Positive Darwinian Selection Drives the Evolution of the Morphology-Related Gene, EPCAM, in Particularly Species-Rich Lineages of African Cichlid Fishes

Shaohua Fan · Kathryn R. Elmer · Axel Meyer

Abstract The study of genetic evolution within the context of adaptive radiations offers insights to genes and selection pressures that result in rapid morphological change. Cichlid fishes are very species-rich and variable in coloration, behavior, and morphology, and so provide a classical model system for studying the genetics of adaptive radiation. In this study, we researched the evolution of the epithelial cell adhesion molecule (EPCAM), a candidate gene for the adaptive evolution of morphology broadly, and skin development specifically, in fishes. We compared EPCAM gene sequences from a rapidly speciating African cichlid lineage (the haplochromines), a species-poor African lineage (Nile tilapia *Oreochromis niloticus*), and a very young adaptive radiation in the Neotropics (sympatric crater lake Midas cichlids, *Amphilophus* sp.). Our results, based on a hierarchy of evolutionary analyses of nucleotide substitution, demonstrate that there are different selection pressures on the EPCAM gene among the cichlid lineages. Several waves of positive natural selection were identified not only on the terminal branches, but also on ancestral branches. Interestingly, significant positive or directional selection was found in the haplochromine cichlids only but not the comparatively species-poor tilapia lineage. We hypothesize that the strong signal of selection in the ancestral African cichlid lineage coincided with the transition from riverine to lacustrine habitat. The two neotropical species for which we collected new sequence data were invariant in the EPCAM locus. Our results suggest that functional changes promoted by positive Darwinian selection are widespread in the EPCAM gene during African cichlid evolution.

Keywords Positive Darwinian selection · Epithelial cell adhesion molecule · Adaptive radiation

S. Fan · K. R. Elmer · A. Meyer (✉)
Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätstrasse 10, 78457 Constance, Germany
e-mail: axel.meyer@uni-konstanz.de

Introduction

Adaptive radiation—the rapid generation of exceptional adaptive diversification within a lineage—is considered one of the central explanations for the diversity of life on earth (Schluter 2000; Gavrilets and Losos 2009). Ecological opportunities play an important role in adaptive radiation, especially when an ancestral species invaded an unutilized niche early in the radiation. The species then diversifies, as new ecological niches provide evolutionary opportunities with relaxed selection or strong disruptive or directional selection (Schluter 2000; Eldredge et al. 2005; Kapralov and Filatov 2006). It is well accepted that divergent natural selection is the primary mechanism of adaptive evolution, and this can be evidenced at the molecular level (Schluter 2001, 2009; Rogers and Bernatchez 2007). Well-known empirical examples of adaptive radiation include: Darwin's finches, Hawaiian honeycreepers (Schluter 2000), cichlid fishes in East African great lakes (Salzburger et al. 2007) and cichlid fishes in the Nicaraguan crater lakes (Elmer et al. 2010b). The genetic basis of adaptive radiation greatly attracts the interest of biologists (Barrier et al. 2001; Kapralov and Filatov 2006; Jeukens et al. 2009) because, among other reasons, it can give insight to the genetic basis of evolution with reduced phylogenetic and geographical noise.

Cichlid fishes are an ideal model system to study the genetic basis of adaptive radiation. First, species richness is extremely high compared with other adaptive radiations; in the great lakes of East Africa almost 2000 species have evolved from a common ancestor within the past few million years (Meyer et al. 1990; Salzburger et al. 2002; Verheyen et al. 2003; Elmer et al. 2009). Second, the age of the adaptive species flocks range among lakes, allowing for comparisons across different temporal scales. The oldest extant radiation began more than 2 million years ago in Lake Tanganyika (Meyer et al. 1990; Meyer 1993; Salzburger et al. 2005; Seehausen 2006). In Lake Victoria, more than 500 endemic species, the renowned 'superflock' of cichlid fishes, evolved within the past 100,000 years (Meyer et al. 1990; Verheyen et al. 2003; Abila et al. 2004; Elmer et al. 2009). Third, a great amount of phenotypic and morphological diversity has arisen with adaptation to new ecological niches. Specifically, cichlids have repeatedly evolved parallel or convergent body shapes and colorations suited to similar but often independent environments (Kocher et al. 1993; Ruber et al. 1999; Koblmuller et al. 2004; Hulsey et al. 2008; Elmer et al. 2010b). Yet many cichlid species remain extremely genetically similar (Sturmbauer and Meyer 1992). Thus, they can be regarded as "natural mutants" (Meyer et al. 1993; Kuraku and Meyer 2008). Recent studies have identified multiple genes that contribute to the adaptive radiation of cichlids by using non-candidate gene (Gerrard and Meyer 2007; Elmer et al. 2010a) and candidate gene approaches to seek the genetic basis of traits related to the visual sensory system (Sugawara et al. 2002; Spady et al. 2005; Seehausen et al. 2008; Hofmann et al. 2009), parental care behavior (Summers and Zhu 2008), reproductive evolution (Gerrard and Meyer 2007), coloration (Salzburger et al. 2007) and jaw morphology (Terai et al. 2002b; Kijimoto et al. 2005; Albertson and Kocher 2006). Fourth, a growing amount of molecular data from African cichlids (Watanabe et al. 2004; Kobayashi et al. 2009; Salzburger et al. 2008; Lee et al. 2010) and neotropical cichlids (Elmer et al. 2010a) are available by which to investigate and examine the molecular evolution of genes that contribute to the adaptive radiation of cichlid fishes. The recent availability of molecular data from neotropical cichlids (Fan, Elmer and Meyer, in prep.) provides reliable outgroup information (Zardoya et al. 1996; Farias et al. 1999) for the tests of positive Darwinian selection detection in African cichlids, since the short divergence time between African and neotropical cichlids is relatively short (85.1 and 40.5 million years, based on gondwana fragmentation and fossil record) (Genner et al. 2007) compared to the evolutionary divergence (>100 million years) between cichlid fishes and other model system fishes (Steinke et al. 2006).

The African cichlid adaptive radiation is famous for the extremely variable morphologies that have arisen and diversified (Meyer 1993). Candidate genes inferred from zebrafish have proven successful at elucidating patterns of molecular evolution and adaptive diversification in cichlid fishes (e.g. Terai et al. 2003; Sugie et al. 2004; Salzburger et al. 2007). In the present study, we analyze the evolution of a candidate gene related to morphological changes and skin development in fishes: the epithelial cell adhesion molecule (EPCAM or CD326). EPCAM belongs to the cell adhesion molecular (CAM) family (Baeuerle and Gires 2007; Trzpis et al. 2007), which not only play a role in cell adhesion, but also in cell proliferation, migration, and differentiation. These processes are known to be fundamental in morphogenesis (Trzpis et al. 2008a). For example, a recent study indicates that the mutant in this gene display defects in both epithelial morphologies and integrity in zebrafish embryo development (Slanchev et al. 2009) though its role in ocular epithelial cells remains unknown (Forrester et al. 2010). We tested the role of Darwinian selection in the molecular evolution of presumably functional changes in the EPCAM gene across diverse lineages of cichlid fishes, based on the ratio of nonsynonymous to synonymous mutations (K_a/K_s ratio, or ω). The ratio of K_a/K_s is a strong signal of positive, or directional, Darwinian evolution and has been used extensively to identify and analyze the role of selection in gene evolution in fishes (e.g. Dann et al. 2004; Gerrard and Meyer 2007; Elmer et al. 2010a).

Using a candidate gene approach based on EPCAM sequences from neotropical and African cichlid lineages, we considered the following hierarchy of hypotheses. First, given that African great lake cichlids show extremely high morphological diversity, we tested for evidence of positive selection on the molecular evolution of EPCAM in the African clade as a whole. Given that this analysis indicated a strong positive signal of positive selection on EPCAM, we then sought in more detail the timing of this selection and the nucleotide regions under selection. Thus, second, we tested the non-neutral molecular evolution in each of the derived lineages of African cichlids (recent evolution) and the ancestral lineages (earlier evolutionary pressures). If EPCAM was only recently under positive selection then we should identify patterns in derived lineages. However, if EPCAM was involved in earlier diversification, then we predicted to find positive selection on the ancestral sequences.

Materials and Methods

The full length protein-coding nucleotide sequence of the EPCAM gene from the stickleback (*Gasterosteus aculeatus*) genome (transcript ID: ENSGACT00000003469; 921 bp) was downloaded from the Ensembl database

(Hubbard et al. 2002) and used as a query sequence to search the NCBI EST database (Boguski et al. 1993), with the BLAST (Altschul et al. 1990) search restricted to the family Cichlidae. The EPCAM gene is composed of nine exons according to the Ensembl database (Hubbard et al. 2002). The EST sequence of EPCAM was downloaded for each cichlid species for which it was available and a single EST sequence per species was assembled by CAP3 (Huang and Madan 1999) using default parameters when multiple ESTs were available for one species. EPCAM sequences from two neotropical crater lake cichlid species *Amphilophus amarillo* and *Amphilophus sagittae* (unpublished data) were included as the outgroup sequences.

The assembled EST sequences were aligned in ClustalX using default parameters (Thompson et al. 2002). After alignment, we identified the open reading frame (ORF) in the assembled cichlid ESTs by comparing with EPCAM protein-coding sequence from stickleback. Non-coding regions were excluded from further analysis. To exclude estimation bias caused by partial sequences, we used only full length protein-coding nucleotide sequences of EPCAM in the further analysis. Based on incomplete tilapia genome sequences that are publicly available (Kocher, unpublished data http://cichlid.umd.edu/blast/, database: RRS5KB-SCAFF.e60.c0.p60), only one copy of the EPCAM gene can be found in cichlids so our analyses will not be complicated by the asymmetrical evolution of paralogues.

The best molecular substitution model for EPCAM was selected by a hierarchical likelihood ratio test after testing in Modeltest version 3.7 (Posada and Crandall 1998). We reconstructed a phylogenetic tree using a maximum likelihood approach with PhyML version 3.0 (Guindon et al. 2009) given an HKY model (Hasegawa et al. 1985) of molecular evolution and gamma-distributed rate variation among sites. The robustness of the topology was tested with the highest likelihood using 100 bootstrap cycles. Trees were visualized in Figtree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

The improved branch-site model (Zhang et al. 2005) by codeml in the PAML package Version 4.4 (Yang 2007) was employed for the positive selection analysis at broad groupings of lineages. The codon-based model is more sensitive to detect positive Darwinian selection on particular lineage comparing with other average sites methods (Yang and Nielsen 2002). We first compared the signal of positive selection between different lineages: (i) African cichlids versus neotropical cichlids and (ii) Lake Victoria 'superflock' *Haplochromis* species versus all other cichlids. The branches of interest were assigned as foreground branches in the branch-site model.

We conducted further analyses to detect whether there is a signal of positive selection in different stages of cichlid evolution: in the derived or ancestral lineages. We used the branch-site model to separate the branches in a phylogenetic tree into two classes: foreground branches (i.e., branches of interest) and background branches. The user-specified foreground branch is tested for a signal of positive selection in comparison with a null model that assumes that signals of negative (class 0) and neutral selection (class 1) on codons. The alternative model is that the foreground branches will show a signal of positive selection relative to the background branches showing negative or neutral selection. To avoid local optima in the maximum likelihood integrations, three different initial omega values ($\omega = 0.5$, $\omega = 1$, $\omega = 2$) were used to evaluate the parameter estimation. The improved branch-site model assumes positive natural selection can only involve a small number of codons, which is a robust method to detect positive selection in recent diverged species (Zhang et al. 2005). The likelihood ratio test (LRT) was used to determine the statistical significance of the signal of positive selection by comparing the likelihood difference of a gene sequence under the null and alternate models (Yang and Nielsen 2002; Zhang et al. 2005). The LRT method has been shown to be both accurate and robust in simulation studies (Anisimova et al. 2001).

Given that the signal of positive selection can be affected by only a small portion of the codons in a gene (Golding and Dean 1998), positively selected sites were inferred using a Bayes empirical Bayes (BEB) method, which reduces sampling error in small data sets (Yang et al. 2005) and therefore improves accuracy and robustness for our analysis. BEB probabilities greater than 0.75 at single codon positions were considered significant values.

Results

Cichlid EMCAP Genes and Phylogeny Construction

The dataset of cichlid ESTs currently publicly available on NCBI is mainly composed of sequences from African cichlid species. Using stickleback full length cDNA of the EPCAM gene as query sequence, we assembled EPCAM sequences from the African basal haplochromine *Astatotilapia burtoni*, which is a riverine and lacustrine species (Salzburger et al. 2008), the Nile tilapia *Oreochromis niloticus*, which is an African riverine habitat species that has recently invaded many lake habitats (Lee et al. 2010), and three African Lake Victoria 'superflock' cichlids *Haplochromis chilotes*, *Haplochromis* sp. 'Matumbi hunter', and *Haplochromis* sp. 'red tail sheller' (Watanabe et al. 2004; Kobayashi et al. 2009) (Supplementary Table 1). Neotropical representatives were two species of Midas cichlid, *Amphilophus amarillo* and *Amphilophus sagittae*

from Nicaraguan crater lake Xiloá (accession number: JN391522, JN391523). The assembled sequence from *A. burtoni* was excluded from further analyses since it did not reach our length criterion (451 bp in the coding region) and would therefore bias the selection analyses.

Maximum likelihood phylogenetic reconstruction of the EPCAM gene produced a topology with high supporting values at branch nodes (Fig. 1). The topology was in agreement with current understanding about the evolutionary relationships and geographical distribution of cichlids (Meyer et al. 1990; Genner et al. 2007). Two divergent clades were identified: one for the neotropical cichlids and the other for the African cichlids. Within the African clade, tilapia is basal to the monophyletic tribe Haplochromini.

Test of Positive Selection in Major Cichlid Clades

First we tested for positive selection on the EPCAM gene between African (*Oreochromis* + *Haplochromis*) and neotropical (*Amphilophus*) cichlids using the modified branch-site model (Zhang et al. 2005). Significant positive selection was detected in the evolution of the EPCAM gene sequence in the African cichlid clade ($\omega_F = 6.08$, $P < 0.001$) (Table 1). EPCAM sequences were identical in both species of *Amphilophus* and therefore showed no sign of selection within the neotropical clade.

Second, we tested for a signal of positive selection on the EPCAM sequences exclusively from the clade of three *Haplochromis* cichlids from Lake Victoria and identified a strong signal of positive selection ($\omega_F = 37.03$, $P < 0.001$, Table 1) on the EPCAM gene in the haplochromine lineage.
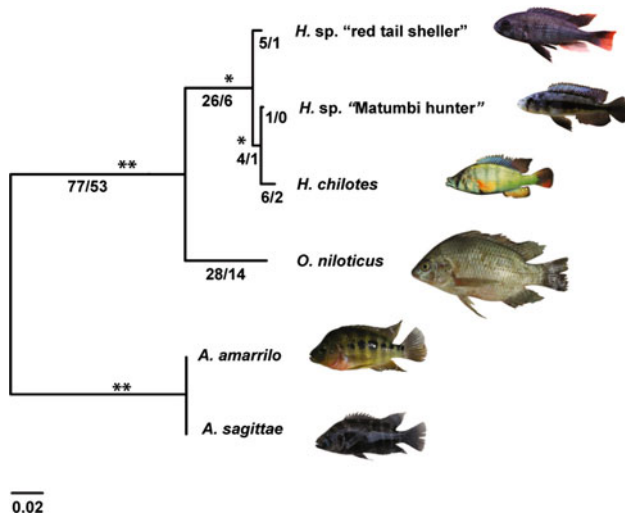
Fig. 1 The phylogeny of African and neotropical cichlids based on the EPCAM gene. The symbol ** and * indicate the bootstrap values equal to 100 or at least 84. The numbers of non-synonymous/synonymous mutations are listed in **bold** (not *italic*) under the branches. The ancestral sequences of all internal nodes were build under the M8 (beta & w) model in PAML

Test of Positive Selection in Derived Cichlid Lineages

Each cichlid species was specified as a foreground branch to test for positive selection separately in each derived lineage. We found significant positive selection on the *Haplochromis* sp. 'red tail sheller' branch ($\omega_F = 999$, $P < 0.001$, Table 1). No signal of positive selection was identified in *Haplochromis* sp. 'Matumbi hunter' or *Haplochromis chilotes* (Table 1). The extreme value of $\omega_F = 999$ indicates rare synonymous substitutions in the foreground branch.

The branch-site model predicted positively selected amino acid sites in *Haplochromis* sp. 'red tail sheller' at positions 72 and 198 (BEB probability > 0.75) which resulted the substitutions from Lysine (K) to Arginine (R) in position 72 and from Proline (P) to Glutamic acid (E) in position 192. To exclude the possibility of an inflated prediction of positive selection that can be caused by polymorphisms in very recently diverged populations (Peterson and Masel 2009), we checked the SNP information for the positively selected site. The novel alleles showing a sign of selection are unique to *Haplochromis* sp. 'red tail sheller' (data not show), which indicates it is not due to shared polymorphism.

Test of Positive Selection on Ancestral African Lineages

We tested for a signal of positive selection in the ancestral sequences of each hierarchical clade of the African cichlids, i.e., (i) ancestor to *H. chilotes* + *Haplochromis* sp. 'Matumbi hunter', (ii) ancestor to all *Haplochromis*, and (iii) ancestor to all African cichlids (*Oreochromis* + *Haplochromis)* (Fig. 1). All three ancestral sequences showed a statistically significant signal of positive selection, though the strength of selection differs across hierarchies (Table 1). The selection pressure in the ancestor sequence of all *Haplochromis* is stronger than that of the evolutionarily deeper grouping of all African cichlids ($\omega_F = 75.84$ vs. 16.34). An ω_F value of 999 (e.g. at the ancestral sequence to *Haplochromis* sp. 'Matumbi hunter' and *H. chilotes*) is caused by sequence divergence comprised only of non-synonymous mutations (i.e., $K_s = 0$).

Interestingly, different, non-overlapping codon sites showed significant signals of positive selection across the three hierarchies of ancestral sequences (Fig. 2). This may reflect that the nucleotide substitutions were driven by temporally and molecularly different selection pressures and, therefore, show different evolutionarily independent signals.

Unfortunately, there is no structural information for the EPCAM gene available in the current protein data bank (Berman et al. 2000). Therefore, we predicted the protein

Table 1 Parameter estimations of the branch-site model for three evolutionary hierarchies of cichlid fishes: total clades, derived lineages, and ancestral lineages

	Tested branch	ω_F	LRT	P value
Clade	All African cichlids	6.08	22.08	<0.001
	Haplochromis	37.03	22.26	<0.001
Derived	*Oreochromis niloticus*	23.20	5.34	ns
	H. sp. 'red tail sheller'	999	14.78	<0.001
	H. sp. 'Matumbi hunter'	1.05	0	ns
	H. chilotes	999	1.43	ns
Ancestral	African ancestor	16.34	12.21	<0.001
	Haplochromis ancestor	75.84	22.26	<0.001
	Ancestor of *H.* sp. 'Matumbi hunter' and *H. chilotes*	999	10.32	<0.01

The absolute numbers of non-synonymous and synonymous mutations for each branch are listed in Fig. 1

$\omega_F = K_a/K_s$ ratio of the foreground branch; LRT = likelihood ratio test. P values > 0.01 are listed as non-significant (ns). $\omega_F = 999$ indicates the presence of non-synonymous mutations in the absence of synonymous mutations (i.e., K_a/K_s is undefined because $K_s = 0$)
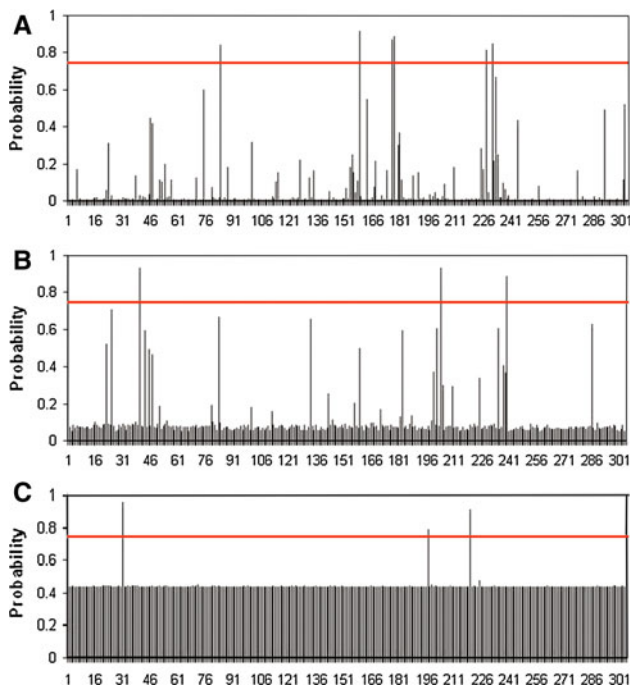
Fig. 2 Sites showing a signal of positive selection in the ancestral branch of all African cichlids (**a**), all Haplochromine cichlids (**b**), and *Haplochromis* sp 'Matumbi hunter' and *H. chilotes* (**c**). The position stands for the codon position on the cDNA sequence. The *horizontal* line indicates the Bayes empirical Bayes probability of positively selected sites equal to 0.75

structure prediction using the interproscan software (Quevillon et al. 2005) online tool (http://www.ebi.ac.uk/Tools/InterProScan/). The thyroglobulin 1 domain was identified (E value = 5.1e−14) and predicted to span from codon position 89–137, a common domain in the EPCAM family (Baeuerle and Gires 2007). However, all of the positively selected nucleotide sites that were determined in

our analysis are located outside of this domain and were without significant domain search hits.

Discussion

In this study, we analyzed the evolution of the EPCAM gene across diverse cichlid lineages. Because of its importance in fish morphogenesis (Slanchev et al. 2009), we considered EPCAM to be a relevant candidate for being involved in the exceptional morphological and phenotypic diversifications of cichlid fishes. We identified a strong signal of positive direction selection in the evolution of the African cichlid lineage, especially that of the Lake Victoria 'superflock' of cichlid fishes, the *Haplochromis* (Table 1; Fig. 1). Selection was less strong in the more species depauperate tilapia lineage. The fact that EPCAM was not always found to be under selection in all lineages indicates it is not obligatory that the gene show such a signal through its evolution.

Local adaptation is one of the driving forces in the evolution of cichlid species (Kocher 2004). One prediction that stems from this is that different, lineage-specific directional selection pressures should be identified on the genes involved in local adaptation. The more than 2,000 cichlid species found in East Africa display a great variation in phenotypic attributes such as body shape, jaw shape, and coloration. Thus, we hypothesized that a gene known to play an important role in morphological variation in other fishes (Slanchev et al. 2009) may a target of positive natural selection during cichlid evolution.

To test this hypothesis, we first compared the evolution of the EPCAM gene between representative African and neotropical cichlids and identified that the EPCAM gene is under significant and strong positive selection in the

African cichlids as a whole (Fig. 1; Table 1). Then we examined in more detail the Lake Victoria 'superflock' of *Haplochromis* cichlids, that are renowned for their spectacular adaptive radiation (Meyer et al. 1990; Verheyen et al. 2003; Salzburger and Meyer 2004) and, indeed, identified a significant signature of positive selection in the EPCAM gene in this group ($\omega_F = 37.03$, $P < 0.001$).

Second, we teased apart the hierarchical level at which Darwinian selection is evidenced in the evolution of the African cichlids. In the derived lineages of four species, that are representative of extremely rapidly evolved Lake Victoria cichlid "super flock" only the species *Haplochromis* sp. 'red tail sheller' showed significant signatures of positive selection in the evolution on EPCAM sequence ($\omega_F = 999$, $P < 0.001$). The extremely high signal of selection pressure (undefined or "999") indicates that only non-synonymous mutations and no synonymous mutations were found in this branch. This small genetic difference could result in considerable phenotypic changes, for example by a single amino acid replacement between closely related species (Hoekstra et al. 2006) or may be a signal of relaxed purifying selection during adaptive evolution (Hughes 2007). We propose that positive selection, rather than relaxed purifying selection, more likely drove the fixation that is only found at these non-synonymous sites based on two reasons. First, the function and amino acid sequences of the EPCAM gene are conserved across vertebrate evolution (Trzpis et al. 2008b; Slanchev et al. 2009). Second, if relaxed purifying selection is caused by small population size (Ohta 1995), we would expect to observe an increased genetic diversity in the genome wide scale. However, other researches show that ESTs and multiple genomic nucleotide markers show very limited genetic differentiations among Lake Victoria cichlid fishes (Elmer et al. 2009; Kobayashi et al. 2009).

We did not detect a signature of positive selection in the other two Lake Victoria cichlids, *H. chilotes* and *Haplochromis* sp. '*Matumbi hunter*' though we did detect selection in the common ancestral branch. Two possible explanations for this could be either the insufficient genetic variability between these recently diverged species, or a signal of purifying selection on derived branches. A similar change of selection regime, from positive selection on the ancestral branch to purifying selection on the derived branch was, for example, shown in the evolution of gut lysozymes in the colobine lineage of primates (Messier and Stewart 1997).

In the processes of speciation in cichlids, it is proposed that the most major morphological changes occur at the initial stages of habitat adaptation, often followed by subsequent subtler evolution into more finely partitioned ecological niches and color divergence based on sexual selection (Kocher 2004). To test whether this model is correct, it is important to determine whether positive selection influenced the sequences earlier than in the derived, extant lineages. Three waves of positive selection were detected in the ancestral sequences of the African cichlids ($\omega_F = 16.34$, $P < 0.001$), *Haplochromis* cichlids family in Lake Victoria ($\omega_F = 75.84$, $P < 0.001$) and *Haplochromis* sp. 'Matumbi hunter' and *H. chilotes* ($\omega_F = 999$, $P < 0.01$). These results suggest that functional changes in the EPCAM gene happened at different and various stages during African cichlid evolution and thus the signal of selection pressures differs along the phylogeny. The strongest positive selection was identified in the ancestral lineage leading to the extraordinarily diverse lineage of haplochromine cichlids, which could be the result of adaptation to the vast set of new open lake niche once the colonization of Lakes Victoria and Malawi from riverine haplochromine lineages took place (Salzburger et al. 2005; Elmer et al. 2009). The positively selected sites among three ancestral sequences were not overlapping, which may reflect the divergent selection caused by adaptations to different environments (Fig. 2a vs. Fig. 2b vs. Fig. 2c).

Cichlids from the African Great Lakes have been found to have accelerated rates of non-synonymous substitution in a number of candidate genes relative to riverine sister lineages. Comparative analyses and functional analyses tended to indicate that this molecular pattern is the result of stronger natural selection, including sometimes sexual selection pressures, in the lacustrine habitat. For example, a high rate of non-synonymous substitution was also identified in the molecular evolution of the *hagoromo* gene in haplochromine cichlids from the endemic adaptive radiations of the Great Lake African cichlids. This gene—also a candidate selected based on its role in zebrafish development—might play an important role in pigmentation, and its evolution might be driven by sexual selection (Terai et al. 2002a). This was further confirmed with analyses that showed that cichlid lineages differ with respect to the complexity of *hagomoro* gene alternative splicing, and that this splicing complexity correlates with lineage speciation rates (Terai et al. 2003). A similar pattern was found with the pigmentation candidate gene *mitf* (Santini and Meyer, unpublished data; Sugie et al. 2004). Genes related to vision are also known to be under positive selection in lacustrine compared to riverine East African haplochromine cichlid species (e.g., Sugawara et al. 2002). Thus, our inferences from the analyses of EPCAM are in agreement with findings from other genes relevant to adaptive phenotypes in the adaptive radiation of cichlids. It is unclear why the signals of natural selection differ among the three extant haplochromines in our study (Table 1), but it may be that their different ecological niches play a role. Further research on the evolution of EPCAM sequence in species

from different lineages and habitats will be required to clarify this pattern.

Theoretical analyses have suggested that K_a/K_s estimates may be elevated in recently diverged species due to incomplete sorting of ancestral polymorphism (Peterson and Masel 2009). That most of the derived African lineages in our study do not show significant positive selection, while the inferred ancestral sequence does, suggests this phenomenon apparently does not play a strong role in our analyses. Furthermore, we found that positively selected alleles in extant species apparently species-specific.

Our current research is limited to EPCAM sequences only from Lake Victoria, the youngest adaptive radiation of the eastern Africa great lakes and rooted with recently determined neotropical cichlid EPCAM sequences. Adding more sequences of this candidate gene from the African lakes that harbor older cichlid species flocks, such as Lake Malawi and Lake Tanganyika, would provide a more comprehensive view of the role of EPCAM during the formation of the adaptive radiations of cichlids. Since cichlid fishes have evolved parallel phenotypes presumably under similar selective regimes due to ecological speciation, one could test the prediction that natural selection drives the same advantageous alleles to fixation independently in closely related species that share similar environments (Schluter 2009). The fixation of different positively selected nucleotide sites within the EPCAM gene suggests that it could be employed as candidate gene to test whether parallel morphologies in East Africa great lake cichlids (Kocher et al. 1993) are caused by parallel, independent substitutions in morphologically relevant genes.

References

Abila R, Barluenga M, Engelken J, Meyer A, Salzburger W (2004) Population-structure and genetic diversity in a haplochromine fish cichlid of a satellite lake of Lake Victoria. Mol Ecol 13:2589–2602

Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. Heredity 97:211–221

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Baeuerle PA, Gires O (2007) EpCAM (CD326) finding its role in cancer. Br J Cancer 96:417–423

Barrier M, Robichaux RH, Purugganan MD (2001) Accelerated regulatory gene evolution in an adaptive radiation. Proc Natl Acad Sci USA 98:10208–10213

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Boguski MS, Lowe TMJ, Tolstoshev CM (1993) Dbest - database for expressed sequence tags. Nat Genet 4:332–333

Dann SG, Allison WT, Levin DB, Taylor JS, Hawryshyn CW (2004) Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in Oncorhynchus. J Mol Evol 58:400–412

Eldredge N, Thompson JN, Brakefield PM, Gavrilets S, Jablonski D, Jackson JBC, Lenski RE, Lieberman BS, McPeek MA, Miller W (2005) The dynamics of evolutionary stasis. Paleobiology 31:133–145

Elmer KR, Reggio C, Wirth T, Verheyen E, Salzburger W, Meyer A (2009) Pleistocene desiccation in East Africa bottlenecked but did not extirpate the adaptive radiation of Lake Victoria haplochromine cichlid fishes. Proc Natl Acad Sci USA 106: 13404–13409

Elmer KR, Fan S, Gunter HM, Jones JC, Boekhoff S, Kuraku S, Meyer A (2010a) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Mol Ecol 19:197–211

Elmer KR, Kusche H, Lehtonen TK, Meyer A (2010b) Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. Philos Trans R Soc Lond B Biol Sci 365: 1763–1782

Farias IP, Orti G, Sampaio I, Schneider H, Meyer A (1999) Mitochondrial DNA phylogeny of the family Cichlidae: monophyly and fast molecular evolution of the neotropical assemblage. J Mol Evol 48:703–711

Forrester JV, Xu HP, Kuffova L, Dick AD, McMenamin PG (2010) Dendritic cell physiology and function in the eye. Immunol Rev 234:282–304

Gavrilets S, Losos JB (2009) Adaptive radiation: contrasting theory with data. Science 323:732–737

Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF (2007) Age of cichlids: new dates for ancient lake fish radiations. Mol Biol Evol 24:1269–1282

Gerrard DT, Meyer A (2007) Positive selection and gene conversion in SPP120, a fertilization-related gene, during the East African cichlid fish radiation. Mol Biol Evol 24:2286–2297

Golding GB, Dean AM (1998) The structural basis of molecular adaptation. Mol Biol Evol 15:355–369

Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol 537:113–137

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101–104

Hofmann CM, O'Quin KE, Marshall NJ, Cronin TW, Seehausen O, Carleton KL (2009) The eyes have it: regulatory and structural changes both underlie cichlid visual pigment diversity. PLoS Biol 7:e1000266

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter

S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M (2002) The Ensembl genome database project. Nucleic Acids Res 30:38–41

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity 99:364–373

Hulsey CD, Roberts RJ, Lin AS, Guldberg R, Streelman JT (2008) Convergence in a mechanically complex phenotype: detecting structural adaptations for crushing in cichlid fish. Evolution 62:1587–1599

Jeukens J, Bittner D, Knudsen R, Bernatchez L (2009) Candidate genes and adaptive radiation: insights from transcriptional adaptation to the limnetic niche among coregonine fishes (Coregonus spp., Salmonidae). Mol Biol Evol 26:155–166

Kapralov MV, Filatov DA (2006) Molecular adaptation during adaptive radiation in the Hawaiian endemic genus Schiedea. PLoS One 1:e8

Kijimoto T, Watanabe M, Fujimura K, Nakazawa M, Murakami Y, Kuratani S, Kohara Y, Gojobori T, Okada N (2005) cimp1, a novel astacin family metalloproteinase gene from East African cichlids, is differentially expressed between species during growth. Mol Biol Evol 22:1649–1660

Kobayashi N, Watanabe M, Horiike T, Kohara Y, Okada N (2009) Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. Gene 441:187–191

Koblmuller S, Salzburger W, Sturmbauer C (2004) Evolutionary relationships in the sand-dwelling cichlid lineage of Lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. J Mol Evol 58:79–96

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat Rev Genet 5:288–298

Kocher TD, Conroy JA, McKaye KR, Stauffer JR (1993) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. Mol Phylogenet Evol 2:158–165

Kuraku S, Meyer A (2008) Genomic analysis of cichlid fish 'natural mutants'. Curr Opin Genet Dev 18:551–558

Lee BY, Howe AE, Conte MA, D'Cotta H, Pepey E, Baroiller JF, di Palma F, Carleton KL, Kocher TD (2010) An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. BMC Genomics 11:278

Messier W, Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. Nature 385:151

Meyer A (1993) Phylogenetic-relationships and evolutionary processes in East-African cichlid fishes. Trends Ecol Evol 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

Meyer A, Biermann CH, Orti G (1993) The phylogenetic position of the zebrafish (Danio-Rerio), a model system in developmental biology—an invitation to the comparative method. Proc R Soc Lond B Biol Sci 252:231–236

Ohta T (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J Mol Evol 40:56–63

Peterson GI, Masel J (2009) Quantitative prediction of molecular clock and Ka/Ks at short timescales. Mol Biol Evol 26:2595–2603

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. Nucleic Acids Res 33:W116–W120

Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (Coregonas sp. Salmonidae) species pairs. Mol Biol Evol 24:1423–1438

Ruber L, Verheyen E, Meyer A (1999) Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. Proc Natl Acad Sci USA 96:10230–10235

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

Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C (2002) Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. Syst Biol 51:113–135

Salzburger W, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol Biol 5:17

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 Biol 5:51

Salzburger W, Renn SC, Steinke D, Braasch I, Hofmann HA, Meyer A (2008) Annotation of expressed sequence tags for the East African cichlid fish Astatotilapia burtoni and evolutionary analyses of cichlid ORFs. BMC Genomics 9:96

Schluter D (2000) The ecology of adaptive radiation. Oxford University Press, Oxford

Schluter D (2001) Ecology and the origin of species. Trends Ecol Evol 16:372–380

Schluter D (2009) Evidence for ecological speciation and its alternative. Science 323:737–741

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. Proc R Soc Lond B Biol Sci 273:1987–1998

Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HD, Miyagi R, van der Sluijs I, Schneider MV, Maan ME, Tachida H, Imai H, Okada N (2008) Speciation through sensory drive in cichlid fish. Nature 455:620–626

Slanchev K, Carney TJ, Stemmler MP, Koschorz B, Amsterdam A, Schwarz H, Hammerschmidt M (2009) The epithelial cell adhesion molecule EPCAM is required for epithelial morphogenesis and integrity during zebrafish epiboly and skin development. PLoS Genet 5:e1000563

Spady TC, Seehausen O, Loew ER, Jordan RC, Kocher TD, Carleton KL (2005) Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. Mol Biol Evol 22:1412–1422

Steinke D, Salzburger W, Meyer A (2006) Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. J Mol Evol 62:772–784

Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature 358:578–581

Sugawara T, Terai Y, Okada N (2002) Natural selection of the rhodopsin gene during the adaptive radiation of East African Great Lakes Cichlid fishes. Mol Biol Evol 19:1807–1811

Sugie A, Terai Y, Ota R, Okada N (2004) The evolution of genes for pigmentation in African cichlid fishes. Gene 343:337–346

Summers K, Zhu Y (2008) Positive selection on a prolactin paralog following gene duplication in cichlids: adaptive evolution in the context of parental care? Copeia 4:872–876

Terai Y, Morikawa N, Kawakami K, Okada N (2002a) Accelerated evolution of the surface amino acids in the WD-repeat domain encoded by the hagoromo gene in an explosively speciated lineage of East African cichlid fishes. Mol Biol Evol 19:574–578

Terai Y, Morikawa N, Kawakami K, Okada N (2003) The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. Proc Natl Acad Sci USA 100:12798–12803

Terai Y, Morikawa N, Okada N (2002b) The evolution of the pro-domain of bone morphogenetic protein 4 (Bmp4) in an explosively speciated lineage of East African cichlid fishes. Mol Biol Evol 19:1628–1632

Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2:Unit 2.3

Trzpis M, McLaughlin PM, de Leij LM, Harmsen MC (2007) Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. Am J Pathol 171:386–395

Trzpis M, Bremer E, McLaughlin PM, de Leij LF, Harmsen MC (2008a) EpCAM in morphogenesis. Front Biosci 13:5050–5055

Trzpis M, McLaughlin PM, Popa ER, Terpstra P, Van Kooten TG, De Leij LM, Harmsen MC (2008b) EpCAM homologues exhibit epithelial-specific but different expression patterns in the kidney. Transgenic Res 17:229–238

Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. Science 300:325–329

Watanabe M, Kobayashi N, Shin-i T, Horiike T, Tateno Y, Kohara Y, Okada N (2004) Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock: identity of EST sequences between *Haplochromis chilotes* and *Haplochromis* sp. "Redtailsheller". Gene 343:263–269

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

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Zardoya R, Vollmer DM, Craddock C, Streelman JT, Karl S, Meyer A (1996) Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). Proc Biol Sci 263:1589–1598

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479