Mapping active promoters by ChIP-seq profiling of H3K4me3 in cichlid fish – a first step to uncover cis-regulatory elements in ecological model teleosts

CLAUDIUS F. KRATOCHWIL*† and AXEL MEYER*

*Chair in Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Universitaetsstrasse 10, 78457 Konstanz, Germany, †Zukunftskolleg, University of Konstanz, Universitaetsstrasse 10, 78457 Konstanz, Germany

Abstract

Evolutionary alterations to cis-regulatory sequences are likely to cause adaptive phenotypic complexity, through orchestrating changes in cellular proliferation, identity and communication. For nonmodel organisms with adaptive key innovations, patterns of regulatory evolution have been predominantly limited to targeted sequence-based analyses. Chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) is a technology that has been primarily used in genetic model systems and is a powerful experimental tool to screen for active cis-regulatory elements. Here, we show that it can also be used in ecological model systems and permits genomewide functional exploration of cis-regulatory elements. As a proof of concept, we use ChIP-seq technology in adult fin tissue of the cichlid fish *Oreochromis niloticus* to map active promoter elements, as indicated by occupancy of trimethylated Histone H3 Lysine 4 (H3K4me3). The fact that cichlids are one of the most phenotypically diverse and species-rich families of vertebrates could make them a perfect model system for the further in-depth analysis of the evolution of transcriptional regulation.

Keywords: cichlids, cis-regulatory elements, epigenetics, evolutionary genetics, histone modifications, phenotypic variation, tilapia, transcriptional regulation

More than 1500 cichlid species evolved in the three Great East African Lakes forming huge adaptive radiations within less than a few million years (Meyer *et al.* 1990; Meyer 1993; Stiassny & Meyer 1999; Kocher 2004; Seehausen 2006). Cichlids are famous for their astonishing rate of phenotypic diversification, making them excellent models for the investigation of the evolutionary role of coding regions but also noncoding elements (Salzburger & Meyer 2004; Kuraku & Meyer 2008; Fan *et al.* 2012; Brawand *et al.* 2014; Henning & Meyer 2014). Their extremely fast rates of speciation and diversification are still puzzling, as they are the only family of fishes in the East African Lakes that has diversified to such an extent. Hence, it has been suggested that especially the evolution of noncoding elements might be one of the driving factors behind the rapid and sustained rate of speciation in the cichlid lineage (Baldo *et al.* 2011).

Recently, many new next-generation DNA-sequencing technologies have allowed for the rapid accumulation of genetic maps (Sanetra *et al.* 2009; Recknagel *et al.* 2013), genome scans (Mattersdorfer *et al.* 2012) and QTL

analyses (Franchini *et al.* 2014); transcriptomes and even genomes (Brawand *et al.* 2014) in cichlid fishes. However, what has been lacking so far are methods for experimental search for cis-regulatory elements. Cis-regulatory regions regulate the transcription of a gene. They can be divided into promoters that are directly 5′ of the first exon in close proximity the transcription start site (TSS) as well as enhancers that amplify transcription and can be located in the distal part of the promoter or up to kilobases or even megabases away from the regulated gene (long-range enhancers) (Wray 2007; Wittkopp & Kalay 2012; Kratochwil & Meyer 2014).

Chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) (Park 2009; Furey 2012) has been used in various mammalian systems to investigate the dynamics of transcription factor binding (Schmidt *et al.* 2010; Stefflova *et al.* 2013), to map transcripts (Cheung *et al.* 2010), to screen for conserved regulatory regions (Tena *et al.* 2014) and to analyse the activity of tissue-specific regulatory elements (Visel *et al.* 2009; Blow *et al.* 2010; Attanasio *et al.* 2013). Increased occupancy of specific histone modifications such as lysine residue methylation and acetylation was found to be linked to active regulatory elements, and active or

Correspondence: Axel Meyer, Fax: +49 7531 883018; E mail: Axel.Meyer@uni konstanz.de

repressed promoters (Barski *et al.* 2007; Heintzman *et al.* 2007; Zhou *et al.* 2011). Trimethylated Histone H3 Lysine 4 (H3K4me3) has been convincingly shown to broadly target active promoters (Mikkelsen *et al.* 2007) and to be associated with transcription throughout cellular differentiation (Guenther *et al.* 2007).

In this study, we sought to examine genomewide patterns of promoter-associated histone mark H3K4me3 occupancy in the fin of Nile tilapia, a species that is basal to the East African radiations of cichlid fishes (Fig. 1). Nile tilapia diverged around 20 45 million years ago, before the East African lake radiations formed, and lacks many of the complex body and fin coloration traits such as egg-spots that characterize haplochromines, the most species-rich lineage of cichlids that comprise the entire species flocks of both Lakes Victoria (with >500 species) and Malawi (800 1000 species) (Hert 1989; Van Alphen 1999; Salzburger & Meyer 2004; Schwarzer *et al.* 2009; Maan & Sefc 2013), so they provide a good outgroup for further investigating the evolution of phenotypically diverse traits. Importantly, Nile tilapia has the best genome (anchored and well annotated) so far of all cichlid species (N50: 29.3 kb), due to dual interests in it as the most important tropical aquaculture species (Gupta & Acosta 2004) and to its evolutionary position (Schwarzer *et al.* 2009). In particular, because of the high quality of the annotations, the Nile tilapia was the best species for a proof of concept of this methodology, as in other cichlid species less transcripts are annotated up to now.

Although mutations and translocation of enhancers have been considered as the main driver of regulatory evolution (Wittkopp & Kalay 2012), also promoter mutations and the recruitment of alternative promoters have been implicated in phenotypic evolution as well as human disease (Landry *et al.* 2003; Carninci *et al.* 2006; Kowalczyk *et al.* 2012; Lenhard *et al.* 2012; Ruiz-Narváez 2013).As, in contrast to enhancers, the position of TSSs is at least to a large extent known for tilapia, we decided to

use the promoter-associated histone mark H3K4me3. ChIP-seq for H3K4me3 is commonly used as a positive control for ChIP-seq experiments in model organism for example, if complicated tissues or new antibodies are tested as functionality can be tested by correlating ChIP-seq read alignment position to TSS position (Schmidt *et al.* 2009). Hence, H3K4me3 was logical choice for a proof of concept of the ChIP-seq technology in a nonmodel teleost.

For ChIP-seq, similarly as for RNA-seq, the choice of tissue is of central importance, especially if they are performed in a comparative framework. Criteria that were considered for the choice of tissue were (i) the diversity of the tissue during the course of cichlid evolution, (ii) the likelihood that a larger fraction of cis-regulatory elements are active, (iii) the possibility that the tissue can be dissected in a quick and reproducible manner, in the best case even without killing the animal, (iv) that the structure is not greatly influenced by phenotypic plasticity and (v) that heterochrony is unlikely to be a major factor explaining the observed diversity.

Based on these five criteria, we picked the anal fin as tissue of choice. Fins are known to exhibit diverse shapes and coloration in many cichlids and surely play an important role in cichlid diversification, in terms of both sexual selection and natural selection (Van Alphen 1999; Salzburger *et al.* 2007; Seehausen *et al.* 2008; Schwarzer *et al.* 2009; Henning & Meyer 2012; Henning *et al.* 2013). For the anal fin, this includes especially the presence of egg spots in the more derived haplochromine cichlids. Egg spots are egg-shaped, brightly coloured circles that are presented to the female during courtship behaviour (Couldridge 2002; Salzburger *et al.* 2007; Egger *et al.* 2011; Maan & Sefc 2013).

Historically, many ChIP-seq experiments have been performed in developing embryos due to the fact that many enhancers are especially active at developmental stages (Visel *et al.* 2009; Blow *et al.* 2010; Attanasio *et al.* 2013). Also in teleosts, recently, histone mark occupancy
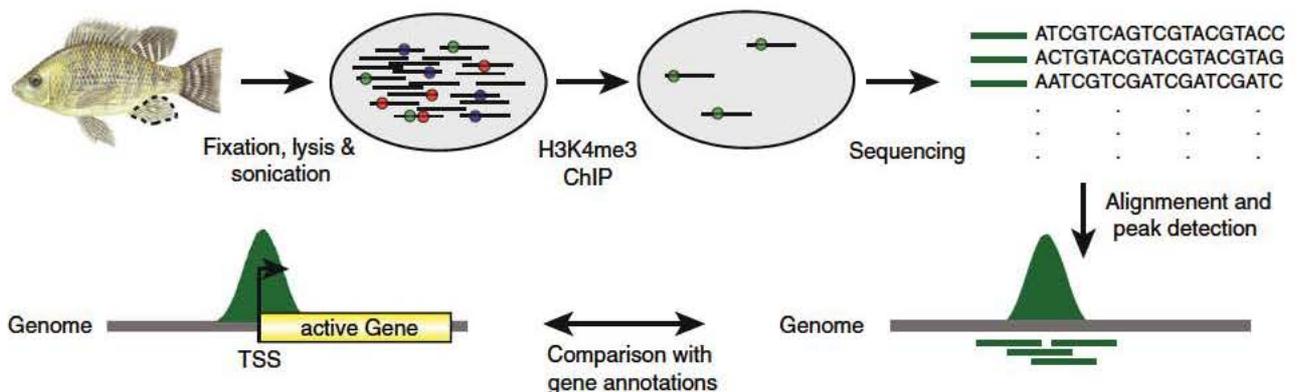

Fig. 1 Experimental overview. Anal Fin tissue of Nile tilapia (*Oreochromis niloticus*) was dissected, fixed, lysed and sonicated, followed by the ChIP seq for H3K4me3. Reads were aligned to the tilapia genome Orenil1.1, and peaks were detected.

has been compared between Medaka and zebrafish giving insight into conserved regulatory modules (Tena *et al.* 2014). However, performing ChIP-seq in a comparative manner between species would not be practicable in embryos. Experimental outcomes would be largely affected by the embryonic stage when embryos are taken, which are barely described in cichlids. Additionally, many phenotypic diverse traits such as jaws might be consequences of evolutionary changes in developmental timing, also called heterochrony (Gunter *et al.* 2014; Keyte & Smith 2014). Furthermore, the spatial resolution would have been reduced dramatically as whole embryos or bigger parts of the embryos would have been necessary to gather enough chromatin for a ChIP-seq experiment.

Fin tissues are highly regenerative structures (Katogi *et al.* 2004), and many of the structural and phenotypic properties have to be actively maintained (van Eeden *et al.* 1996). In particular, for cichlids, it is known that hormone levels drastically affect coloration in body and fins (Fernald 1976; Oliveira & Canário 2000; Parikh *et al.* 2006) and can even induce male coloration in females if applied for a few days (Wapler-Leong & Reinboth 1974; Oliveira & Canário 2000). Both regenerative potential of the tissue and inducibility by hormones led us assume that many promoters are still in an active or poised state.

Additionally, the use of fin tissue has further advantages. Compared to other tissues such as jaws, fins are not as dramatically affected by phenotypic plasticity (Gunter *et al.* 2013). Hormone levels are dependent on the social status, but can be kept stable by keeping animals alone (Grosenick *et al.* 2007; Fernald 2012). An additional advantage of fins is that the tissue can be dissected quickly and reproducibly, potentially even under field conditions. As the cutting of the fin does not affect survival, fishes can be theoretically even kept for further experiments or crossings.

This study provides an excellent proof of concept and forms the baseline for larger studies of the role of regulatory landscapes for phenotypic diversification of cichlids and other teleosts. Furthermore, we think that ChIP-seq technology can be easily adapted for other sequenced nonmodel teleosts such as sticklebacks, poeciliids and Mexican tetra (blind cave fish, *Astyanax mexicanus*) and that ChIP-seq is truly an opportunity to add depth to our understanding of the genetic and epigenetic factors influencing phenotypic variation.

Materials and methods

Chromatin immunoprecipitation

Tilapia fin tissue from laboratory stocks (kept at the Animal Research Facility at the University of Kon-

stanz, Germany) was dissected, immediately frozen on dry ice and stored at 80 °C. ChIP-seq was performed based on Active Motif's HistonePath method (Active Motif) as previously described (Ramagopalan *et al.* 2010). The frozen tissue was crushed in liquid nitrogen and fixed in PBS + 1% formaldehyde at room temperature for 10 min. Fixation was stopped by the addition of 0.125 M glycine (final concentration), and the fin pieces were washed 2× in PBS and 1× in deoxycholate sonication buffer. Lysates were sonicated with a microtip in order to shear the DNA to an average length of 300 500 bp. Lysates were cleared by centrifugation and stored at 80 °C. Genomic DNA (input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-cross-linking, followed by phenol chloroform extraction and ethanol precipitation. Purified DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield. See Fig. 1 for diagrammatic presentation of the protocol.

For each ChIP reaction, 15 µg of chromatin was precleared with protein A agarose beads (Invitrogen). ChIP reactions were set up using precleared chromatin and antibody H3K4me3 (Active Motif, cat. 39159, Lot# 4) and incubated overnight at 4 °C. Protein A agarose beads were added, and incubation at 4 °C was continued for another 3 h. Immune complexes were washed, eluted from the beads with SDS buffer and subjected to RNase treatment and proteinase K treatment. Cross-links were reversed by incubation overnight at 65 °C, and ChIP DNA was purified by phenol chloroform extraction and ethanol precipitation.

ChIP sequencing

ChIP DNA was prepared for amplification by converting overhangs into phosphorylated blunt ends and the addition of an adenine to the 3′-ends. Illumina genomic adapters were ligated, and the sample was size-fractionated (200 300 bp) on an agarose gel. After a final PCR amplification step (18 cycles), the resulting DNA libraries were quantified and sequenced on Illumina MiSeq. Sequences (50-nt reads, single end).

Of 6 367 693 reads were obtained with a GC content of 45.91%. 86.17% of the sequences had an average PHRED score between 38 and 36; 98.34% of the sequences had an average PHRED score ≥30. Reads were aligned to the anchored *Oreochromis niloticus* genome Orenil1.1, (GenBank, BioProject Identifier: PRJNA59571). Read mapping was performed using CLC Genomics Workbench 7.0.3 using standard setting of the 'Map Reads to Reference' Package (no masking, mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5,

similarity fraction 0.8). 0.13% of the reference genome was covered by the reads. Chip-Seq peaks were detected using the 'ChIP-Seq Analysis' package with standard settings (Window Size 250, maximum false discovery rate (FDR) 5%, shift reads based on fragment length 250, with and without boundary refinement, filter peaks on probability of identical locations on forward and reverse reads (Wilcoxon *P*-value 0.0001) (Table S1, Supporting information).

To obtain a version of the genome annotated directly with gene names, all predicted peptides from *Oreochromis niloticus* (GCF 000188235.2) were blasted against the NR blast database from NCBI using blastp (E-value <0.001), and annotations were replaced by the top blast hit (original annotation in parentheses e.g. LG1.123). Annotated genes without hits were not excluded from further analysis, except for gene ontology analysis. Kfl5 and the two copies of Kfl6 in Table 1 were confirmed to be Klf5b (ENSONIGP00000008543), Klf6a (ENSONIGP00000004586) and Klf6b (ENSONIP00000018824) using BLASTN at Ensembl genome browser.

Cluster analysis

Reads and reference coordinates (position of 31220 transcription starts sites, assumed to be the 5′-end of the annotated transcript, of all annotated genes of all linkage groups and the 100 largest unknown scaffolds (UNK1-100) of the *Oreochromis niloticus* genome Orenil1.1) were imported into seqMINER 1.3.3 (Ye *et al.* 2011). Windows of negative strand transcripts were inverted so that they have the same orientation as positive strand transcripts. Clustering normalization was performed using *k*-means raw, setting the expected number of clusters after at three (Fig. 2, Table S2, Supporting information) or five (Fig. S1, Supporting information). We chose a number of three clusters because we expected genes to fall in a cluster with very low/no H3K4me3 occupancy, a cluster with high H3K4me3 occupancy close to the predicted TSSs and a cluster with high H3K4me3 occupancy at a position more distal of the predicted TSSs, which was confirmed by the data. Choosing five clusters for the subsequent analysis (Fig. S1, Supporting information) was empirically determined, as the use of a higher number of clusters did in our opinion not provide further information. The discrepancy between the TSSs that overlap with H3K4me3 reads (27%) and the low number of peaks identified is due to the fact that the peak calling algorithm is more stringent. It reduces the chance of false discoveries and filters based on forward and reverse read ratio, but does not use information about TSS position. Therefore, it also allows peak detection outside of the

Table 1 List of genes with implications for fin development and regeneration with enrichment of H3K4me3. Linkage groups are noted in parentheses after the gene name

| Gene	Accession	FDR (%)	*P* value	Process involved	Citation
Beta catenin like protein 1, ctnnb1 (LG20.600)	XP 003449172.1	6.98306E 25	<0.0001	Fin regeneration	Katogi *et al.* 2004
Distal less homeobox protein 3a, dlx3a (LG4.390)	XP 005468638.1	4.51325E 54	0.007	Fin regeneration	Katogi *et al.* 2004
Distal less homeobox protein 3b, dlx3b (LG8 24.275)	XP 003438791.1	2.86119E 43	0.0455	Fin regeneration	Katogi *et al.* 2004
Probable E3 ubiquitin protein ligase, dtx2 (LG10.438)	XP 003453484.1	4.29823E 71	<0.0001	Fin development	Offen *et al.* 2009
Fras1 related extracellular matrix protein 2, frem2 (LG14.594)	XP 003450227.1	4.16807E 11	0.002	Structural maintainance	van Eeden *et al.* 1996
DnaJ homolog subfamily B member 1, hsp40 (LG6.682)	XP 003450019.1	2.86119E 43	0.0006	Fin regeneration	Tawk *et al.* 2000
Interleukin 10 receptor subunit beta like, il10rb (LG16 21.671)	XP 005450418.1	5.81452E 47	0.0097	Fin regeneration	Tawk *et al.* 2000
Integrin alpha 3, itga3/hsc70 (LG8 24.276)	XP 005448228.1	2.75861E 20	0.0054	Fin regeneration	Katogi *et al.* 2004
Krueppel like factor 5b, klf5b (LG16 21.251)	XP 005475502.1	1.83312E 37	0.0165	Fin regeneration	Katogi *et al.* 2004
Krueppel like factor 6a, klf6a (LG18.748)	XP 005476862.1	2.10469E 59	0.002	Fin regeneration	Katogi *et al.* 2004
Krueppel like factor 6b, klf6b (LG9.51)	XP 003450279.1	6.36587E 45	0.0442	Structural maintainance	van Eeden *et al.* 1996
Retinoic acid receptor gamma, rar γ (LG20.704)	XP 008286513.1	4.70689E 39	0.001	Fin regeneration	Katogi *et al.* 2004
Secreted frizzled related protein 1, sfrp1 (LG12.71)	XP 003448656.1	4.4756E 34	<0.0001	Fin regeneration	Katogi *et al.* 2004
Suppressor of fused homolog, sufu (LG13.468)	XP 003441070.2	1.66371E 11	0.0228	Fin regeneration	Katogi *et al.* 2004
Mitogen activated protein kinase kinase kinase 7, tak1 (LG19.807)	XP 005477768.1	8.79323E 31	<0.0001	Fin regeneration	Katogi *et al.* 2004
E3 ubiquitin protein ligase trim33 (LG20.608)	XP 003449176.2	6.99453E 52	0.0459	Fin regeneration	Katogi *et al.* 2004
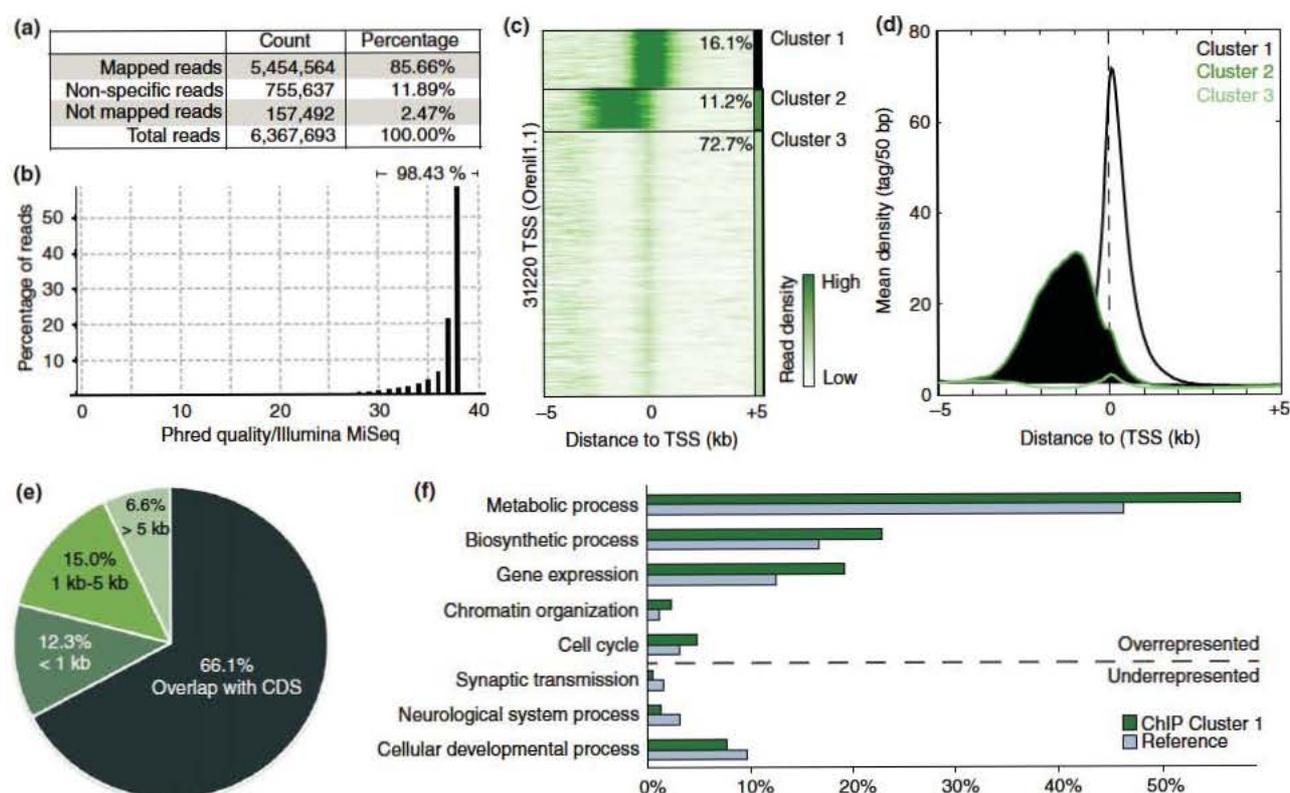
Fig. 2 Quality assessment and genomewide analysis of H3K4me3 ChIP seq. (a b). Summary of read mapping (a) and read quality (b). (c d). Cluster analysis of mapped read position in relation to the annotated transcription start sites (TSS) of tilapia genome Orenil1.1 as heatmap (c) and graph (d). (e). Distribution of the detected ChIP seq peaks. (f). Over and under represented gene ontology (GO) terms in Cluster 1 compared to reference (all annotated genes). Genes involved in developmental or nervous system processes are under rep resented, those controlling metabolic processes, gene expression, chromatin organization and cell cycle are over represented (P < 0.001).

10 kb window around the TSSs. Cluster analysis is less stringent as only the number of aligned reads is put into account, but it uses the TSS information and thereby gives a clear good readout about the quality of the ChIP-seq using the relative read density at TSSs as a proxy. High read-density at a TSS is unlikely to occur by chance and H3K4me3 is a hallmark of active promoters, suggesting that clusters 1 and 2 are composed of genes with active promoters (Guenther *et al.* 2007; Mikkelsen *et al.* 2007).

Gene ontology analysis

Gene ontology (GO) analysis was performed using BLAST2GO V.2.7.1 (Conesa *et al.* 2005). Genes of subsets (transcripts of Cluster 1) and references (all tilapia transcripts) were mapped and annotated (E-Value-Hit-Filter: 1.0E-6; annotation cut-off: 55, GO Weight: 5, HSP-Hit Coverage cut-off: 0). Enrichment analysis was performed using Fisher's exact test with Tern Filter Mode set to FDR <0.05 and a P-value <0.001.

Results

After H3K4me3 antibody precipitation of the chromatin and subsequent Illumina sequencing (see Fig. 1 and Materials and methods), the resulting reads were mapped onto the Nile tilapia genome. 5 454 564 (85.66%) of the 6 367 693 obtained reads could be mapped unambiguously to the genome. While nonspecific matches were ignored intentionally (755.637 reads, 11.89%), only 157.492 reads did not map at all (2.47%) (Fig. 2a). The quality of the reads was very high, with 98.34% of the sequences having an average PHRED score ≥30 (Fig. 2b).

To first determine whether the aligned ChIP-seq read mappings were selectively enriched at TSSs, as expect for the H3K4me3 histone modification, cluster analysis (number of clusters: 3) was used to evaluate the distribution of aligned reads in a 10-kb window around the TSSs of all annotated genes (Fig. 2c,d). In Cluster 1, a clear peak around the TSS of the genes was obtained. The peak had a slight 3′ shift (<100 bp) that is comparable to ChIP-seq results from mice (Ye *et al.* 2011). Cluster 1

encompassed 16.1% of all TSSs. Cluster 2 (11.2% of the TSSs) also has a clear but broader peak, suggesting higher peak position heterogeneity relative to the TSS. The peak is shifted by ~500 bp to a 5′ direction (Figs 2d, S1 and S2, Supporting information). Shift and heterogeneity are likely to be explained by the lack of 5′-UTR annotations in the genome data, as well as missing annotations of alternative TSSs. Cluster 3 included the highest proportion transcripts, which had only a very slight peak around the TSS. Based on these data, it would be suggested that 27.3% of the promoters are either active or poised (clusters 1 and 2), while the 72.7% of the promoter loci are inactive or weak (Cluster 3) (Figs 2c, S1 and S2, Supporting information).

In an effort to further investigate the heterogeneity of Cluster 2 and to be able to differentiate more clearly between weak and inactive promoters in Cluster 3, both were broken down further by cluster analysis (Fig. S1, Supporting information). In Cluster 2, the average peak position of the subclusters was located between 0.6 (Cluster 2.1) and 2.7 kb (Cluster 2.5) to the 5′ direction of the promoter. In Cluster 3, the majority of peaks (78.58% of Cluster 3; 57.13% of all TSSs) had no clear peak (Cluster 3.5), while the rest had low peaks (Cluster 3.2; 8.38% of all TSSs) or more distal peaks 5′ (Cluster 3.3 and 3.4; 5.47% of all TSSs) or 3′ of the TSS (Cluster 3.1; 1.74% of all TSSs) (Fig. S1a, Supporting information). To better illustrate the character of the clusters, representative examples for clusters 1, 2.2, 3.2 and 3.5 are shown as coverage plots for the same window (±5 kb around TSS) (Fig. S2, Supporting information).

The association between H3K4me3 occupancy and promoter position was further investigated by the analysis of the location of all ChIP-seq peaks according to their distance to all genes. Using stringent peak detection algorithms (Wilcoxon P-value <0.0001; CLC Genomics Workbench ChIP-seq Analysis package), 1677 different peaks were identified. 66.1% of them were overlapping with genes, and 27.3% were within a 5 kb window, 5′ direction to the TSS. Only 6.6% of the peaks had a distance of more than 5 kb away from the TSSs of annotated genes (Fig. 2e). To determine whether it is possible to identify nonannotated genes in close proximity to this 6.6% subset of peaks, we screened for open reading frames which would result in a peptide with more than 50 amino acids within a 10-kb region around a highly significant peak on LG17 (P-value: 4.76×10^{-5}) (Fig. S3, Supporting information). Using this approach, we identified 2 putative peptides, one of which showed a BLAST match to a known gene (PPWP domain-containing protein 2A-like), suggesting that this approach can indeed help to discover nonannotated transcripts (Fig. S3, Supporting information).

As a further line of evidence to assess the quality of the Chip-Seq results, we tested by gene ontology analysis whether genes involved in processes expected to be enriched in fin tissue such as metabolic processes or biosynthetic processes are over-represented in the subset of genes with H3K4me3-enriched promoters (Cluster 1), while genes involved in, for example, nervous system-related processes are under-represented. Indeed, we confirmed that genes involved in metabolic and biosynthetic processes but also cell cycle, gene expression and chromatin organization are significantly enriched (P-value <0.001; FDR<0.05%; Fig. 2f, Table S3, Supporting information), while genes involved in processes as synaptic transmission, neurological system processes and cellular developmental processes were significantly under-represented (P-value <0.001; FDR<0.05%; Fig. 2f, Table S3, Supporting information).

To examine whether genes with putatively active promoters include genes that are of importance for fin-specific cellular processes, a review of recent literature was performed to find genes involved in the development (Offen et al. 2009), regeneration (Tawk et al. 2000; Katogi et al. 2004) and structural maintenance of fins (van Eeden et al. 1996) (Table 1, Fig. S4, Supporting information). Many of these genes coding for transcription factors such as dlx3a, dlx3b, klf5b, klf6a and klf6b, genes involved in signalling pathways of Shh (sufu), retinoic acid (rar-γ), Wnts (ctnnb1, sfrp1), Bmps (trim33), Notch (dtx2) and involved in maintaining tissue integrity (frem2, tak1) exhibited high H3K4me3 occupancy (Table 1). The coverage plots of three of these genes (dlx3b, klf5b and rar-γ), including their transcript annotations, are shown in S4. Rar-γ showed two annotated TSSs, of which only the second TSS possesses a H3K4me3 peak (Fig. S4, Supporting information), suggesting that the other transcript might not be expressed in this context.

Regulation of hox genes is a classical model to study the functions of chromatin, chromatin marks such as H3K4me3 and their modifiers (Soshnikova & Duboule 2009; Soshnikova 2014). The transcriptional outputs of genes in hox clusters follow a collinear distribution, meaning that they are expressed along the anterior posterior body axis in a sequence corresponding to their respective positions on the chromosome. While genes situated at one end of the cluster are transcribed more anteriorly, genes located at the opposite end are progressively expressed more posteriorly. As the in vivo dynamics of chromatin marks can be well visualized at the spatially compartmentalized hox gene clusters, we wanted to analyse whether colinearity can be observed at the level of H3K4me3 occupancy in Nile tilapia anal fin tissue.

Sharp TSS-associated peaks could be detected in all hox clusters except hoxBb and hoxDb (Fig. 3). Strikingly, according to the very posterior position of the anal fin along the body axis, peaks were found mainly in posterior genes, especially abdominal-B-type hox paralog

groups 9 13, particularly in the *hoxAa, hoxAb, hoxCa* and *hoxDa* clusters (Fig. 3). In the *hoxBa* cluster, peaks extended also to more anterior genes including *hoxB4a, hoxB5a, hoxB6a, hoxB7a, hoxB8a* and an unknown gene that inserted between hoxB6a and hoxB7a (Fig. 3). Hox genes that are very specific for anterior body regions as the paralog groups 1 and 2 completely lacked H3K4me3 peaks.

Discussion

We present a protocol for conducting H3K4me ChIP-seq on a nonmodel teleost, opening up a powerful opportunity to examine the role of alterations to genomic regulatory landscapes that probably underlie the most compelling evolutionary novelties (Prud'homme *et al.* 2006; Wray 2007; Levine 2010; Lowe *et al.* 2011). This was accomplished on a species for which only a draft version of the genome is available. The efficacy of our protocol is demonstrated by several lines of evidence: (i) 27.3% of the annotated TSS had an enrichment of reads in close proximity H3K4me3 (Fig. 2b); (ii) 93.4% of the peaks were in close proximity (<5 kb) to annotated genes (Fig. 2e); (iii) gene ontology analysis shows an over-representation of processes expected to be enriched in fin tissue and an under-representation of processes not to be enriched in fin tissue (Fig. 2f); and (iv) many genes could
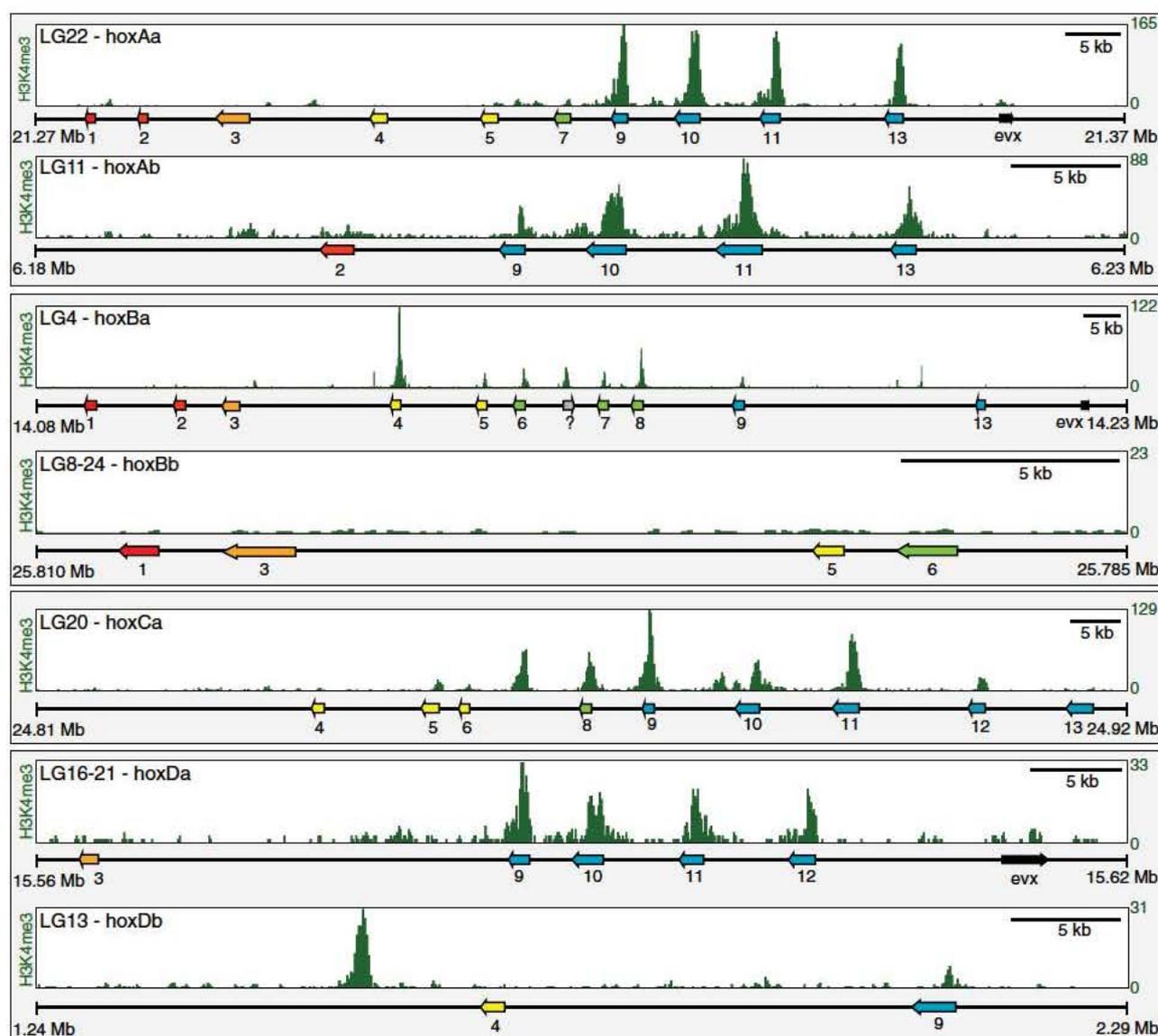

Fig. 3 Coverage plots of H3K4me3 occupancy in the seven hox clusters of the Nile tilapia. Clear peaks can be detected throughout the hox clusters, with exception of hoxBb and hoxDb, with the tendency to be restricted to the abdominal B type hox paralog groups 9 13. This is in accordance with the expected expression patterns at this axial position.

be found that are expected to be expressed in anal fin tissue based on literature (Table 1).

These results provide first insights into the genes and regulatory networks involved in fin-specific cellular processes from cichlid fishes, adding to previous, more detailed work on model organisms such as zebrafish (Aday *et al.* 2011; Bogdanovic *et al.* 2012). In addition to the many genes that are likely to be involved in structural maintenance of the fin tissue such as *fras1* and *klf6b* that are known to result in fin degeneration in zebrafish (van Eeden *et al.* 1996), promoter activation for many transcription factors (including *hox* and *dlx* genes) and signalling-related factors was observed.

It has been previously shown that the hox gene expression patterns that are established during embryogenesis are actively maintained (Chang *et al.* 2002; Rinn *et al.* 2006), most likely by epigenetic mechanisms influencing histone marks such as H3K4me3 and the repressive mark H3K27me3 (Noordermeer & Duboule 2013). Furthermore, genes involved in signalling pathways involved during fin development and regeneration including sonic hedgehog (*shh*) (Avaron *et al.* 2007), Wnt/ß-catenin (Wehner *et al.* 2014), retinoic acid (Gibert *et al.* 2006), Bmp (Smith *et al.* 2006) and Notch signalling (Münch *et al.* 2013) have H3K4me3-enriched promoters and a similar scenario of sustained epigenetic activation during adulthood is likely. This suite of genes might provide the cells with a programme that is poised for activation in the case of injury, by equipping the cells with a positional memory. It is likely that these genes are actively 'reused' during regenerative processes, where complex and intricate tissues (such as the anal fin) can be faithfully restored (Knopf *et al.* 2011; Blum & Begemann 2012; Nachtrab *et al.* 2013).

Mapping genomewide occupancy of H3K4me3 in tilapia fin tissue encourages further investigations. First, as was previously observed in humans (Cheung *et al.* 2010), H3K4me3 maps can be used as a guide to find novel nonannotated transcripts in the genome (Fig. S3, Supporting information). Furthermore, for many genes, the actual promoter as well as the 5′-UTR is not annotated (see Cluster 2 in Fig. 2c,d), because annotations are very often solely based on gene predictions. H3K4me3 maps would be therefore particularly useful for improving the annotation of genomes of nonmodel organisms. Lastly, many protein-coding genes use alternative promoters [around 20% in mammals (Carninci *et al.* 2006)], which are also barely mapped and which are a powerful mechanism for selectively modifying gene expression of specific transcripts during the course of evolution (Landry *et al.* 2003; Carninci *et al.* 2006). Alternative promoter evolution has been under-analysed in the context of teleost or cichlid evolution and diversification.

Second, the use of ChIP-seq for examining the basis of evolutionary novelties extends the search for regulatory evolution beyond promoter sequence evolution (Main *et al.* 2013; Nepal *et al.* 2013), towards epigenetic alterations, which are also likely to play a role in stably altering gene expression. Natural variation of histone mark distribution of H3K27me3 has been shown to occur even within species (Dong *et al.* 2012), so it is plausible that even population-level differences are underlain by such epigenetic alterations.

Third, further histone marks exist that have been convincingly shown to be enriched at other functional noncoding elements (Zhou *et al.* 2011) for which ChIP-seq is very likely to work as well. Regulatory landscapes of genomes are highly complex. It has been estimated that around 1 million regulatory elements are controlling the expression of the ca. 25.000 genes we find in mammals (de Laat & Duboule 2013). ChIP-seq has been recently used to map tissue-specific regulatory landscapes (Visel *et al.* 2009; Attanasio *et al.* 2013), making it even possible to map and test the involvement of single regulatory elements in phenotypic variation of the cranial structure of laboratory mice (Attanasio *et al.* 2013). A further aim will be to compare and expand the ChIP-seq analysis from promoter-associated histone marks such as H3K4me3 to the enhancer-associated histone marks H3K4me1 and H3K27ac to be able to map more distal regulatory regions. The distance between a regulatory element and the controlling gene is, on average, around 120 kb (Sanyal *et al.* 2012) and can be in excess of 1000 kb (Benko *et al.* 2009; Kleinjan & Coutinho 2009), which makes it almost impossible to find them with comparative *in silico* approaches.

An intrinsic problem of *in vivo* studies using next-generation sequencing techniques such as ChIP-seq and RNA-seq that are performed in a comparative manner between species is the heterogeneity of analysed tissues. Fin tissues consist of many different cell types including pigment cells (melanophores, xanthophores, iridophores), osteoblasts, fibroblasts, epidermal cells and vascular cell types (Tu & Johnson 2011), making it difficult to draw direct conclusions to single-cell types. Still, ChIP-seq has the power to help in finding cis-regulatory elements throughout the genome. In nonmodel organisms, especially vertebrates, most comparative analyses have focused on coding regions, simply because noncoding elements are not annotated and therefore unknown. This is a major drawback if one does research in these species, and ChIP-seq could significantly aid in the analysis of noncoding DNA. In the future, ChIP-seq could especially help to identify cis-regulatory elements such as enhancers and analyse these sequences between the sequenced (and unsequenced) cichlid genomes for positive or purifying selection.

The genomes of five African (Brawand *et al.* 2014) and one Neotropical cichlid species (Fan S, Meyer A, unpublished data) have been recently sequenced, providing a powerful means for identifying the genomic bases of the extreme phenotypic variability observed within this family. We consider the cichlid family to be an excellent 'model-family' of vertebrates in which to study the role of cis-regulatory elements such as promoter regions and enhancers during evolution and phenotypic variation (Kratochwil & Meyer 2014). Expression divergence associated with transposable element insertions has been observed in the cichlid lineage (Brawand *et al.* 2014), and it is likely that the transposition of regulatory regions contributes to genomic, regulatory and thereby phenotypic diversity and eventually speciation. ChIP-seq for histone modifications will be a promising technique to map and comparatively analyse cis-regulatory elements on a genomewide scale in an extremely species-rich and phenotypically diverse vertebrate family.

Acknowledgements

References

Aday AW, Zhu LJ, Lakshmanan A, Wang J, Lawson ND (2011) Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Developmental Biology*, **357**, 450 462.

Attanasio C, Nord AS, Zhu Y *et al.* (2013) Fine tuning of craniofacial morphology by distant-acting enhancers. *Science*, **342**, 1241006.

Avaron F, Smith A, Akimenko M-A (2007) Sonic Hedgehog signalling in the developing and regenerating fins of zebrafish. In: *. . . and Gli Signal ling and Development Molecular Biology Intelligence Unit*, pp. 93 106. Springer New York, New York, NY.

Baldo L, Santos ME, Salzburger W (2011) Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome Biology and Evolution*, **3**, 443 455.

Barski A, Cuddapah S, Cui K *et al.* (2007) High-resolution profiling of his tone methylations in the human genome. *Cell*, **129**, 823 837.

Benko S, Fantes JA, Amiel J *et al.* (2009) Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics*, **41**, 359 364.

Blow MJ, McCulley DJ, Li Z *et al.* (2010) ChIP-Seq identification of weakly conserved heart enhancers. *Nature Genetics*, **42**, 806 810.

Blum N, Begemann G (2012) Retinoic acid signaling controls the formation, proliferation and survival of the blastema during adult zebrafish fin regeneration. *Development*, **139**, 107 116.

Bogdanovic O, Fernandez-Miñán A, Tena JJ *et al.* (2012) Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*, **22**, 2043 2053.

Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375 381.

Carninci P, Sandelin A, Lenhard B *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, **38**, 626 635.

Chang HY, Chi J-T, Dudoit S *et al.* (2002) Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proceedings of the National Academy of Sciences*, **99**, 12877 12882.

Cheung I, Shulha HP, Jiang Y *et al.* (2010) Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proceedings of the National Academy of Sciences*, **107**, 8824 8829.

Conesa A, Götz S, García-Gómez JM, *et al.* (2005) Blast2GO : a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674 3676.

Couldridge V (2002) Experimental manipulation of male eggspots demonstrates female preference for one large spot in Pseudotvopheus lombavdoi. *Journal of Fish Biology*, **60**, 726 730.

Dong X, Reimer J, Göbel U *et al.* (2012) Natural variation of H3K27me3 distribution between two *Arabidopsis* accessions and its association with flanking transposable elements. *Genome Biology*, **13**, R117.

van Eeden FJ, Granato M, Schach U *et al.* (1996) Genetic analysis of fin formation in the zebrafish, *Danio rerio*. *Development*, **123**, 255 262.

Egger B, Klaefiger Y, Theis A, Salzburger W (2011) A sensory bias has triggered the evolution of egg-spots in cichlid fishes. *PLoS ONE*, **6**, e25601.

Fan S, Elmer KR, Meyer A (2012) Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philosophical transactions of the Royal Society of London Ser ies B, Biological Sciences*, **367**, 385 394.

Fernald RD (1976) The effect of testosterone on the behavior and coloration of adult male cichlid fish (Haplochromis burtoni, Günther). *Hor mone Research*, **7**, 172 178.

Fernald RD (2012) Social control of the brain. *Annual Review of Neurosci ence*, **35**, 133 151.

Franchini P, Fruciano C, Spreitzer ML *et al.* (2014) Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Molecular Ecology*, **23**, 1828 1845.

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature Reviews Genetics*, **13**, 840 852.

Gibert Y, Gajewski A, Meyer A, Begemann G (2006) Induction and prep-atterning of the zebrafish pectoral fin bud requires axial retinoic acid signaling. *Development*, **133**, 2649 2659.

Grosenick L, Clement TS, Fernald RD (2007) Fish can infer social rank by observation alone. *Nature*, **445**, 429 432.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, **130**, 77 88.

Gunter HM, Fan S, Xiong F *et al.* (2013) Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. *Molecular Ecology*, **22**, 4516 4531.

Gunter HM, Koppermann C, Meyer A (2014) Revisiting de Beer's text-book example of heterochrony and jaw elongation in fish: calmodulin expression reflects heterochronic growth, and underlies morphological innovation in the jaws of belonoid fishes. *EvoDevo*, **5**, 8.

Gupta MV, Acosta BO (2004) A review of global tilapia farming practices. *Aquaculture Asia*, **9**, 7 12.

Heintzman ND, Stuart RK, Hon G *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, **39**, 311 318.

Henning F, Meyer A (2012) Eggspot number and sexual selection in the cichlid fish Astatotilapia burtoni. *PLoS ONE*, **7**, e43695.

Henning F, Meyer A (2014) The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. *Annual Review of Genomics and Human Genetics*, **15**, 417 441.

Henning F, Jones JC, Franchini P, Meyer A (2013) Transcriptomics of morphological color change in polychromatic Midas cichlids. *BMC Genomics*, **14**, 171.

Hert E (1989) The function of egg-spots in an African mouth-brooding cichlid fish. *Animal Behaviour*, **37**, 726 732.

Katogi R, Nakatani Y, Shin-i T *et al.* (2004) Large-scale analysis of the genes involved in fin regeneration and blastema formation in the medaka, Oryzias latipes. *Mechanisms of Development*, **121**, 861 872.

Keyte AL, Smith KK (2014) Heterochrony and developmental timing mechanisms: changing ontogenies in evolution. *Seminars in Cell and Developmental Biology*, **34**, 99 107.

Kleinjan D-J, Coutinho P (2009) Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. *Briefings in Functional Genomics and Proteomics*, **8**, 317 332.

Knopf F, Hammond C, Chekuru A *et al.* (2011) Bone regenerates via dedifferentiation of osteoblasts in the zebrafish fin. *Developmental Cell*, **20**, 713 724.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **5**, 288 298.

Kowalczyk MS, Hughes JR, Garrick D *et al.* (2012) Intragenic enhancers act as alternative promoters. *Molecular Cell*, **45**, 447 458.

Kratochwil CF, Meyer A (2014) Closing the genotype-phenotype gap: emerging technologies for evolutionary genetics in ecological model vertebrate systems. *BioEssays*. doi:10.1002/bies.201400142 [Epub ahead of print].

Kuraku S, Meyer A (2008) Genomic analysis of cichlid fish 'natural mutants'. *Current Opinion in Genetics and Development*, **18**, 551 558.

de Laat W, Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, **502**, 499 506.

Landry J-R, Mager DL, Wilhelm BT (2003) Complex controls: the role of alternative promoters in mammalian genomes. *Trends in Genetics*, **19**, 640 648.

Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, **13**, 233 245.

Levine M (2010) Transcriptional enhancers in animal development and evolution. *Current Biology*, **20**, R754 R763.

Lowe CB, Kellis M, Siepel A *et al.* (2011) Three periods of regulatory innovation during vertebrate evolution. *Science*, **333**, 1019 1024.

Maan ME, Sefc KM (2013) Colour variation in cichlid fish: developmental mechanisms, selective pressures and evolutionary consequences. *Seminars in Cell and Developmental Biology*, **24**, 516 528.

Main BJ, Smith AD, Jang H, Nuzhdin SV (2013) Transcription start site evolution in *Drosophila*. *Molecular Biology and Evolution*, **30**, 1966 1974.

Mattersdorfer K, Koblmüller S, Sefc KM (2012) AFLP genome scans suggest divergent selection on colour patterning in allopatric colour morphs of a cichlid fish. *Molecular Ecology*, **21**, 3531 3544.

Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in Ecology and Evolution*, **8**, 279 284.

Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, **347**, 550 553.

Mikkelsen TS, Ku M, Jaffe DB *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553 560.

Münch J, González-Rajal A, la Pompa de JL (2013) Notch regulates blastema proliferation and prevents differentiation during adult zebrafish fin regeneration. *Development*, **140**, 1402 1411.

Nachtrab G, Kikuchi K, Tornini VA, Poss KD (2013) Transcriptional components of anteroposterior positional information during zebrafish fin regeneration. *Development*, **140**, 3754 3764.

Nepal C, Hadzhiev Y, Previti C *et al.* (2013) Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Research*, **23**, 1938 1950.

Noordermeer D, Duboule D (2013) Chromatin architectures and Hox gene collinearity. *Current Topics in Developmental Biology*, **104**, 113 148.

Offen N, Meyer A, Begemann G (2009) Identification of novel genes involved in the development of the sword and gonopodium in swordtail fish. *Developmental Dynamics*, **238**, 1674 1687.

Oliveira RF, Canário AV (2000) Hormones and social behavior of cichlid fishes: a case study in the *Mozambique tilapia*. *Journal of Aquaculture and Aquatic Sciences*, **9**, 187 207.

Parikh VN, Clement TS, Fernald RD (2006) Androgen level and male social status in the African cichlid, Astatotilapia burtoni. *Behavioural Brain Research*, **166**, 291 295.

Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*, **10**, 669 680.

Prud'homme B, Gompel N, Rokas A *et al.* (2006) Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, **440**, 1050 1053.

Ramagopalan SV, Heger A, Berlanga AJ *et al.* (2010) A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Research*, **20**, 1352 1360.

Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3*, **3**, 65 74.

Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY (2006) Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genetics*, **2**, e119.

Ruiz-Narváez EA (2013) Use of alternative promoters may hide genetic effects on phenotypic traits. *Journal of Human Genetics*, **58**, 47 50.

Salzburger W, Meyer A (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Die Naturwissenschaften*, **91**, 277 290.

Salzburger W, Braasch I, Meyer A (2007) Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. *BMC Biology*, **5**, 51.

Sanetra M, Henning F, Fukamachi S, Meyer A (2009) A microsatellite-based genetic linkage map of the cichlid fish, Astatotilapia burtoni (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids. *Genetics*, **182**, 387 397.

Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. *Nature*, **489**, 109 113.

Schmidt D, Wilson MD, Spyrou C *et al.* (2009) ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods*, **48**, 240 248.

Schmidt D, Wilson MD, Ballester B *et al.* (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, **328**, 1036 1040.

Schwarzer J, Misof B, Tautz D, Schliewen UK (2009) The root of the East African cichlid radiations. *BMC Evolutionary Biology*, **9**, 186.

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research.. *Proceedings Biological Sciences/The Royal Society*, **273**, 1987 1998.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620 626.

Smith A, Avaron F, Guay D, Padhi BK, Akimenko MA (2006) Inhibition of BMP signaling during zebrafish fin regeneration disrupts fin growth and scleroblasts differentiation and function. *Developmental Biology*, **299**, 438 454.

Soshnikova N (2014) Hox genes regulation in vertebrates. *Developmental Dynamics*, **243**, 49 58.

Soshnikova N, Duboule D (2009) Epigenetic temporal control of mouse Hox genes *in vivo*. *Science*, **324**, 1320 1323.

Stefflova K, Thybert D, Wilson MD *et al.* (2013) Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell*, **154**, 530 540.

Stiassny ML, Meyer A (1999) Cichlids of the rift lakes. *Scientific American*, **280**, 64 69.

Tawk M, Joulie C, Vriz S (2000) Zebrafish Hsp40 and Hsc70 genes are both induced during caudal fin regeneration. *Mechanisms of Development*, **99**, 183 186.

Tena JJ, González-Aguilera C, Fernandez-Miñán A *et al.* (2014) Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period. *Genome Research*, **24**, 1075 1085.

Tu S, Johnson SL (2011) Fate restriction in the growing and regenerating zebrafish fin. *Developmental Cell*, **20**, 725 732.

Van Alphen J (1999) Evolution of colour patterns in East African cichlid fish. *Journal of Evolutionary Biology*, **12**, 514 534.

Visel A, Blow MJ, Li Z *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854 858.

Wapler-Leong DC, Reinboth R (1974) The influence of androgenic hormone on the behavior of *Haplochromis burtoni* (Cichlidae). *Fortschritte der Zoologie*, **22**, 334 339.

Wehner D, Cizelsky W, Vasudevaro MD *et al.* (2014) Wnt/β-catenin signaling defines organizing centers that orchestrate growth and differentiation of the regenerating zebrafish caudal fin. *Cell Reports*, **6**, 467 481.

Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, **13**, 59 69.

Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, **8**, 206 216.

Ye T, Krebs AR, Choukrallah M-A *et al.* (2011) seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Research*, **39**, e35.

Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews Genetics*, **12**, 7 18.

Data accessibility

The ChIP-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under Accession no. GSE62791.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig S1 Clustal analysis of cluster 2 and 3 were further analyzed using a second level clustal analysis.

Fig S2 Coverage plots of representive examples of genes fallling into the clusters 1, 2.2, 3.2 and 3.5.

Fig S3 Example how ChiP seq can be used to find a new non annotated genes.

Fig S4 Examplary coverage plots of genes with implications for fin development and regeneration that show enrichment of H3K4me3.

Table S1 List of ChIP seq peaks with position and neighboring genes.

Table S2 Gene list of Cluster 1 3 from the cluster analysis.

Table S3 List with all over and underrepresented GO Terms in cluster 1.