), which is consistent with morphological differentiation; 2) neurogenetics (e.g., synaptic circuitry and neurological processes), which is consistent with previous studies showing the role of neurogenetics in adaptation and speciation (Li et al. 2015); and 3) nutrition. (e.g., vitamin transmembrane transport) (fig. 5*A*). More interestingly, *G. scleracanthus* genes were particularly enriched in "transport" (GO:0006813), "organic anion transport" (GO:0015711), "vitamin transmembrane transport" (GO:0035461), and "establishment of localization" (GO:0051234) terms. The different dietary preferences elicit biological functions of nutrition. These two sympatric schizothoracine fish have totally different GO terms, which is consistent with the fact that they are exposed to different food types (Zhao et al. 2009).

Genes play pivotal roles in organismal adaptation to local environments. Expression divergence for genes and key gene functions is potentially important in sympatric speciation. In this study, several genes were identified as putatively selected genes, which are involved in carbohydrate metabolism. To date, genes coding for lactase (LCT) and the amylase (AMY1) enzymes have been identified as involved in dietary specializations in biological metabolism (Luca et al. 2010). LGALS913, orthologous to human LGALS9 (galectin-9) and LGALS9C (galectin 9C), participates in the carbohydrate sulfate metabolic pathway and monosaccharide metabolism (Suzuki et al. 2001). The gene RXYLT1 encodes a type II transmembrane protein that is thought to have glycosyltransferase function. ST3GAL1 (ST3 beta-galactoside alpha-2,3-sialyltransferase 1) is a type II membrane protein that catalyzes the transfer of sialic acid from CMP-sialic acid to galactose-containing substrates. Sialyltransferases are biosynthetic enzymes of the sialic acid metabolic pathway that mediate the transfer of sialic acid residues to terminal nonreducing positions of a variety of oligosaccharide chains found on glycoproteins and glycolipids (Petit et al. 2015). ST6GALNAC4 (ST6 [alpha-*N*-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-*N*-acetylgalactosaminide alpha-2,6-sialyltransferase 4]) is a type II membrane protein that catalyzes the transfer of sialic acid from CMP-sialic acid to galactose-containing substrates (Miao et al. 2016). POMT1 encodes protein *O*-mannosyltransferase, an enzyme that catalyzes *O*-mannosylation of proteins, and is present in many different tissues in the body but is particularly abundant in the muscles used for movement (skeletal muscles). Taken together, these selected genes suggest that divergent selection in carbohydrate metabolism attributed the adaptation of different niches with dietary differentiation, which indirectly unveils the genetic basis of sympatric speciation.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Author Contributions

S. H. and K. Z. lead the project. J. C. and L. Y. performed all computational analyses. R. Z., C. Z., and F. T. and collected the samples. S. U. provided statistical advice. J. C. wrote and revised the paper with input from the other authors.

Literature Cited

April J, Hanner RH, Dion-Cote AM, Bernatchez L. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. Mol Ecol. 22(2):409–422.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439(7077):719–723.

Berlocher SH, Feder JL. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annu Rev Entomol. 47(1):773–815.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. Annu Rev Ecol Evol Syst. 38(1):459–487.

Brawand D, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478(7369):343–348.

Cao WX, Chen YY, Wu YF, Zhu SQ. 1981. Origin and evolution of schizothoracine fishes in relation to the upheaval of the Xizang Plateau (in Chinese) In: The Comprehensive Scientific Expedition to the Qinghai-Xizang Plateau, Chinese Academy of Sciences, ed. Collection in Studies on the Period, Amplitude and Type of the Uplift of the Qinghai-Xizang Plateau. Beijing: Science Press, pp. 118–130.

Cao YB, Chen XQ, Wang S, Wang YX, Du JZ. 2008. Evolution and regulation of the downstream gene of hypoxia-inducible factor-1alpha in naked carp (Gymnocypris przewalskii) from Lake Qinghai, China. J Mol Evol. 67(5):570–580.

Cardoso AL, Sales KA, Nagamachi CY, Pieczarka JC, Noronha RC. 2013. Comparative cytogenetics of two species of genus Scobinancistrus (Siluriformes, Loricariidae, Ancistrini) from the Xingu River, Brazil. Comp Cytogenet. 7:43–51.

Chen YF, Cao WY. 2000. Schizothoracinae. In: Yue PQ, editor. Fauna Sinica, Osteichthyes, Cypriniformes III. Beijing (China): Science Press. p. 273–390.

Chen YY, Zhang W, Hwang SY. 1982. Speciation in schizothoracid fishes of Lake Lugu. Acta Zool Sin. 28:217–225.

Coyne J, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.

Darwin C. 1859. On the Origins of Species by Means of Natural Selection. London: John Murray.

Dieckmann U, Doebeli M. 1999. On the origin of species by sympatric speciation. Nature 400(6742):354–357.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10(1):48.

Friesen VL, et al. 2007. Sympatric speciation by allochrony in a seabird. Proc Natl Acad Sci U S A. 104(47):18589–18594.

Gavrilets S, Li H, Vose MD. 2000. Patterns of parapatric speciation. Evolution 54(4):1126–1134.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29(7):644–652.

Guan L, Chi W, Xiao W, Chen L, He S. 2014. Analysis of hypoxia-inducible factor alpha polyploidization reveals adaptation to Tibetan Plateau in the evolution of schizothoracine fish. BMC Evol Biol. 14(1):192.

Hadid Y, et al. 2013. Possible incipient sympatric ecological speciation in blind mole rats (Spalax). Proc Natl Acad Sci U S A. 110(7):2587–2592.

Hadid Y, et al. 2014. Sympatric incipient speciation of spiny mice Acomys at "Evolution Canyon," Israel. Proc Natl Acad Sci U S A. 111(3):1043–1048.

Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. 2013. Finding the active genes in deep RNA-seq gene expression studies. BMC Genomics 14(1):778.

He DK, Chen YF. 2006. Biogeography and molecular phylogeny of the genus Schizothorax (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. J Biogeogr. 33(8):1448–1460.

Higashi M, Takimoto G, Yamamura N. 1999. Sympatric speciation by sexual selection. Nature 402(6761):523–526.

Hubner S, et al. 2013. Genome differentiation of Drosophila melanogaster from a microclimate contrast in Evolution Canyon, Israel. Proc Natl Acad Sci U S A. 110(52):21059–21064.

Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L. 2010. The transcriptomics of sympatric dwarf and normal lake whitefish (Coregonus clupeaformis spp., Salmonidae) divergence as revealed by next-generation sequencing. Mol Ecol. 19(24):5389–5403.

Kautt AF, Machado-Schiaffino G, Meyer A. 2016. Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. PLoS Genet. 12(6):e1006157.

Kautt AF, Machado-Schiaffino G, Torres-Dowdall J, Meyer A. 2016. Incipient sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater lakes in Nicaragua due to differential use of the benthic and limnetic habitats? Ecol Evol. 6(15):5342–5357.

Khaitovich P, et al. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309(5742):1850–1854.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science 188(4184):107–116.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12(1):323.

Li BY. 2000. The last greatest lakes on the Xizang (Tibetan) Plateau. Acta Geogr Sin. 55(2):174–182.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.

Li H, et al.; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li K, et al. 2015. Sympatric speciation revealed by genome-wide divergence in the blind mole rat Spalax. Proc Natl Acad Sci U S A. 112(38):11905–11910.

Li KX, et al. 2016. Sympatric speciation of spiny mice, Acomys, unfolded transcriptomically at Evolution Canyon, Israel. Proc Natl Acad Sci U S A. 113(29):8254–8259.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22(13):1658–1659.

Li YL, et al. 2013. High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. Gene 517(2):169–178.

Luca F, Perry GH, Di Rienzo A. 2010. Evolutionary adaptations to dietary changes. Annu Rev Nutr. 30(1):291–314.

Malinsky M, et al. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350(6267):1493–1498.

Martin M. 2001.Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:3.

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.

Miao X, et al. 2016. miR-4299 mediates the invasive properties and tumorigenicity of human follicular thyroid carcinoma by targeting ST6GALNAC4. IUBMB Life 68(2):136–144.

Mirza MR. 1991. A contribution to the systematics of the schizothoracine fishes (Pisces: Cyprinidae) with the description of three new tribes. Pak J Zool. 23:339–341.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A. 76(10):5269–5273.

Nevo E. 2014. Evolution in action: adaptation and incipient sympatric speciation with gene flow across life at "Evolution Canyon", Israel. Israel J Ecol Evol. 60:85–98.

Odeen A, Florin AB. 2002. Sexual selection and peripatric speciation: the Kaneshiro model revisited. J Evol Biol. 15:301–306.

Petit D, et al. 2015. Integrative view of alpha2,3-sialyltransferases (ST3Gal) molecular and functional evolution in deuterostomes: significance of lineage-specific losses. Mol Biol Evol. 32(4):906–927.

Qi DL, et al. 2018. Transcriptome analysis provides insights into the adaptive responses to hypoxia of a schizothoracine fish (Gymnocypris eckloni). Front Physiol. 9:1326.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140.

Rundle HD, Nagel L, Wenrick Boughman J, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287(5451):306–308.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406–425.

Savolainen V, et al. 2006. Sympatric speciation in palms on an oceanic island. Nature 441(7090):210–213.

Schmidt D, Pool J. 2002. The effect of population history on the distribution of the Tajima's D statistic. http://www.cam.cornell.edu~deena/TajimasD.pdf.

Sharaf K, Hadid Y, Pavlicek T, Eviatar N. 2013. Local genetic population divergence in a saw-toothed grain beetle, Oryzaephilus surinamensis (L.) (Coleoptera, Cucujidae). J Stored Prod Res. 53:72–76.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Suzuki A, et al. 2001. Molecular cloning and expression of a novel human beta-Gal-3-O-sulfotransferase that acts preferentially on N-acetyllactosamine in N- and O-glycans. J Biol Chem. 276(26):24388–24395.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595.

Tang Y, et al. 2019. Convergent evolution misled taxonomy in schizothoracine fishes (Cypriniformes: Cyprinidae). Mol Phylogenet Evol. 134:323–337.

Tregenza T, Butlin RK. 1999. Speciation without isolation. Nature 400(6742):311–312.

Uebbing S, et al. 2016. Divergence in gene expression within and between two closely related flycatcher species. Mol Ecol. 25(9):2015–2028.

Venables WN, Ripley BD. 2002. Modern applied statistics with S. New York: Springer Science+Business Media.

Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol Evol. 16(7):381–390.

Vilella AJ, et al. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19(2):327–335.

Wang S, Dou H. 1998. China lakes record. Beijing (China): Science Press Ltd.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Wolf JB, Lindell J, Backstrom N. 2010. Speciation genetics: current status and evolving approaches. Philos Trans R Soc B. 365(1547):1717–1733.

Wu YF, Wu CZ. 1992. The fishes of the Qinghai-Xizang plateau. Chengdu (China): Sichuan Publishing House of Science & Technology.

Xia M, et al. 2016. Changes of hemoglobin expression in response to hypoxia in a Tibetan schizothoracine fish, Schizopygopsis pylzovi. J Comp Physiol B. 186(8):1033–1043.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Yonezawa T, Hasegawa M, Zhong Y. 2014. Polyphyletic origins of schizothoracine fish (Cyprinidae, Osteichthyes) and adaptive evolution in their mitochondrial genomes. Genes Genet Syst. 89:187–191.

Yun-Fei WU. 1991. Characteristics of the fish-fauna of the characteristics of Qinghai-Xlzang Plateau and its geological distribution and formation. Acta Zool Sin. 37(2):135–152.

Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35(10):1786–1788.

Zhang CG, Cai B, Xu TQ. 1995. Fishes and fish resources in Xizang, China. Beijing (China): China Agriculture Press.

Zhang J, Chen Z, Zhou CJ, Kong XH. 2016. Molecular phylogeny of the subfamily Schizothoracinae (Teleostei: Cypriniformes: Cyprinidae) inferred from complete mitochondrial genomes. Biochem Syst Ecol. 64:6–13.

Zhang RY, Tang YT, Feng CG, Liu SJ, Zhao K. 2016. Comparative cytogenetics of two sympatric species (Gymnocypris, Schizothoracinae) from the Lake Langcuo, Qinghai-Tibet Plateau. Caryologia 69(3):196–200.

Zhang W, Sun Z. 2008. Random local neighbor joining: a new method for reconstructing phylogenetic trees. Mol Phylogenet Evol. 47(1):117–128.

Zhao K, et al. 2009. The youngest split in sympatric schizothoracine fish (Cyprinidae) is shaped by ecological adaptations in a Tibetan Plateau glacier lake. Mol Ecol. 18(17):3616–3628.

Zheng BR, Zhang YP, Xiao CJ, Xiao H, Zan RG. 2004. The genetic evidence for sympatric speciation pattern of Cyprinus from Erhai Lake. Yi Chuan Xue Bao 31(9):976–982.

Associate editor: Maria Costantini