Volume 37 No. 2 February 2015 · ISSN 0265-9247

BioEssays

2/15

Ideas that Push the Boundaries

Hypotheses
Perspectives
Reviews

www.bioessays-journal.com

WILEY

Closing the genotype–phenotype gap: Emerging technologies for evolutionary genetics in ecological model vertebrate systems

Claudius F. Kratochwil[1)2)] *and Axel Meyer*[1)*]

The analysis of genetic and epigenetic mechanisms of the genotype–phenotypic connection has, so far, only been possible in a handful of genetic model systems. Recent technological advances, including next-generation sequencing methods such as RNA-seq, ChIP-seq and RAD-seq, and genome-editing approaches including CRISPR-Cas, now permit to address these fundamental questions of biology also in organisms that have been studied in their natural habitats. We provide an overview of the benefits and drawbacks of these novel techniques and experimental approaches that can now be applied to ecological and evolutionary vertebrate models such as sticklebacks and cichlid fish. We can anticipate that these new methods will increase the understanding of the genetic and epigenetic factors influencing adaptations and phenotypic variation in ecological settings. These new arrows in the methodological quiver of ecologist will drastically increase the understanding of the genetic basis of adaptive traits – leading to a further closing of the genotype–phenotype gap.

Keywords:
adaptation; cichlids; epigenetics; gene regulation; genome editing; model organisms; phenotypic variation

Introduction

One of the main goals of evolutionary genetics is to understand the molecular basis of phenotypic traits that are key-innovations and drive phenotypic diversification and speciation. Many genes involved in quantitative traits such as coloration patterns [1], body shapes [2], form and length of body appendages including fins [3] or other skeletal structures [4] have been identified in mutagenesis screens and by targeted mutations in model organisms such as zebrafish [5] and mouse [6]. But, the majority of traits observed in nature can obviously not be 'modelled' by using a few supposedly representative organisms, that have been inbred for many generations to minimize their genetic variability and that may not have exactly those structures of ecological importance in terms of adaptations and speciation. Systematic approaches such as comparative genomics, association studies and QTL analyses are very powerful methods for identifying genomic regions or even genes and regulatory sequence stretches that are likely to be causally involved in certain traits. In the best case these approaches might identify the actual causal mutations – the quantitative trait nucleotides (QTNs). But, many species cannot be bred at all or in sufficient numbers and within the lifetime of a researcher for QTL analyses. And, until recently, functional validation of targets that are purportedly involved in phenotypic variation,

DOI 10.1002/bies.201400142

[1)] Chair in Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Konstanz, Germany
[2)] Zukunftskolleg, University of Konstanz, Konstanz, Germany

***Corresponding author:**
Axel Meyer
E-mail: Axel.Meyer@uni-konstanz.de

Abbreviations:
ChIP-seq, chromatin-immuno-precipiation with high-throughput sequencing; **CRISPR-Cas,** clustered regulatory interspaced short palindromic repeat associated proteins; **ISH,** in situ hybridization; **lncRNAs,** long non-coding RNAs; **miRNA,** microRNA; **qPCR,** quantitative PCR; **QTL,** quantitative trait loci; **RAD-seq,** restriction site-associated DNA sequencing; **RNA-seq,** RNA-sequencing; **SNP,** single nucleotide polymorphism; **TALENs,** transcription activator-like effector nucleases; **ZFNs,** zinc finger nucleases.

has been extremely challenging, or had to be performed in, ideally, closely related model organisms [7].

During the first half of the last century large segments of the scientific community started to concentrate research on only a handful of species and also attempted to reduce their genetic variance. The idea of the 'model organism' was born [8]. While *Drosophila melanogaster* is by far the most widely-used genetic model system, the house mouse, *Mus musculus* remained for a long time the only vertebrate model organisms that was used for genetic studies until the zebrafish, *Danio rerio* entered the research scene in the 70's [9]. The restriction to just a few species and inbred lines with highly similar genetic backgrounds had strong advantages for the analysis of genetic mechanisms, reducing biological variation, thus enhancing experimental reproducibility. Over the years the advantages of this approach grew further due to the possibility to share resources such as experimental methods, techniques, mutant strains, genetic maps and, more recently, the availability of sequenced genomes [8]. Forward genetics screens have classically been applied to studies of model organisms to better understand genotype–phenotype relationships. By generating random mutations in mutagenesis screens, linking them to aberrant phenotypes and localizing the mutation in the genome, research on model organisms has dramatically increased the understanding of gene functions as well as lead to the discovery of previously unknown genes. The function of particular genes could be more efficiently analysed in a homogenous genetic background. Nevertheless there are many limitations to this approach, including, for example, difficulties in the analysis of genes with redundant function or pleiotropic genes, that in many cases induce early lethality [5]. Surely, the random nature of the screens that differs greatly from the natural occurring variation also pose a limitation – most genes that have been found are large effect mutations that are very unlikely to play a role during phenotypic diversification [10].

A further drawback was the strong limitation caused by the intrinsically low degree of phenotypic diversity that can be investigated in a homogeneous genetic background by, for example, mutagenesis screens. Using only model organisms it was therefore almost impossible to analyse adaptively relevant phenotypic diversity. Another reason is, of course, that the 'ecology' of *Drosophila* as well as other model organisms such as *Caenorhabditis* is famously unknown. This lack of knowledge of the natural habitat and the organisms' real biology outside the laboratory was a hindrance to understanding the whole organism. On the other hand evolutionary and population genetics in vertebrates remained restricted to mostly correlative analyses. These disciplines mostly focused – due to methodological limits – on the coding portions of the genome. They lacked experimental approaches for the identification of stretches of non-coding DNA such as regulatory elements that are involved in phenotypically diverse traits [11].

The exponentially increasing number of sequenced genomes [12, 13] and the plethora of resources acquired by next-generation sequencing methods for genome-wide analyses including RNA-seq, ChIP-seq and RAD-seq is about to drastically expand the scope of species amenable for studying the genotype–phenotype connection beyond a handful of highly genetically-uniform model organisms [14–17]. The current rise of comparative genomics and epigenomics should help to broaden the view of molecular biologists, by allowing them now to study intra- and inter-species diversity [17]. Furthermore, more and more critics begin to stress the fact that even standardizing experimental conditions, leads to a deterioration of the reproducibility of experiments rather than improving it – a caveat that is especially relevant for research on laboratory organisms. While experimental and genetic standardization indeed reduces 'within-experiment variation', it might not increase the applicability of a result to other conditions, species, strains or populations and may reveal 'local truths' that cannot be generalized [18].

The number of organisms used for genetic studies is steadily increasing [19]. It is obvious that additional model species are needed to understand the basis of phenotypic and organismal diversity. Detailed evolutionary analyses are increasingly recognized as being advantageous to clarify the functions of coding and especially non-coding DNA [11]. Here, we provide an overview of emerging technologies, which, until now, have mainly been used only in model organisms and human cell lines. We will discuss new next-generation-sequencing applications such as ChIP-seq, which might ultimately help to find non-coding DNA stretches involved in phenotypic variation in non-model organisms as well as transgenic and genome editing approaches – in particular CRISPR-Cas – and how they can be combined with classical approaches such as QTLs (Box 1).

Accessing the role of non-coding DNA and epigenetic marks

Only a few non-coding elements implicated in phenotypic evolution are known so far

Previously, the misnomer 'junk DNA' was attached to non-coding portions of the genome. Now, the role of non-coding DNA (Fig. 1) is at the focus of much research [20]. It was recently suggested that it indeed constitutes a major catalyst for phenotypic evolution and might explain many or even most phenotypic differences [21, 22]. Most micro-evolutionary comparisons have relied on SNP-based analyses and the detection of modified sequence stretches [23]. Due to technical limitations, comparative genome analyses historically focused on uncovering changes in protein-coding portions and not on non-coding DNA. But, also because of the methodological difficulties only a few non-coding regions with large, ecologically relevant effects, have been identified so far. Research on non-coding DNA has been hampered due to the limited annotation of non-coding elements such as regulatory elements.

The regulatory landscapes of genomes are highly complex; it has been estimated that around 1 million regulatory elements control the expression of the 20,000–25,000 genes found in mammals [24]. The distance between a regulatory element and the controlling gene is, on average, around 120 kb [25], but can be in excess of 1,000 kb [26]. Beside the promoter regions that are proximal to the transcription start

Box 1: Glossary of key technologies

Next generation sequencing (NGS) technologies

NGS, also called deep sequencing, massive parallel sequencing or high-throughput sequencing are technologies that differ from the traditional sequencing approaches as, for example, Sanger sequencing by the possibility to sequence multiple fragments in parallel, reducing time and costs dramatically.

- **RNA-sequencing (RNA-seq)**: RNA sequencing is an approach to obtain a snapshot of all transcribed RNAs present in a certain cell or tissue type. The RNA is converted into a cDNA fragment library with adaptors attached to it. Those fragments are sequenced by next-generation sequencing technology. The reads can be aligned to a genome or assembled de novo to obtain the transcriptome of a species for which no sequenced genome is available.
  Example: *Differences in coding and non-coding RNA transcriptomes between tissues and species* [31, 48]
  Can be used for: Screen for differentially expressed genes, sequence variations, splicing variants, new genes in species with and without available genomes. Can ultimately help to find genes that underlie the phenotypic diversities between species by comparison of the transcriptomes of selective tissues during ontogeny or adulthood. Results can be confirmed by in situ hybridisation or quantitative PCRs and validated by genome engineering or transgenesis.
- **Chromatin-immuno-precipiation with high-throughput sequencing (ChIP-seq)**: DNA is cross-linked to proteins associated with it, sonicated (to generate small DNA fragments) and immunoprecipitated (purification based on an antibody binding to a specific protein or protein modification). Hereby DNA fragments that bind to a certain transcription factor, histone modification, polymerase or other DNA associated proteins can be selected. Those fragments are sequenced using next-generation sequencing and aligned to the genome of the respective species. Peaks of alignments suggest enrichment of the selected protein at this position. Using this approach it is possible to map, for example, putative promoters or regulatory elements such as enhancers or insulators.
  Example: *Differences in transcription factor binding between species* [35, 47]
  Can be used for: Screen for and comparative analysis of regulatory regions in sequenced genomes to find potential loci of evolution in non-coding DNA.

Computational genotype–phenotype association methods

Methods that associate the presence of specific nucleotides, genes or genomic regions to phenotypes. While there are many other bioinformatical approaches to screen genomes, transcriptomes and population data for phenotype–genotype association we focused on three of the most important approaches.

- **Quantitative Trait Locus Analysis (QTL-Analysis)**: Using QTL analysis loci can be found that affect the variation of a quantitative trait and contain the causal QTNs. QTLs are identified by their linkage to a polymorphic marker such as molecular tags (e. g. microsatellites or SNPs). For a QTL analysis two individuals with different variations of a quantitative trait (e.g. different coloration or shape) are crossed. In the segregating progenies of this cross the phenotypic variation is then subsequently linked to one or multiple loci (identified by the polymorphic markers).
  Example: *More than hundreds of studies for example the QTL to study the genomic basis of pelvic fin loss in sticklebacks* [7].
  Can be used for: Will be further used in species that allow crossing with sufficient offspring numbers in the laboratory. NGS technologies and functional validation techniques such as transgenesis and genome engineering might assist in finding the causal mutations within QTLs.
- **Genome wide association studies (GWAS) and low-density SNP arrays**: In GWAS the association between a catalogue of common SNP variants and a phenotype is evaluated across a large number of unrelated individuals. While the resolution of GWAS is much higher then in QTLs, the number of false positives is also higher making the approaches complementary. However, SNP arrays with high densities that enable fully effective GWAS studies are up to now limited to humans, domestic animals, invertebrates and plants.
  Example: *Molecular basis of European hair coloration by GWAS* [41] *or screen for loci contributing to marine–freshwater and benthic–limnetic divergence in stickleBacks by low density SNP array* [78].
  Can be used for: Since more and more SNPs are known and SNP ChIPs are decreasing in cost, the technique might spread amongst evolutionary model organisms that are used by larger communities such as wild mice, sticklebacks or cichlids to map the genomic basis of traits. However, genome resequencing technologies are, if costs are further decreasing, the superior approach.
- **Genome resequencing based association studies**: In future, due to decreasing costs, SNP arrays and QTLs are likely to be largely replaced by genome resequencing of populations and/or individuals. Hereby, while being able to decrease sample size, resolution can be increased to efficiently screen for causative loci that underlie phenotypic traits.
  Example: *Resequencing of 40 individuals from a set of different rock pigeon breeds and populations to find loci contributing phenotypic traits such as the EphB2 locus in case of the head feather crest* [80].
  Can be used for: Genome resequencing will probably replace SNP based methods and QTLs due to its higher flexibility in the selection of individuals and its higher resolution. However, limiting factor are still the costs, if one aims at high sequencing coverage and high number of individuals.

Genome editing technologies

Genome editing refers to methods that use engineered nucleases to modify genomes by the targeted insertion, removal or replacement of genomic DNA stretches including ZFNs, TALENs and CRISPR-Cas-based RNA-guided DNA endonucleases (CRISPR-Cas).

Example: (Have not been used in the context of evolutionary biology yet).

Can be used for: Functional validation of the involvement of genes, cis-regulatory regions or genomic regions that have been found using comparative analyses (RNA-seq, ChIP-seq, genome comparisons) or QTL analyses.

- **CRISPR-Cas:** Clustered regulatory interspaced short palindromic repeat (CRISPR)-Cas RNA-guided DNA endonucleases have been found 1987 in *E. coli*, but can be also found in many other eubacteria and archaea. CRISPR can be considered as part of the prokaryotic immune system that mutates exogenous DNA sequences in the genome. This system has been modified in a way that one single RNA, called single-guide RNA (gRNA or sgRNA) together with an endonuclease protein (Cas9) is able to induce double strand breaks at a targeted position, which are then repaired by error-prone endogenous repair mechanisms, inducing small deletions at this position. The specificity is hereby given by a short sequence in the gRNA (23 base-pairs).
- **TALENs:** Transcription activator-like effector nucleases are artificial proteins that consist of multiple repeats of a TALE DNA binding domains and a DNA cleavage domain. Each of the TALE binding domain has 34 amino acids of which number 12 and 13 are highly variable and their amino acid compositions shows a strong correlation to the binding of specific nucleotides. Hence, a combination of TALE domains can guide the DNA cleavage domain to a specific genomic sequence.

Transgenesis technologies

The introduction of a gene from an organism into another's genome is called transgenesis. A popular example is the introduction of the green fluorescent protein (GFP) from the jellyfish *Aquaria victoria* into other organisms as mice or zebrafish to label cells or to analyse promoters. Alternative approaches are the over-expression of genes (to test their function). Most approaches to generate transgenes trigger random insertions into the genome.

Example: Functional validation using Tol2 that a regulatory element of Pitx1 is active in pelvic fins of sticklebacks. The lack of the enhancers explains the pelvic fin loss in fresh-water populations of sticklebacks [37].

Can be used for: Similar to gene engineering; functional validation of the involvement of genes, cis-regulatory regions that have been found using comparative analyses (RNA-seq, ChIP-seq, Genome comparisons) or QTL analyses by testing regulatory element activity and over-expression of genes or gene variants of interest.

- **Tol2:** The Tol2 transposase system uses an autonomous transposon from the medaka fish *Oryzias latipes* that encodes the fully functional transposase Tol2. From this transposon two constructs can be generated: (1) A construct just consisting of the coding region of the transposon, which can be used to generate mRNA for the Tol2 transposase and (2) a construct which lacks the coding region and just has the flanking sites, which are targeted by the transposase. The sequence between these sites can now be replaced by any DNA fragment of choice and will be inserted into the DNA if Tol2 transposase (which can be supplied by adding RNA synthesized from construct 1) is present.

sites, regulatory elements can be divided into three major groups: Enhancers, silencers [27] and insulators [28]. Silencers and enhancers are binding sites for transcription-factors that upon binding activate (enhancers) or suppress (silencers) gene expression. Insulators function as genetic boundaries that prevent enhancers and silencers from interfering with the regulation of other genes.

Recently, driven by the development of high-throughput sequencing techniques, there have been more and more attempts made to map regulatory elements using various methods [27, 29]. Based on those studies it was attempted to better understand the diversity of gene regulation in different species (also in non-model organisms) [30]. Comparative analyses of gene expression using RNA-seq and transcription factor binding using ChIP-seq have the power to reveal both patterns and processes of transcriptome and gene regulation evolution. RNA-seq experiments provided the first ideas about the evolutionary dynamics of mammalian transcriptomes and the extent of transcriptome variation between species [31]. RNA-seq is now widely used and has made substantial contributions to our understanding of genome expression and regulation, in an evolutionary context. RNA-seq can be used independently of a sequenced genome and is thereby a crucial technique to fill the evolutionary gaps between the sequenced organisms. It not only allows to rapidly compare genetic sequences encoding for genes but also to detect differential gene expression between species [14, 32, 33]. Both can be ultimately used to screen for the molecular bases of species-specific traits. Although purely correlative, it is probably one of the most powerful and straightforward approaches for non-model organisms.

Comparative ChIP-seq analysis between vertebrates has shown that the binding of transcription factors (TFs) to regulatory elements is extremely dynamic, revealing significant interspecies differences in transcriptional regulation on a genome-wide scale. A large number of TF binding sites are species-specific, even when analysed in highly conserved tissues such as liver, using highly conserved TFs for ChIP [34]. Interestingly, already among very closely related species of rodents or strains of 'laboratory mice', strong differences in transcription factor binding can be observed, even in tissues with very conserved regulatory networks [35].

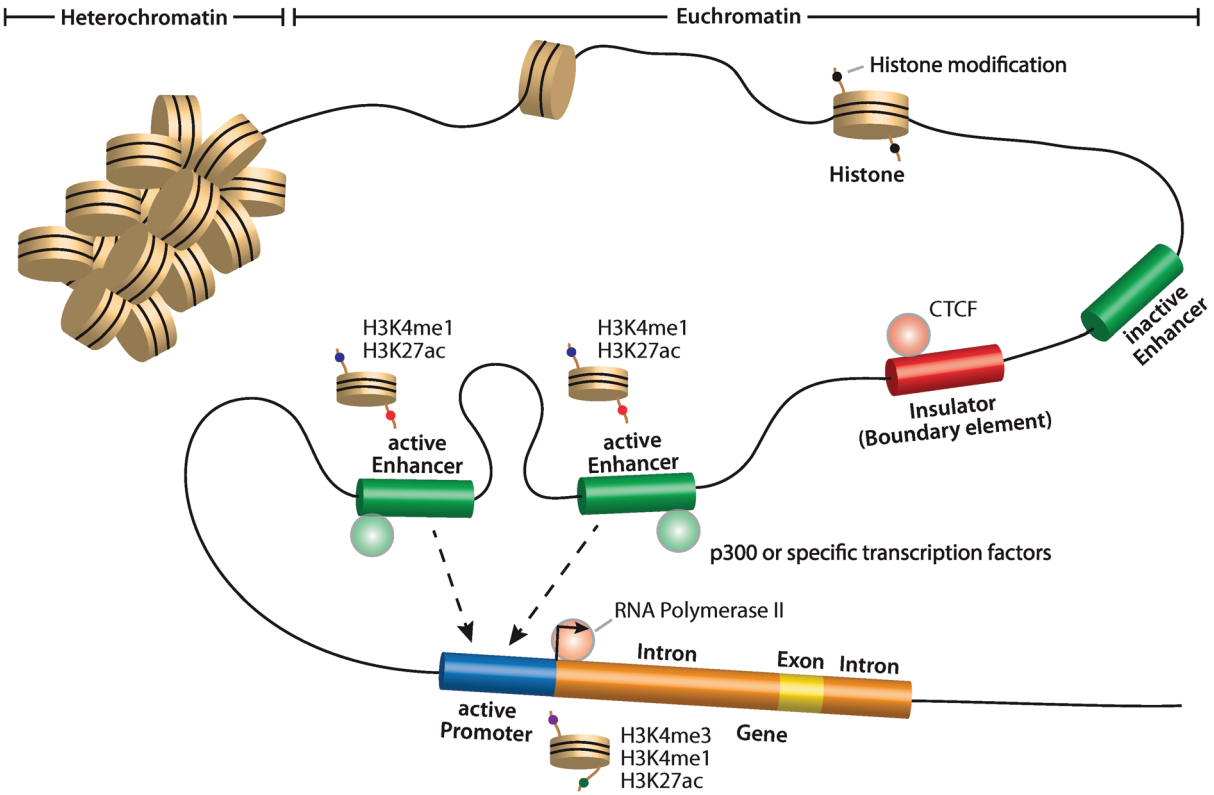

Figure 1. Overview of cis-regulatory elements involved in gene regulation. Beside mutations in coding regions also mutations in or deletions of regulatory elements such as enhancers, insulators or promoter regions can result in changes in gene expression and can eventually lead to phenotypic diversification. Sequence stretches as active enhancers, promoters or insulators can be detected by ChIP-seq using antibodies for proteins (such as CTCF, p300, specific transcription factors or RNA Polymerase II) or histone modifications (H3K4me1, H3K4me3, H3K27ac) that are enriched at these sequences.

Besides the aforementioned analyses, several studies have directly demonstrated the role of regulatory mutations in generating adaptive traits. One of the most prominent examples from vertebrates is the loss of pelvic fins in sticklebacks, which is driven by the mutation of a pelvic fin-specific regulatory element of the transcription factor Pitx1 [7, 36]. Other examples of regulatory elements include differences in limb length in mammals, triggered by regulatory mutations of Prx1 [37], differences in axial morphology triggered by mutations in hox gene enhancers [38, 39], or differences in hair coloration in humans that are explained by a regulatory mutation of KITLG [40]. A selection of recent studies analysing the roles of non-coding DNA in an evolutionary context is highlighted in Table 1. The studies range from analyses focusing on specific traits and how their expression is influenced by cis-regulatory elements [36–46], to studies that compare genome-wide gene expression and gene regulation between species and might be the basis for further work on the genomic bases of lineage-specific adaptive traits [31, 35, 47–49].

Also regulatory RNAs have attracted increasingly more attention for evolutionary questions. MiRNAs are short noncoding RNA molecules that bind to complementary sequences in messenger RNAs (mRNAs), promoting mRNA translational repression or degradation. Expansions of miRNAs are suggested to have contributed significantly to phenotypic evolution in vertebrates by the modification of post-translational regulation [50]. The haplochromine lineage of African cichlid fishes shows enrichment for novel miRNAs suggesting a role in their extreme adaptive radiation [51]. Also the role of lncRNAs is beginning to be analysed in the context of evolutionary biology. Despite their modest sequence conservation, lncRNAs have been annotated in 11 tetrapod species including 2,500 highly conserved lncRNAs and many more lineage-specific lncRNAs in primates [48]. The description of lncRNAs is a further step towards understanding their role in development, disease, as well as vertebrate evolution.

Changes in epigenetic modifications have the power to change phenotypes – but how important are they for evolution?

Beside promoters, enhancers, silencers and insulators further elements control gene regulation that are not directly caused by changes in the DNA sequence. But, they are mediated by epigenetic modifications such as DNA modifications (e.g. DNA methylation) or chromatin modifications (e.g. histone modifications). Epigenetic changes can be mitotically and meiotically stable, but are potentially reversible [52]. They are becoming easier to map and analyse, mainly driven by key technological advances (RNA-seq and ChIP-seq), and through the rapid advancement of studies on model organisms. We anticipate that these techniques will also soon be applicable to

Table 1. Selection of comparative studies analysing gene regulation and non-coding DNA in the context of vertebrate evolution.

Sort of non-coding DNA	Regulated gene(s)	Organism(s)	Tissue	Function	Used techniques	References
Enhancer/regulatory element	Bmp6	Bony fish (stickleback)	Teeth	Tooth gain	QTL, expression analysis (ISH)	Cleves at al. [45]
Enhancer/regulatory element	HoxD13	Bony fish (zebrafish)	Fin	Fin growth and fin/limb morphology	Comparative sequence analysis, zebrafish transgenics (overexpression, enhancer assay), expression analysis	Freitas et al. [42]
Enhancer/regulatory element	Kit ligand (Kitlg)	Bony fish (stickleback)	Gill, skin	Pigmentation	QTL, QTL finemapping, expression analysis (ISH, qPCR)	Miller et al. [41]
Enhancer/regulatory element	Kit ligand (Kitlg)	Humans, mice (for validation)	Hairs	Pigmentation/hair colour	Genome wide association study (GWAS), transgenesis in mice	Guenther et al. [40]
Enhancer/regulatory element	Pax7	Bony fish (cichlids: *Maylandia zebra*)	Skin	Coloration	QTL, expression analysis by qPCR	Roberts et al. [44]
Enhancer/regulatory element	Pituitary homeobox 1 (Pitx1)	Bony fish (stickleback)	Fin	Pelvic fin reduction in freshwater sticklebacks	QTL, QTL finemapping, expression analysis (ISH, qPCR), comparative sequence analysis, Mouse knockouts, stickleback transgenesis	Shapiro et al., Chan et al. [7, 36]
Enhancer/regulatory element	Ptch1	Mammals (mouse, bovine)	Limbs	Bovine limb evolution	Expression analysis (ISH), ChIP-seq, DNAseI hypersensitivity assay, 4C, Mouse transgenesis	Lopez-Rios et al. [43]
Enhancer/regulatory element	Rx1	Bony fish (cichlids: *Aulonocara baenschi* and *Tramitichromis intermedius*)	Retina	Opsin expression/ visual system diversity	QTL, computational binding site analysis, expression analysis by qPCR, resequencing in other species	Schulte et al. [46]
Enhancer/regulatory element	Myogenic factor 5 (Myf5)	Reptiles (corn snake, python, and boa), mammals (manatee, mouse)	Vertebrae	Vertebrae identity, rib Formation	Comparative Sequence Analysis, Mouse transgenics, Protein-DNA binding assay (EMSA)	Guerreiro et al. [38]
Enhancer for liver specific transcription factors	Genome wide	Mammals (4 mouse species (+1 strain); human, dog, opossum), bird (chicken)	Liver	Phenotypic differences not analysed	ChIP-seq, genome-wide comparative analysis	Schmidt et al., Stefflova et al. [35, 47]
Long noncoding RNAs (lncRNAs)	Genome wide	Mammals (human, chimpanzee, bonobo, gorilla, orangutan, macaque, mouse, opossum, platypus), birds (chicken), amphibians (xenopus)	Cortex, cerebellum, heart, kidney, liver, placenta, ovary, testis	Phenotypic differences not analysed	RNA-seq, genome-wide comparative analysis	Necsulea et al. [48]
Regulatory elements (indirectly via looking at gene expression levels)	Genome wide + special focus on peroxisomal genes	Mammals (including humans and 11 nonhuman primates)	Liver	Causal relationship between peroximal gene expression and nutrition suggested	RNA-seq, genome-wide comparative analysis	Perry et al. [49]
Regulatory elements (indirectly via looking at gene expression levels)	Genome wide	Mammals (human, chimpanzee, bonobo, gorilla, orangutan, macaque, mouse, opossum, platypus), birds (chicken)	Cortex, cerebellum, heart, kidney, liver, testis	Phenotypic differences not analysed	RNA-seq, genome-wide comparative analysis	Brawand et al. [31]

non-model organism, since the sequencing of their genomes is rapidly becoming more affordable and, as more and more genomes are known, also the assembly will become increasingly easier.

It has been shown that methylation patterns change drastically during evolution. Comparative analyses on humans and great apes provide insight into alterations in DNA methylation that might have contributed to evolutionary changes in gene regulation and thereby to phenotypic diversification [53]. Other examples suggest that DNA methylation at promoters, which can permanently block transcription during ontogeny, seems to rebalance dosage effects that are induced by gene duplication [54]. Epigenetic modifications to histones, around which the DNA is wrapped, provide a highly modifiable DNA scaffold that dictates patterns of gene expression through altering the accessibility of DNA to proteins that in turn influence gene regulation [55]. There are many known modifications to specific amino acid residues of histone proteins, including methylation or acetylation. These are driven by a large set of histone modifier proteins including polycomb and histone acetyltransferase proteins [56]. The role of these modifications in evolutionary processes has been barely analysed and remains largely unknown. There have been attempts to analyse the evolutionary history of some of the proteins involved in chromatin modification [57], but it is still unclear what role they play during vertebrate evolution. Histone modifiers orchestrate transcriptional programs throughout ontogeny from early development up to the formation of complex adult tissues such as neural circuitry and their lack induces strong misregulation of gene expression and phenotypic malformations [58, 59].

Although there is up to now little evidence that histone modifications are transgenerationally stable, they are greatly influenced by environmental variations (cellularly and externally) and DNA mutations [52]. Compared to plants, heritable epigenetic changes do not seem as common in animals; mainly due to the fact that most epigenetic tags are reset in the germline and therefore irrelevant for the next generation [60]. In fact, although this hypothesis is hard to test, their variety amongst populations and species might in many cases be simply based on genetic variation such as transposon insertions or SNPs that recruit or block epigenetic modifier enzymes – either cell-type independent or solely in specific cell lineages [52]. Still, many of these genetic changes that result in differential gene regulation might be very often mediated by epigenetic mechanisms. Changes in cis-gene-regulation are – due to a reduced number epistatic effects – rather caused by mutations in the cis-regulatory element than in the coding-sequence of the transcription factor that binds to it. Similarly, epigenetic changes might be more likely induced by changes in the DNA that recruits epigenetic modifiers. The use of wild populations or closely related species with higher genetic diversity than laboratory strains might actually facilitate the search for the genomic and epigenetic bases of differential DNA methylation and histone modification patterns. And, more importantly, it will be an exiting focus to study if and to what extent changes in epigenetic regulation are actually involved in phenotypic diversification.

Making the step from correlation to functional validation

Combining the knowledge and tool sets of model and non-model organisms

Intraspecific variation has increasingly been analysed at the genomic level. This provided insights into a broad array of regulatory elements that control the development of adaptive phenotypic traits and illustrates the complexity of the standing genetic variation of regulatory elements. Recently, the genetic regulatory network controlling craniofacial development and thereby controlling the structure and form of head and face were analysed using ChIP-seq in laboratory mice (Fig. 2) [61]. In this genome-wide analysis, the regulatory landscape of over 4,300 putative enhancers was found to be active in the course of craniofacial development. Some candidate enhancers were located up to 1.4 Mb from the next transcription start (median distance 44 kb), which makes it almost impossible to find them by in silico approaches. Although an extremely large number of these elements showed evolutionary constraints in their activity (87.5%), it is plausible that variations in this set of regulatory elements partially explains the intra- and interspecies variation observed in complex structures such as the craniofacial skeleton. Hence, data sets such as this represent an extremely valuable resource, not only for human geneticists but also especially for evolutionary biologists, who are interested in comparative analyses.

Studies of complex traits such as craniofacial morphology provide an excellent demonstration of how many mutations of small effect can tinker with the regulatory gears of genetic networks during development, and how their selection might generate diversity in structures that ultimately permit the occupation of alternative ecological niches [62]. It is likely that similar mechanisms can explain the morphological diversity of body shape or the form of body appendixes such as fins and limbs.

One of the key challenges of the next decade will be the annotation of genetic elements involved in gene regulation including regulatory elements such as enhancers (gene-activating regulatory elements) and silencers (gene-silencing regulatory elements) in whole-genome sequenced organisms. Beside regulatory elements also the evolutionary relevance of insulators, the elements that blocks the interaction of an enhancer with a promoter (Fig. 1) has been suggested in vertebrates [63] and their importance for adaptive evolution was indicated by experiments in *Drosophila* [64].

ChIP-seq for histone marks that are associated with active regulatory elements (such as H3K27ac and H3K4me1; Fig. 1) is the most promising techniques for identifying elements involved in gene regulation, since ChIP-seq for histone marks has been shown to also work in other vertebrates such as zebrafish [65] and Medaka [66]. Additional techniques that are able to reveal cis-regulatory interactions such as STARR-Seq [67], DNaseI assays [68], FAIRE-seq [69], ATAC [70] and chromosome conformation capture techniques such as 3C, 4C, 5C or Hi-C [24, 71] will eventually contribute to the understanding of genome evolution. And although these techniques require genomic resources, they are becoming

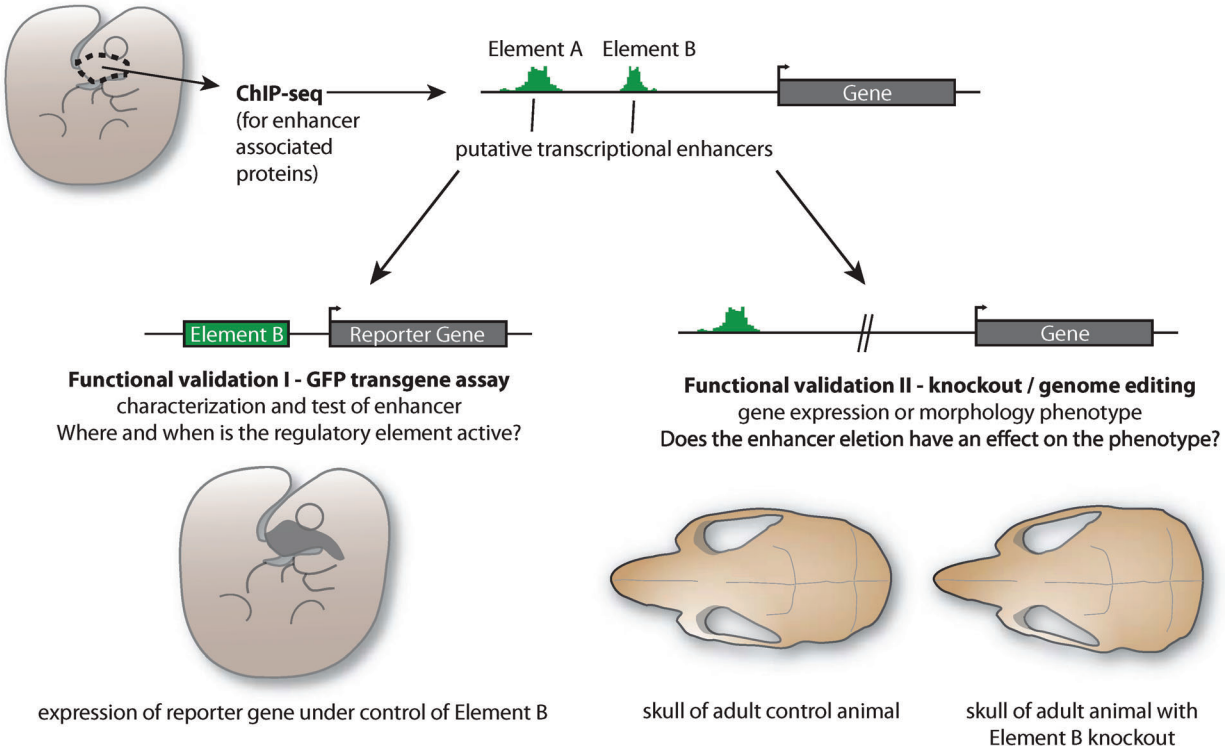

Figure 2. Screen and functional validation of enhancers involved in craniofacial morphogenesis. A recent publication from Attanasio et al. [61] showed how regulatory elements involved in certain processes (here craniofacial development) can be mapped throughout the genome using ChIP-seq. The activity of these elements can be tested by reporter gene assays using trans-genesis. It turned out that the knockout of single regulatory elements resulted in quantifiable morphological changes. If studies like this would be extended to other related species, valuable information about the involvement of regulatory elements for phenotypic evolution could be obtained.

increasingly standardized and affordable and commercial services that perform ChIP- or RNA-sequencing are available as well.

Two classical approaches to understanding the genetic basis of phenotypic variation are Mendelian trait loci (MTL) and QTL analyses [72]. By correlating genetic variation with trait variation it is possible to identify stretches of DNA that are involved in the expression of a quantitative trait. However, it is extremely challenging to find the causal mutations, especially if non-coding DNA stretches functionally related with the trait. Next-generation sequencing techniques such as RAD-Seq [16] can help in the identification of genes and cis-regulatory elements that cause phenotypic variation. RAD-seq is a powerful new technique that can help to create dense linkage maps. Those markers can then be used for fine-scale genetic mapping of MTLs and QTLs. A further approach directly includes expression levels into the QTL analyses (gene expression Quantitative Trait Locus; eQTL) and can therefore provide information on which genes might experience differential cis-regulation. Currently, eQTL analyses cannot only be performed using microarrays but also by RNA-seq [73].

Correlating phenotypes and genotypes in the postgenomic era

Beside that also computational analyses can help to find phenotype genotype links. Comparative genomic approaches together with the steadily increasing amount of genomic information on individuals, populations and different species can greatly contribute to extent association studies from single candidate genes to genome-wide approaches. A major challenge is, however, to find the causal genetic variants amongst the millions of nucleotide polymorphisms and small-scale variations (insertions, deletions and inversions) that might be involved in the expression of a trait. Already within species, there are millions of SNPs, in vertebrates around one per 300 bp [74]. In humans roughly 10 million SNPs have been reported that segregate in human populations of which 500,000 are common variations therefore most SNP genotyping chips cover only 500,000 to 2.5 million SNPs [74, 75]. The association of phenotypic and genotypic variance in GWAS helped to reveal the genetic basis of many human traits from hair coloration [40] to diseases [74]. Amongst vertebrates, GWAS has only been adapted to domestic animals including cattle, pigs and chicken using between 22,000 and 800,000 SNPs [76]. GWAS has not filtered down yet to non-model vertebrates – due to the immense costs for large-scale collection of SNPs and the need for cost-effective methods such as SNP chips to analyse these across large sets of individuals. Still, genome wide SNP genotyping arrays have been used to study patterns of genetic variation at lower resolution in non-model organisms using lower SNP densities. In example, an array for roughly 3,000 SNPs could show that the repeated adaptations to similar environments that can be found in sticklebacks seem to be triggered by

the same genetic elements [77]. However, with decreasing costs genome resequencing approaches as performed recently in sticklebacks [78] and wild pigeons [79] or reduced genome representation methods such as double digest RAD-seq [80] will be powerful alternatives to GWAS and SNP genotyping arrays [81].

Across less related species computational approaches largely focus on conserved parts of the genome including coding regions or on highly conserved non-coding elements [82], but due to the manifold of genetic changes associated to specific phenotypes this is usually infeasible. A strategy that has been taken to overcome this problem was to focus on phenotypic differences that occurred independently across lineages [83]. However, although this is a powerful approach to detect genes that are linked to phenotypic losses, since very often the same genes are implicated, phenotypic novelties that evolved in parallel often underlie different genetic causes [32] and can be therefore not detected using this methodology. In general, all association studies are purely correlative and also give no information about the relative contribution to the phenotype. Still, they are powerful means to screen for genetic targets that can be further validated by additional functional tests.

A plethora of new methods enable researchers to apply gold standard techniques now also to non-model organisms. Particularly, methods for genome engineering, such as the integration of foreign DNA into the genome (transgenesis), as well as the targeted knockout of genes and other genomic fragments (TALEN and CRISPR-Cas technologies) are increasingly applicable to less established laboratory organisms [84, 85].

Two fundamentally distinct approaches, namely insertional transgenesis and genome editing, provide powerful, targeted approaches for editing genomic DNA. Insertional transgenesis involves the introduction of DNA from a different species into the genome of a second organism, usually in a random position. Genome editing involves the modification of a predetermined sequence in the chromosomal DNA [86, 87]. Prior to modern genetic engineering, transgenesis in vertebrates was only performed in models such as mice, zebrafish, *Xenopus* and medaka.

Transgenesis allows testing of genes and regulatory elements 'out-of-context'

Insertional transgenesis can be used to verify and analyse the activity of a regulatory element (e.g. an enhancer) by combining it with reporter gene such as GFP. In this case GFP would only be expressed in tissues in which the enhancer is active. Transgenesis also allows for the testing of gene function by (1) expressing a cDNA from another organism or (2) by the expression of an endogenous gene in an ectopic position [88]. A further approach is the random integration of a reporter gene, which is then able to 'hijack' enhancers of other genes, thereby allowing screening for functional regulatory elements (enhancer trap approach [5]).

For the generation of transgenic mice, a linearized DNA construct is injected into one of the pronuclei and the injected eggs are then transferred into foster mice. Usually 10–25% of injected eggs integrate the DNA into their genomes [89]. In zebrafish, medaka and amphibians, however, the efficiency is lower. The most commonly used approaches to increase efficiency are the I-SceI meganuclease system [90] or the transposon-mediated systems Tol2 [84] and *sleeping beauty* [91]. The Tol2 system is the most efficient system for insertional transgenesis and has by now been used successfully in non-model organisms such as the Nile tilapia (*Oreochromis niloticus*) [92], the haplochromine cichlid *Astatotilapia burtoni* [93], the short-lived African killifish (*Nothobranchius furzeri*) [94] and the threespine stickleback (*Gasterosteus aculeatus*) [36].

CRISPR-Cas allows knockouts in almost all vertebrates

The number of tools available for targeted genome editing has increased dramatically during the last years. Historically, generating knockouts in vertebrates was limited to organisms possessing embryonic stem cells (ES cells), since the modification has been achieved by homologous recombination. In recent years a variety of new techniques including target-designed ZFNs [95], TALENs [96] and, a particularly new promising technique: CRISPR-Cas-based RNA-guided DNA endonucleases (Fig. 3) [97] have been introduced. These novel genome-editing approaches extended the possibility of modifying genomes to zebrafish and other vertebrate model organisms and even allow for the modification of non-model organism genomes [98].

The versatility of all three techniques comes from their ability to customize the DNA-binding domain of the protein (for ZFNs and TALENs) or the DNA-binding RNA (in case of CRISPR-Cas). Through modifying the DNA-binding protein domain or RNA, virtually any sequence can be targeted. After DNA binding, all three proteins induce DNA double-strand breaks that stimulate error-prone DNA repair at the targeted genomic locations [85]. While for ZFNs and TALENs the DNA binding domain has to be modified, CRISPR-Cas uses an RNA (referred to as guide RNA or gRNA), which provides the target specificity for the Cas protein possessing the endonuclease domain. As binding may occur with imperfectly matching sequences, the modification of 'off-targets' and its detection has been analysed and discussed recently [99]. Despite this drawback, the technique is extremely powerful in analysing gene function or non-coding DNA in many organisms, even under mosaic conditions (i.e. when a gene is not knocked out in all cells). In zebrafish it was recently shown that one can successfully knockout of up to five different genes (corresponding to ten different alleles) [100]. Similarly, in mice five genes could be knocked out at once [101]. Such a concurrent knockout is especially important if paralogous, functionally redundant genes are present in the genome, as is the case in many fish species due to their genome duplication [102] or if epistatic relationships are analysed [103].

Furthermore, by using CRISPR-Cas it is not only possible to trigger deleterious mutations, but also to introduce DNA from other species at specific positions into the genome using homologous recombination or non-homologous end joining [104]. This allows for testing genes or regulatory genomic

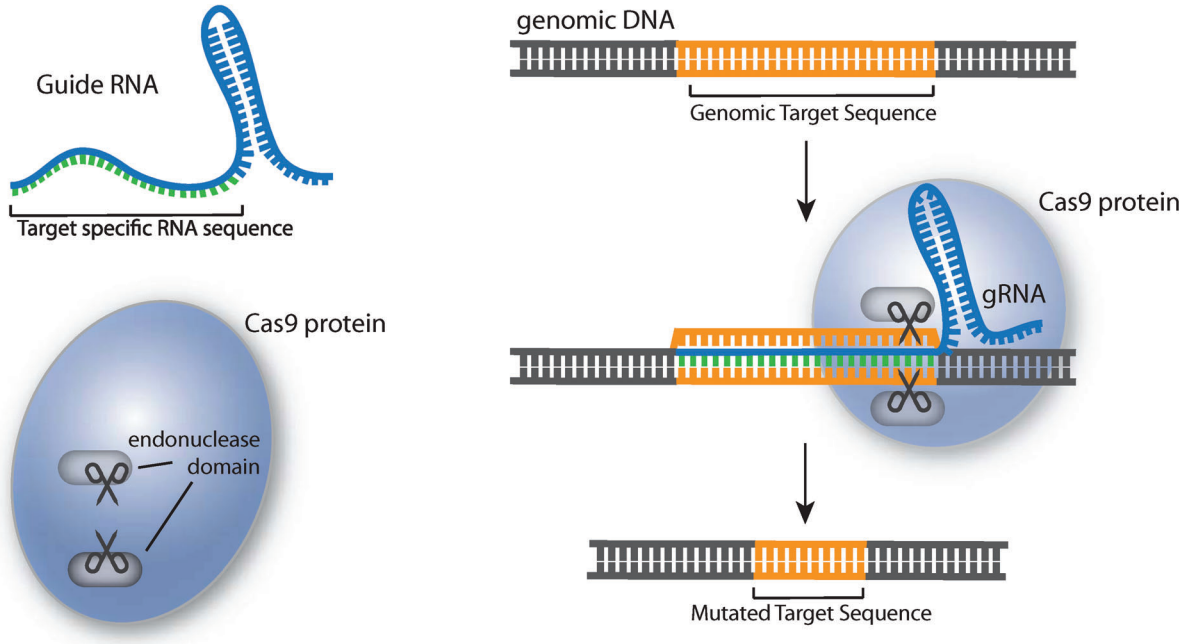

Figure 3. Overview of the CRISPR-Cas mechanism to knockout genomic sequences. The CRISPR-Cas system consists of a protein (Cas9) with endonuclease domain and the guide RNA (gRNA). The Cas9-gRNA complex is able to bind to genomic DNA sequences, which are complementary to the 23 bp target sequence on the gRNA and induce a double-strand-break at this position. The mutation frequencies at the break points are high, resulting in small deletions in the targeted sequence.

regions in the genomic context of another species. Even transcriptional activation and repression of regulatory elements or gene expression can directly be regulated by CRISPR-Cas, as the endonuclease domain can be replaced by a transcriptional regulator [105]. Further, more sophisticated genetic approaches such as those used in mice and zebrafish can, for example, restrict knockouts to certain cell-types to assess gene function in an even more precise manner (Cre/lox and Gal4/UAS system) [106].

From phenotypes to genotypes to causal mutations

The identification of the gene(s) underlying MTLs and QTLs is still very challenging. Despite considerable efforts, the causal sequence changes have been described only a handful of loci [107]. The precise identification of causal coding or even non-coding DNA stretches could benefit hugely from the use of genome editing approaches. Especially for organisms with large progenies such as many fish and amphibians, one could screen for the causal genes within a MTL or QTL interval. While a MTL with a small number of genes inside the interval would be the easiest scenario, more complicated scenarios including (1) multiple loci, (2) large loci with many genes and (3) the involvement of non-coding elements might be still be resolvable using genome editing approaches. Multiple genes could be targeted by a cocktail of gRNAs, resulting in multiple knockouts as previously shown in zebrafish and mice [100, 101]. This would enable the analysis of more complex, multigenic phenotypic traits. Since genome editing using TALENs and CRISPR-Cas also allows for the deletion of bigger genomic regions [108], it would be possible to narrow down the QTL interval by removal of larger fractions of the QTL and the evaluation of the resulting phenotypes. This method might also be applicable to screen for the position of potentially involved regulatory elements, by systematically knocking out genomic regions and screening for phenotypes resembling the corresponding parental trait.

Cichlid fishes as models for closing the phenotype–genotype gap

Although genomic resources such as genetic maps, genomes and transcriptomes are available for more and more species, means to perform genetic mappings (such as QTL), embryology or functional experiments (such as transgenesis and genome editing) are limited to organisms that can be successfully bred and interbred in sufficient numbers under laboratory conditions. Therefore, amongst vertebrates, those experiments will most probably stay restricted to rodents (since functional experiments can be performed in laboratory mice) and diverse teleost families such as sticklebacks, cichlids and killifish in which transgenesis, genome editing technologies and genetic mappings are applicable [109].

The family of cichlid fishes (*Cichlidae*) are an especially excellent model system to understand phenotypic diversification from a genomic standpoint (Fig. 4). Cichlids are a famous example for explosive adaptive radiation – in less then a few million years over 1,000 species evolved in the three east African lakes Lake Victoria, Lake Tanganyika and Lake Malawi [110]. The astonishing rate of diversification makes

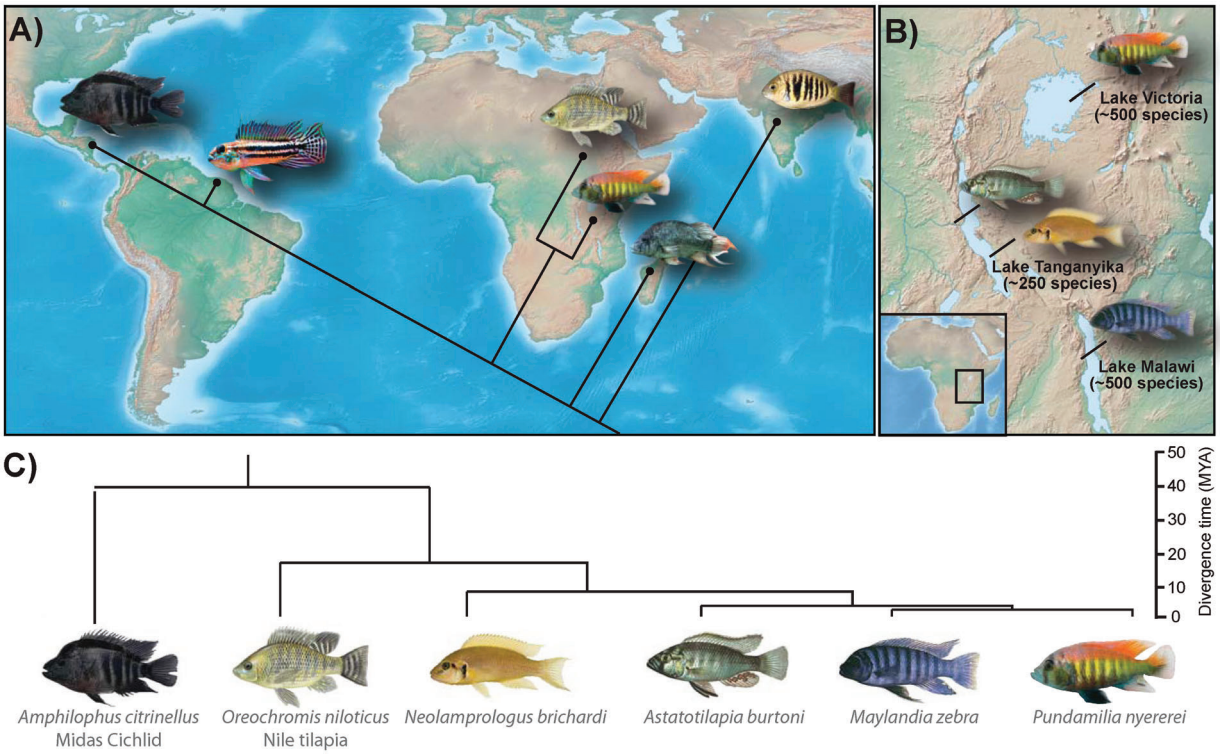

Figure 4. Overview of cichlid fishes. **A:** Worldwide geographic distribution of cichlids with representatives from India and Madagascar forming the most basal lineages and the monophyletic African and South and Central American lineages as sister-groups. **B:** The cichlids' center of biodiversity is East Africa, where more than 1,500 cichlid species are recognized. Estimated species numbers for the big lakes Victoria, Tanganyika and Malawi are given in brackets. **C:** Phylogenetic tree of the six cichlid species from Central America (*A. citrinellus*) and Africa with high-quality genome drafts. Divergence times are based on the lower timescale of [51].

them an exceptionally suitable family of vertebrates to investigate the evolutionary role of coding regions. But, the more likely cause for their extremely rapid rates of diversification are non-coding elements such as cis-regulatory elements. The deluge of phenotypes that can be found in cichlids include an enormous diversity of pigmentation and coloration patterns, body shapes and jaw- and lip forms [111]. Also highly complex behaviours evolved in cichlids. These include elaborate dominance behaviours, astonishingly diverse strategies for predation and predator avoidance, as well as courtship and brood care [111]. The best-known example of extensive brood care is probably the maternal mouth brooding of the haplochromine cichlids. Cichlids cannot be only found in Africa, but also in India, Madagascar and in South- and Central America [112]. Neotropical cichlids from Nicaragua, Central-America, are becoming a model system for evolutionary research since they include one of the few empirical examples of sympatric speciation [113]. In Nicaragua, several crater lakes have been colonized independently from the big lakes Managua and Nicaragua and evolved within a very short period of time (<25,000 years, some <2,000 years), making them an excellent model for the study of early stages of phenotypic diversification. Cichlids might be seen as a 'natural mutagenesis screen', allowing testing for the involvement of genes and cis-regulatory elements in adaptive traits [110, 114].

The practical advantages of cichlid fishes as a model system are numerous. In between, high-quality draft genomes are available for five African species *Oreochromis niloticus*, *Neolamprologus brichardi*, *Astatotilapia burtoni*, *Pundamilia nyererei and Maylandia zebra* [51]. Four more African species from lake Malawi, *Rhamphochromis esox*, *Melanochromis auratus*, *Mchenga conophoros* and *Labeotropheus fuelleborni* have been sequenced at low coverage [115], for 16 species from lake Victoria cichlids RAD-seq data is available [51, 116]. Also one Neotropical cichlid, *Amphilophus citrinellus* has a unpublished high-quality draft genome (Meyer Laboratory, University of Konstanz, Germany) and SNPs have been extensively described genome-wide in eleven more species of the Midas cichlid complex [117]. The availability of these genomic resources provides powerful means for identifying the genomic bases of their phenotypic traits. Recently, proof-of concept experiments were successfully executed for Tol2 Transgenesis [92, 93], Sleeping Beauty [118] and Crispr-Cas [119] (Tol2, Sleeping Beauty and Crispr-Cas in *Oreochromis niloticus*, Tol2 in *Astatotilapia burtoni* and *Amphilophus citrinellus)*. Also ChIP-seq was recently tested successfully in *Oreochromis niloticus* and *Pundamilia nyererei* (Kratochwil et al., unpublished results). Together with the possibility to perform QTL-crosses on these species [120] as well as to confirm mutations by PCR-screening of natural populations, cichlids might be a set of species in which evolutionary genetics could largely benefit from technologies established in model organisms such as transgenesis, genome engineering and ChIP-seq.

Conclusions and prospect

Ecologists and evolutionary geneticists are in an especially favourable position since they have access to a treasure trove of spectacular phenotypes in nature whose underlying genetic bases have yet to be unlocked. The discovery of the genetic bases of those phenotypes are likely to be, driven by four main factors: (1) by the availability of genomic and transcriptomic data, (2) by the availability of annotations for coding and non-coding regions, (3) by applicability of methods that facilitate linking genotypic and phenotypic variation and (4) and methods that functionally determine and validate the genetic basis of phenotypic traits. We are entering a new and exciting era in biology where we can expect great advances in understanding the mutations that drive the evolution of diversity.

Acknowledgments

CFK is supported by the Swiss National Science Foundation (P2BSP3_148629) and the EU FP7 Marie Curie Zukunftskolleg Incoming Fellowship Program, University of Konstanz (grant no. 291784). AM is funded by several grants of the Deutsche Forschungsgesellschaft (DFG) and advanced grant 297300 'Gen-adapt' by the European Research Council. The authors declare that they have no competing interests. Helen Gunter, Frederico Henning, Joost Woltering, Andreas Kautt, Paolo Franchini, Maggie Sefton are acknowledged for providing valuable comments on the manuscript. We appreciate the efforts of the editor and two anonymous reviewers for their constructive comments, which helped us to improve the manuscript. Lastly, we apologize to the colleagues whose work could not be cited because of space and reference restrictions.

References

1. **Schier AF**. 2013. Genomics: Zebrafish earns its stripes. *Nature* **496**: 443–4.
2. **Haffter P, Odenthal J, Mullins MC, Lin S**, et al. 1996. Mutations affecting pigmentation and shape of the adult zebrafish. *Dev Genes Evol* **206**: 260–76.
3. **van Eeden FJ, Granato M, Schach U, Brand M**, et al. 1996. Genetic analysis of fin formation in the zebrafish, *Danio rerio*. *Development* **123**: 255–62.
4. **Yelick PC, Schilling TF**. 2002. Molecular dissection of craniofacial development using zebrafish. *Crit Rev Oral Biol Med* **13**: 308–22.
5. **Lawson ND, Wolfe SA**. 2011. Forward and reverse genetic approaches for the analysis of vertebrate development in the zebrafish. *Dev Cell* **21**: 48–64.
6. **Caspary T**. 2010. Phenotype-driven mouse ENU mutagenesis screens. *Methods Enzymol* **477**: 313–27.
7. **Shapiro MD, Marks ME, Peichel CL, Blackman BK**, et al. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**: 717–23.
8. **Davis RH**. 2004. The age of model organisms. *Nat Rev Genet* **5**: 69–76.
9. **Grunwald DJ, Eisen JS**. 2002. Headwaters of the zebrafish—emergence of a new model vertebrate. *Nat Rev Genet* **3**: 717–24.
10. **Belknap JK, Hitzemann R, Crabbe JC, Phillips TJ**, et al. 2001. QTL analysis and genomewide mutagenesis in mice: Complementary genetic approaches to the dissection of complex traits. *Behav Genet* **31**: 5–15.
11. **Boffelli D, Nobrega MA, Rubin EM**. 2004. Comparative genomics at the vertebrate extremes. *Nat Rev Genet* **5**: 456–65.
12. 10K Genome Community of Scientists. 2009. Genome 10K: A proposal to obtain whole-genome sequence for 10,000 vertebrate species. *J Hered* **100**: 659–74.
13. i5K Consortium. 2013. The i5K initiative: Advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered* **104**: 595–600.
14. **Wang Z, Gerstein M, Snyder M**. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57–63.
15. **Park PJ**. 2009. ChIP-seq: Advantages and challenges of a maturing technology. *Nat Rev Genet* **10**: 669–80.
16. **Davey JW, Davey JL, Blaxter ML, Blaxter MW**. 2010. RADSeq: Next-generation population genetics. *Brief Funct Genomics* **9**: 416–23.
17. **Ekblom R, Galindo J**. 2010. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**: 1–15.
18. **Richter SH, Garner JP, Auer C, Kunert J**, et al. 2010. Systematic variation improves reproducibility of animal experiments. *Nat Methods* **7**: 167–8.
19. **Abzhanov A, Extavour CG, Groover A, Hodges SA**, et al. 2008. Are we there yet? Tracking the development of new model systems. *Trends Genet* **24**: 353–60.
20. **Hoekstra HE, Coyne JA**. 2007. The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* **61**: 995–1016.
21. **Wray GA**. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**: 206–16.
22. **Wittkopp PJ, Kalay G**. 2012. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**: 59–69.
23. **Wray NR, Yang J, Hayes BJ, Price AL**, et al. 2013. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* **14**: 507–15.
24. **de Laat W, Duboule D**. 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**: 499–506.
25. **Sanyal A, Lajoie BR, Jain G, Dekker J**. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**: 109–13.
26. **Benko S, Fantes JA, Amiel J, Kleinjan D-J**, et al. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet* **41**: 359–64.
27. **Pennacchio LA, Bickmore W, Dean A, Nobrega MA**, et al. 2013. Enhancers: Five essential questions. *Nat Rev Genet* **14**: 288–95.
28. **Raab JR, Kamakaka RT**. 2010. Insulators and promoters: Closer than we think. *Nat Rev Genet* **11**: 439–46.
29. **Flintoft L**. 2013. Gene regulation: Enhancing the hunt for enhancers. *Nat Rev Genet* **14**: 151.
30. **Villar D, Flicek P, Odom DT**. 2014. Evolution of transcription factor binding in metazoans—mechanisms and functional implications. *Nat Rev Genet* **15**: 221–33.
31. **Brawand D, Soumillon M, Necsulea A, Julien P**, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* **478**: 343–8.
32. **Stern DL**. 2013. The genetic causes of convergent evolution. *Nat Rev Genet* **14**: 751–64.
33. **Romero IG, Ruvinsky I, Gilad Y**. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13**: 505–16.
34. **Schmidt D, Wilson MD, Ballester B, Schwalie PC**, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**: 1036–40.
35. **Stefflova K, Thybert D, Wilson MD, Streeter I**, et al. 2013. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* **154**: 530–40.
36. **Chan YF, Marks ME, Jones FC, Villarreal G**, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302–5.
37. **Cretekos CJ, Wang Y, Green ED, Martin JF**, et al. 2008. Regulatory divergence modifies limb length between mammals. *Genes Dev* **22**: 141–51.
38. **Guerreiro I, Nunes A, Woltering JM, Casaca A**, et al. 2013. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci USA* **110**: 10682–6.
39. **Belting HG, Shashikant CS, Ruddle FH**. 1998. Modification of expression and cis-regulation of Hoxc8 in the evolution of diverged axial morphology. *Proc Natl Acad Sci USA* **95**: 2355–60.
40. **Guenther CA, Tasic B, Luo L, Bedell MA**, et al. 2014. A molecular basis for classic blond hair color in Europeans. *Nat Genet* **46**: 748–52.
41. **Miller CT, Beleza S, Pollen AA, Schluter D**, et al. 2007. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**: 1179–89.
42. **Freitas R, Gómez-Marín C, Wilson JM, Casares F**, et al. 2012. Hoxd13 contribution to the evolution of vertebrate appendages. *Dev Cell* **23**: 1219–29.

43. **Lopez-Rios J, Duchesne A, Speziale D, Andrey G**, et al. 2014. Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* **511**: 46–51.
44. **Roberts RB, Ser JR, Kocher TD**. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science* **326**: 998–1001.
45. **Cleves PA, Ellis NA, Jimenez MT, Nunez SM**, et al. 2014. Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. *Proc Natl Acad Sci USA* **111**: 13912–7.
46. **Schulte JE, O'Brien CS, Conte MA, O'Quin KE**, et al. 2014. Interspecific variation in rx1 expression controls opsin expression and causes visual system diversity in African cichlid fishes. *Mol Biol Evol* **31**: 2297–308.
47. **Schmidt D, Wilson MD, Spyrou C, Brown GD**, et al. 2009. ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**: 240–8.
48. **Necsulea A, Soumillon M, Warnefors M, Liechti A**, et al. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**: 635–40.
49. **Perry GH, Melsted P, Marioni JC, Wang Y**, et al. 2012. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res* **22**: 602–10.
50. **Heimberg AM, Sempere LF, Moy VN, Donoghue PCJ**, et al. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci USA* **105**: 2946–50.
51. **Brawand D, Wagner CE, Li YI, Malinsky M**, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**: 375–81.
52. **Heard E, Martienssen RA**. 2014. Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell* **157**: 95–109.
53. **Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M**, et al. 2013. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* **9**: e1003763.
54. **Chang AY-F, Liao B-Y**. 2012. DNA methylation rebalances gene dosage after mammalian gene duplications. *Mol Biol Evol* **29**: 133–44.
55. **Bannister AJ, Kouzarides T**. 2011. Regulation of chromatin by histone modifications. *Cell Res* **21**: 381–95.
56. **Zhou VW, Goren A, Bernstein BE**. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**: 7–18.
57. **Whitcomb SJ, Basu A, Allis CD, Bernstein E**. 2007. Polycomb group proteins: An evolutionary perspective. *Trends Genet* **23**: 494–502.
58. **Schwarz D, Varum S, Zemke M, Schöler A**, et al. 2014. Ezh2 is required for neural crest-derived cartilage and bone formation. *Development* **141**: 867–77.
59. **Di Meglio T, Kratochwil CF, Vilain N, Loche A**, et al. 2013. Ezh2 orchestrates topographic migration and connectivity of mouse precerebellar neurons. *Science* **339**: 204–7.
60. **Daxinger L, Whitelaw E**. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13**: 153–62.
61. **Attanasio C, Nord AS, Zhu Y, Blow MJ**, et al. 2013. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**: 1241006.
62. **Young NM, Hu D, Lainoff AJ, Smith FJ**, et al. 2014. Embryonic bauplans and the developmental origins of facial diversity and constraint. *Development* **141**: 1059–63.
63. **Schmidt D, Schwalie PC, Wilson MD, Ballester B**, et al. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**: 335–48.
64. **Ni X, Zhang YE, Nègre N, Chen S**, et al. 2012. Adaptive evolution and the birth of CTCF binding sites in the Drosophila genome. *PLoS Biol* **10**: e1001420.
65. **Bogdanovic O, Fernandez-Miñán A, Tena JJ, la Calle-Mustienes de E**, et al. 2013. The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos. *Methods* **62**: 207–15.
66. **Tena JJ, González-Aguilera C, Fernandez-Miñán A, Vázquez-Marín J**, et al. 2014. Comparative epigenomics in distantly related teleost species identifies conserved cis-regulatory nodes active during the vertebrate phylotypic period. *Genome Res* **24**: 1075–85.
67. **Arnold CD, Gerlach D, Stelzer C, Boryń ŁM**, et al. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**: 1074–7.
68. **Thurman RE, Rynes E, Humbert R, Vierstra J**, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.
69. **Giresi PG, Kim J, McDaniell RM, Iyer VR**, et al. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**: 877–85.
70. **Buenrostro JD, Giresi PG, Zaba LC, Chang HY**, et al. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–8.
71. **de Wit E, de Laat W**. 2012. A decade of 3C technologies: Insights into nuclear organization. *Genes Dev* **26**: 11–24.
72. **Mackay TFC, Stone EA, Ayroles JF**. 2009. The genetics of quantitative traits: Challenges and prospects. *Nat Rev Genet* **10**: 565–77.
73. **Sun W, Hu Y**. 2013. eQTL Mapping using RNA-seq data. *Stat Biosci* **5**: 198–219.
74. **Visscher PM, Brown MA, McCarthy MI, Yang J**. 2012. Five years of GWAS discovery. *Am J Hum Genet* **90**: 7–24.
75. **Ha N-T, Freytag S, Bickeboeller H**. 2014. Coverage and efficiency in current SNP chips. *Eur J Hum Genet* **22**: 1124–30.
76. **Zhang H, Wang Z, Wang S, Li H**. 2012. Progress of genome wide association study in domestic animals. *J Anim Sci Biotechnol* **3**: 26.
77. **Jones FC, Chan YF, Schmutz J, Grimwood J**, et al. 2012. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol* **22**: 83–90.
78. **Jones FC, Grabherr MG, Chan YF, Russell P**, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.
79. **Shapiro MD, Kronenberg Z, Li C, Domyan ET**, et al. 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**: 1063–7.
80. **Peterson BK, Weber JN, Kay EH, Fisher HS**, et al. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**: e37135.
81. **Ellegren H**. 2014. Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol* **29**: 51–63.
82. **Kim SY, Pritchard JK**. 2007. Adaptive evolution of conserved noncoding elements in mammals. *PLoS Genet* **3**: 1572–86.
83. **Hiller M, Schaar BT, Bejerano G**. 2012. Hundreds of conserved non-coding genomic regions are independently lost in mammals. *Nucleic Acids Res* **40**: 11463–76.
84. **Kawakami K**. 2007. Tol2: A versatile gene transfer vector in vertebrates. *Genome Biol* **8** (Suppl 1): S7.
85. **Gaj T, Gersbach CA, Barbas CF**. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**: 397–405.
86. **Perez-Pinera P, Ousterout DG, Gersbach CA**. 2012. Advances in targeted genome editing. *Curr Opin Chem Biol* **16**: 268–77.
87. **Davis J, Maillet M, Miano JM, Molkentin JD**. 2012. Lost in transgenesis: A user's guide for genetically manipulating the mouse in cardiac research. *Circ Res* **111**: 761–77.
88. **Clark KJ, Urban MD, Skuster KJ, Ekker SC**. 2011. Transgenic zebrafish using transposable elements. *Methods Cell Biol* **104**: 137–49.
89. **Ittner LM, Götz J**. 2007. Pronuclear injection for the production of transgenic mice. *Nat Protoc* **2**: 1206–15.
90. **Grabher C, Wittbrodt J**. 2007. Meganuclease and transposon mediated transgenesis in medaka. *Genome Biol* **8** (Suppl 1): S10.
91. **Balciunas D, Davidson AE, Sivasubbu S, Hermanson SB**, et al. 2004. Enhancer trapping in zebrafish using the sleeping beauty transposon. *BMC Genomics* **5**: 62.
92. **Fujimura K, Kocher TD**. 2011. Tol2-mediated transgenesis in tilapia (*Oreochromis niloticus*). *Aquaculture* **319**: 342–6.
93. **Juntti SA, Hu CK, Fernald RD**. 2013. Tol2-mediated generation of a transgenic haplochromine cichlid, *Astatotilapia burtoni*. *PLoS One* **8**: e77647.
94. **Valenzano DR, Sharp S, Brunet A**. 2011. Transposon-mediated transgenesis in the short-lived African killifish *Nothobranchius furzeri*, a vertebrate model for aging. *G3 (Bethesda)* **1**: 531–8.
95. **Urnov FD, Rebar EJ, Holmes MC, Zhang HS**, et al. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**: 636–46.
96. **Joung JK, Sander JD**. 2013. TALENs: A widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**: 49–55.
97. **Cong L, Ran FA, Cox D, Lin S**, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–23.
98. **Niu Y, Shen B, Cui Y, Chen Y**, et al. 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156**: 836–43.

99. **Fu Y, Foden JA, Khayter C, Maeder ML**, et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**: 822–6.
100. **Jao L-E, Wente SR, Chen W**. 2013. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci USA* **110**: 13904–9.
101. **Wang H, Yang H, Shivalila CS, Dawlaty MM**, et al. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**: 910–8.
102. **Taylor JS, Braasch I, Frickey T, Meyer A**, et al. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. *Genome Res* **13**: 382–90.
103. **Mackay TFC**. 2014. Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nat Rev Genet* **15**: 22–33.
104. **Auer TO, Duroure K, De Cian A, Concordet J-P**, et al. 2014. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res* **24**: 142–53.
105. **Gilbert LA, Larson MH, Morsut L, Liu Z**, et al. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**: 442–51.
106. **Kratochwil CF, Rijli FM**. 2014. The Cre/Lox system to assess the development of the mouse brain. *Methods Mol Biol* **1082**: 295–313.
107. **Stinchcombe JR, Hoekstra HE**. 2008. Combining population genomics quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity (Edinb)* **100**: 158–70.
108. **Xiao A, Wang Z, Hu Y, Wu Y**, et al. 2013. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res* **41**: e141.
109. **Braasch I, Peterson SM, Desvignes T, McCluskey BM**, et al. A new model army: Emerging fish models to study the genomics of vertebrate Evo-Devo. *J Exp Zool B Mol Dev Evol*, in press, doi: 10.1002/jez.b.22589.
110. **Henning F, Meyer A**. 2014. The evolutionary genomics of cichlid fishes: Explosive speciation and adaptation in the postgenomic era. *Annu Rev Genomics Hum Genet* **15**: 417–41.
111. **Kocher TD**. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nat Rev Genet* **5**: 288–98.
112. **Salzburger W, Meyer A**. 2004. The species flocks of East African cichlid fishes: Recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften* **91**: 277–90.
113. **Barluenga M, Stölting KN, Salzburger W, Muschick M**, et al. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* **439**: 719–23.
114. **Kuraku S, Meyer A**. 2008. Genomic analysis of cichlid fish 'natural mutants'. *Curr Opin Genet Dev* **18**: 551–8.
115. **Loh Y-HE, Katz LS, Mims MC, Kocher TD**, et al. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. *Genome Biol* **9**: R113.
116. **Wagner CE, Keller I, Wittwer S, Selz OM**, et al. 2012. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol Ecol* **22**: 787–98.
117. **Elmer KR, Fan S, Kusche H, Spreitzer ML**, et al. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat Commun* **5**: 5168.
118. **He X, Li J, Long Y, Song G**, et al. 2013. Gene transfer and mutagenesis mediated by Sleeping Beauty transposon in Nile tilapia (*Oreochromis niloticus*). *Transgenic Res* **22**: 913–24.
119. **Li M, Yang H, Zhao J, Fang L**, et al. 2014. Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics* **197**: 591–9.
120. **Henning F, Lee HJ, Franchini P, Meyer A**. 2014. Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: Benefits and pitfalls of using RAD markers for dense linkage mapping. *Mol Ecol* **23**: 5224–40.