EVOLUTIONARY GENETICS OF NATURALLY AND SEXUALLY SELECTED TRAITS IN CICHLID FISHES

INAUGURALDISSERTATION

ZUR
ERLANGUNG DER WÜRDE EINES DOKTORS DER PHILOSOPHIE
VORGELEGT DER
PHILOSOPHISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER UNIVERSITÄT BASEL

VON

EVELINE TAMARA DIEPEVEEN
AUS AMSTERDAM, DIE NIEDERLANDE

BASEL, 2014

EVOLUTIONARY GENETICS OF NATURALLY AND SEXUALLY SELECTED TRAITS IN CICHLID FISHES

GENEHMIGT VON DER
PHILOSOPHISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
AUF ANTRAG VON

PROF. DR. WALTER SALZBURGER
(UNIVERSITÄT BASEL)
UND
PROF. DR. IR. DICK T.F.M. ROELOFS
(VU UNIVERSITY)

BASEL, 11 DEZEMBER, 2012

PROF. DR. JÖRG SCHIBLER
THE DEAN OF FACULTY

to my parents

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Schweiz
(CC BY-NC-ND 3.0 CH)

Sie dürfen: **Teilen** — den Inhalt kopieren, verbreiten und zugänglich machen

Unter den folgenden Bedingungen:

Namensnennung — Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Keine kommerzielle Nutzung — Sie dürfen diesen Inhalt nicht für kommerzielle Zwecke nutzen.

Keine Bearbeitung erlaubt — Sie dürfen diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.

Wobei gilt:

- **Verzichtserklärung —** Jede der vorgenannten Bedingungen kann **aufgehoben** werden, sofern Sie die ausdrückliche Einwilligung des Rechteinhabers dazu erhalten.
- **Public Domain (gemeinfreie oder nicht-schützbare Inhalte) —** Soweit das Werk, der Inhalt oder irgendein Teil davon zur Public Domain der jeweiligen Rechtsordnung gehört, wird dieser Status von der Lizenz in keiner Weise berührt.
- **Sonstige Rechte —** Die Lizenz hat keinerlei Einfluss auf die folgenden Rechte:
    - Die Rechte, die jedermann wegen der Schranken des Urheberrechts oder aufgrund gesetzlicher Erlaubnisse zustehen (in einigen Ländern als grundsätzliche Doktrin des **fair use** bekannt);
    - Die **Persönlichkeitsrechte** des Urhebers;
    - Rechte anderer Personen, entweder am Lizenzgegenstand selber oder bezüglich seiner Verwendung, zum Beispiel für **Werbung** oder Privatsphärenschutz.
- **Hinweis —** Bei jeder Nutzung oder Verbreitung müssen Sie anderen alle Lizenzbedingungen mitteilen, die für diesen Inhalt gelten. Am einfachsten ist es, an entsprechender Stelle einen Link auf diese Seite einzubinden.

Quelle: http://creativecommons.org/licenses/by-nc-nd/3.0/ch/ Datum: 12.11.2013

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

> "*THERE IS GRANDEUR IN THIS VIEW OF LIFE, WITH ITS SEVERAL POWERS, HAVING BEEN ORIGINALLY BREATHED INTO A FEW FORMS OR INTO ONE; AND THAT, WHILST THIS PLANET HAS GONE CYCLING ON ACCORDING TO THE FIXED LAW OF GRAVITY, FROM SO SIMPLE A BEGINNING ENDLESS FORMS MOST BEAUTIFUL AND MOST WONDERFUL HAVE BEEN, AND ARE BEING, EVOLVED.*"
>
> -Charles Darwin, 1859-

In this closing sentence of the most renowned book in evolutionary history, *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, Charles Darwin (1859) captures brilliantly the stunning process of evolution and shows his admiration for the great diversity surrounding us. It is exactly this biological diversity[1] and the shared ancestries of organisms that have interested (evolutionary) biologists for more than a century and that still form the main pillars of evolutionary biology as research field.

The evolution of the field of evolutionary biology

To study (the evolutionary processes underlying) the diversity of life forms surrounding us, many tools, theories and methods have been generated. Some of these fundamental ideas, discoveries and theories date back several centuries. Already before his influential book on the origin of species in 1859 (Darwin 1859) and joint publication with Alfred Russel Wallace in 1858 (Darwin & Wallace 1858), in which the theory of natural selection was presented, Charles Darwin's grandfather Erasmus put forward the idea of descent from a common ancestor (Darwin 1818). Half a century later Gregor Mendel discovered the typical inheritance patterns of dominant and recessive traits in pea plants (Mendel 1866). So both the original theory of natural selection and the first empirical proof of a mechanism of inheritance date back to the 19th century, although a broad acceptance and integration of the two would last until the 20th century.

The field of evolutionary biology greatly revolutionized in the 20th century with the discovery of e.g., the molecular structure of DNA in 1953 (Watson & Crick 1953) and the introduction and development of statistical genetics and population genetics (e.g., Fisher 1919; Haldane 1932; Wright 1932). It was now possible –and customary- to study gene mutations, their effect on phenotype and their distribution throughout, mostly, populations in the laboratory.

---

[1] Biodiversity or biological diversity is defined as *the variability among living organisms from all resources including inter alia, terrestrial marine and other aquatic systems and the ecological complexes of which they are part [ ]* (Convention on Biological Diversity, Article 2. United Nations Conference on the Environment and development, 'The Earth Summit' 1992.)

Within the last decade, further integration of subfields, such as evolutionary developmental biology, bio-informatics and genomics, together with the development of new technologies, such as more efficient sequencing techniques (i.e., Next Generation Sequencing, Third Generation Sequencing) made it possible to study a broad range of questions directly in wild populations. In the 21st century evolutionary biology is a truly interdisciplinary field, from field ecology to comparable genomics. With these technological advances and integration of different subfields, the main focus of the field is shifted towards understanding and finding the exact changes in the genome underlying a particular natural occurring phenotype of interest. By doing this, scientists are now able to explain the differences not only between populations or species, but ultimately also between two individuals.

Well-studied examples of this approach focus on adaptive traits in response to new or altered environments, such as the colonization of novel freshwater lake and stream habitats by marine three-spined sticklebacks (*Gasterosteus aculeatus*) after the last Ice Age and the colonization of newly formed islands in the Gulf of Florida and of the Atlantic coast by the oldfield mice (*Peromyscus polionotus*). These species are known for their phenotypic adaptations to the novel environments. Freshwater sticklebacks have evolved a reduction in armor plates and pelvic skeleton, possibly through lowered predation pressure, compared to their marine relatives. The mice inhabiting the light beach habitats have reduced pigment patterns in comparison to the melanic pigment patterns of the inland mice. In both cases the underlying genetic basis of the adaptive phenotype has been largely determined (Shapiro *et al*. 2004; Colosimo 2005; Steiner *et al*. 2007; 2009; Chan *et al*. 2010; Jones *et al*. 2012).

The focus of evolutionary biologists nowadays, is not only aimed at determining the genetic changes underlying (adaptive) phenotypes, but also on disentangling the effects and contribution of different evolutionary processes (i.e., migration, mutation, genetic drift and selection) on the genome-wide genetic diversity of both wild and laboratory populations in response to specific environmental conditions. For instance, migration rates can be estimated from multilocus sequence data to give insight into (ongoing) gene flow between populations inhabiting divergent habitats or between sympatric species (see e.g., Hey & Nielsen 2004; Rosenblum 2006). In addition, direct measurements of predation pressures on mice showed that natural selection, through visual hunting predators, is responsible for the cryptic coat color of *Peromyscus* mice inhabiting both the light-substrate beach habitats and the dark-substrate inland habitats along the Gulf of Florida (Vignieri *et al*. 2010). These approaches, thus, facilitate the reconstruction of the evolutionary history of populations, the roles of the evolutionary processes during adaptations to new environments and ultimately in the evolution of species and biological diversity.

Since one of the major goals of evolutionary biology is to understand the processes shaping natural populations, many studies focus on organisms other than the traditional laboratory model systems such the fruit fly *Drosophila melanogaster*, the brassicacaen *Arabidopsis thaliana* and the house mouse *Mus musculus* (see e.g., Abzhanov *et al*. 2008). Natural populations that (convergently) adapted to new or altered conditions are particular suitable systems to study (the genes involved in) adaptation and speciation. Besides the three-spined sticklebacks and *Peromyscus*

mice, great scientific interest has been shown in e.g., *Mimulus* monkey flowers, *Poecilia* guppy, and *Heliconius* and *Bicyclus* butterflies (see e.g., Abzhanov *et al*. 2008). Furthermore, species originating from an adaptive radiation, which is described as the evolution of ecological diversity within a rapidly multiplying lineage (Simpson 1953; Schluter 2000), are particular interesting model systems, since they comprise many closely-related species with a broad diversity in ecological and phenotypic traits. Most compelling examples of adaptive radiations are the Darwin finches on the Galapagos islands (Grant & Grant 2011), the Anoles lizards of the greater Antilles (Losos *et al*. 1997; Butler *et al*. 2007; Losos 2009) and the cichlid fishes of the East African great lakes (Kocher 2004; Seehausen 2006; Salzburger 2009).

The cichlid fishes as model system for evolutionary biology

> "*I've never met an animal, or plant for that matter, that wasn't interesting, but some stand out as special. Cichlid fishes are right up there.*"
>
> -George Barlow, 2000-

In my PhD thesis I use cichlid fishes as model system to study the evolutionary genetics of naturally and sexually selected traits. As the above quote of Barlow (Barlow 2000) illustrates, cichlid fishes are remarkable animals that have furthermore been proposed as model systems for evolutionary biology (Kornfield & Smith 2000; Kocher 2004; Salzburger 2009). The latter is caused by the fact that there are thousands of cichlid species that are characterized by a great diversity in phenotypic and ecological traits.

Cichlid fishes (Cichlidae) are ray-finned fishes (Actinopterygii) belonging to the Teleostei. They are found in numerous lake and river systems across Africa and Madagascar, Central and South America, the Middle East, India and Sri Lanka (see (Salzburger & Meyer 2004; Salzburger 2009). This biogeography corresponds with an initial Gondwanan distribution (excluding Australia and Antarctica) and its following fragmentation of landmasses, which is furthermore, supported by phylogenetic and biogeographic studies (Sparks & Smith 2004; 2005). However, species numbers in these regions are not equally distributed. The most basal and species-poor lineages are found in Asia, Madagascar and the Middle East (3, 32 and 7, respectively), while the lineages in Central and South America are relatively species-rich (110, 450; Turner 2007). Notably, the African lineages are particularly species-rich, with estimated species numbers of 180-250 in Lake Tanganyika, 600-700 in Lake Malawi, around 500-700 in Lake Victoria, plus another 300 in the remainder of the African continent (Turner *et al*. 2001; Turner 2007).

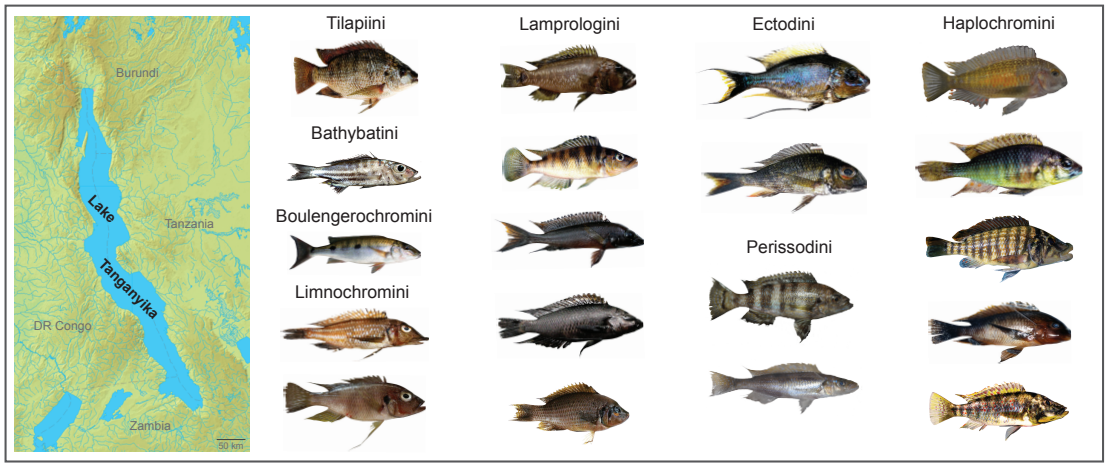


Figure 1.1. Lake Tanganyika and its cichlids' diversity. Lake Tanganyika (left) is a huge rift valley lake in East Africa. It is bordering Zambia in the south, Tanzania in east, Burundi in the north and the Democratic Republic of the Congo in the west and is inhabited by roughly 200 different species of cichlid fish. Displayed in the right part of the figure are eight of the Lake Tanganyikan lineages. The represented species vary in e.g., their body shape and size, coloration, diet and reproductive behavior. [Photographs by the Salzburger Lab]

The exceptional high numbers of species in the East African Great Lakes, Malawi, Tanganyika and Victoria, are thought to have evolved within the last 10 million years (Kocher 2004; Sturmbauer *et al*. 2011), although some estimates are as low as 12,000 years for Lake Victoria (Johnson *et al*. 1996; Fryer 1997). These geologically and evolutionary short time periods together with the species-richness are proof for explosive and rapid speciation events.

Interestingly, virtually all species from Lakes Malawi and Victoria belong to the haplochromine cichlid lineage (Fryer & Iles 1972; Greenwood 1979; Meyer *et al*. 1990) and phylogenetic reconstructions based on mitochondrial DNA have shown that these species, as well as the haplochromines of Lake Tanganyika (in this case Tropheini), are ultimately derived from Lake Tanganyikan ancestors (Salzburger *et al*. 2005). Lake Tanganyika is furthermore known for its diverse radiation, the primary lacustrine radiation, that resulted in the phylogenetic distinct lineages of the Trematocarini, Bathybatini, Eretmodini, Lamprologini, Ectodini, Cyprichromini, Perissodini, Limnochromini, Benthochromini, Cyphotilapiini and the Haplochromini (Salzburger *et al*. 2002; Sturmbauer *et al*. 2011; see Figure 1.1). As a consequence of these events and the occurrence and/or colonization of several non-radiated lineages (i.e., Tylochromini, Tilapiini, Boulengerochromini and Hemibatini), Lake Tanganyika is now populated by fifteen different lineages (Salzburger *et al*. 2002; Day *et al*. 2008; Sturmbauer *et al*. 2011).

Within these different lineages an astonishing high level of morphological, phenotypical and behavioral diversity evolved (Kornfield & Smith 2000; Barlow 2000; Salzburger & Meyer 2004; Kocher 2004; Seehausen 2006; Salzburger 2009; see Figure 1.1), forming a 'natural mutagenic screen' (Kocher 2004) that offers great opportunities to study the genetic basis of speciation and adaptation. Examples of such studies are given in later chapters as well as elsewhere (Salzburger 2009; Santos & Salzburger 2012).

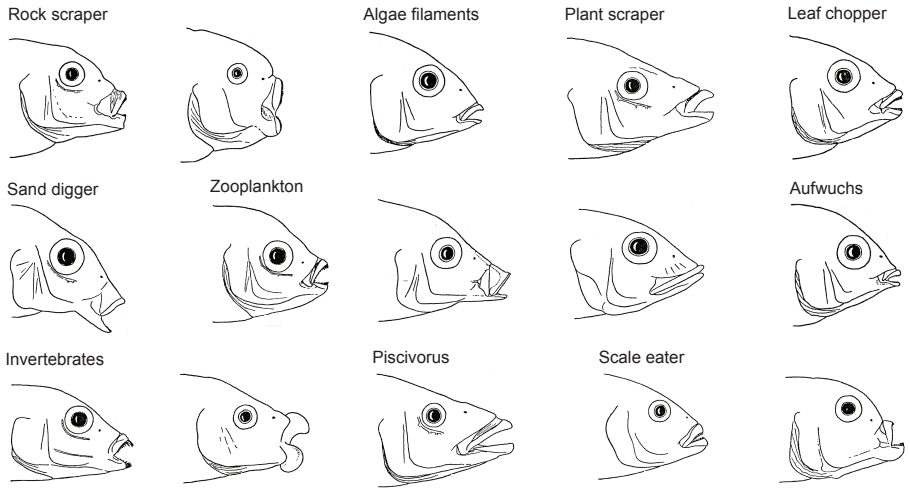


Figure 1.2. Variation in cichlid mouth morphologies which corresponds with diet. After Fryer & Iles 1972

One of the traits that is particular diverse is the cichlids' mouth morphology (Figure 1.2). The main parts of the trophic apparatus are the oral jaws and the pharyngeal jaws, which have evolved from highly modified parts of the gill branches. It has been hypothesized that by processing the food items with the pharyngeal jaws, the oral jaws were freed to evolve diverse specialization in respect to the collection of these items (Liem 1973). The cichlids' mouth morphology (including the pharyngeal jaw) highly correlates with its diet (see Takahashi & Koblmüller 2011 for references; Muschick *et al*. 2012), which ranges from algae, insects, other fish, other fishes' scales or eyes to mollusks (see also Figure 1.2). Furthermore, it has been shown that diet or mode of feeding can change the pharyngeal jaw shape, implying that the trait is plastic and can respond quickly to an alternative food source (Hellig *et al*. 2010; Muschick *et al*. 2011).

Pigmentation is also highly diverse in cichlid fishes (Figure 1.1). Especially, body coloration and patterns between sexes or between sympatric species can be highly different. While female body coloration is often dull, males are characterized by bright coloration, which forms the basis of many behavioral experiments in which typically female mate choice gets studied. These studies revealed that the female mate choice contributes to reproductive isolation and ultimately to speciation (see e.g., Turner *et al*. 1995; Knight & Turner 2004; Deutsch 2008).

Interestingly, another pigmentation trait is believed to play a role in the mating behavior of the most species-rich cichlid lineage: the mouth brooding modern haplochromines (Wickler 1962; Hert 1989). The males of these species possess colorful spots, resembling real eggs – to some degree - on their anal fins close to their genital opening. It is thought that after the female has laid a clutch of eggs, the male presents her the egg-spots on his anal fin, to which the female reacts by picking up the eggs with her mouth. At this point, the male releases sperm and the eggs get fertilized in the female's mouth. Variation in size, number and color of egg-spot might be evolutionary important and its function can be studied with e.g., mate choice experiments (see e.g., Hert 1989; Couldridge 2002; Egger *et al*. 2011; Theis *et al*. 2012).

A third highly diverse trait among cichlid species is breeding behavior, including courtship and parental care systems. For instance, the above mentioned haplochromine lineage is characterized by maternal mouthbrooding behavior. This means that the fertilized eggs, and later on the fry, are kept in the female's mouth. Mating systems across species range from monogamy to harem polygyny, while parental care in cichlids includes maternal/paternal/biparental mouthbrooding, biparental substrate-breeding and cooperative breeding (Sefc 2011).
Remaining traits and the role of natural and sexual selection on all of them, are described elsewhere (e.g., Salzburger 2009; Takahashi & Koblmüller 2011; Sturmbauer *et al*. 2011) as well as in the next chapters.

Thus, cichlid fishes of the East African lakes, as well as several lineages in Central and South America, are characterized by great ecological and phenotypical diversity that evolved in rapidly multiplying lineages, which are the two main components of the classical definition of adaptive radiations (Simpson 1953; Schluter 2000). Cichlid fishes are great examples of adaptive radiations and excellent model systems for the study of adaptation, speciation and evolutionary innovation (Kornfield & Smith 2000; Kocher 2004; Seehausen 2006; Salzburger 2009).

Research goals of my doctoral thesis

Several traits of cichlid fishes such as; the pharyngeal jaw apparatus, haplochromine maternal mouthbrooding, egg-spots and color polymorphisms have been proposed as evolutionary key innovations (Liem 1973; Salzburger *et al*. 2005). Key innovations are described as novel traits that allow its bearer to exploit new ecological niches, and are considered as one of the possible causes leading to an adaptive radiation (Schluter 2000; Gavrilets & Losos 2009; Losos 2010). Furthermore, these traits could (partly) be responsible for the evolutionary success of cichlid lineages.
Besides these key innovations, other adaptive traits could also have increased the fitness of an individual and subsequently spread within a population. These adaptive traits can be of a morphological, behavioral as well as physiological kind and are eminent candidates to study the genetic basis of adaptation and diversification. Furthermore, they can be used to determine the role of natural and/or sexual selection in adaptive processes (see Salzburger 2009; Sturmbauer *et al*. 2011).

In my doctoral work I studied the molecular evolutionary genetics of candidate gene families that are associated with either key innovations or other adaptations in cichlids or specific lineages thereof. The goal was to get a better understanding of the sequence evolution and type of selection (i.e., neutral, purifying or positive selection) acting upon genes that are involved in evolutionary important traits. Furthermore, functional implications of these loci can then be validated by means of e.g., expression assays.
The goal of the second part of my thesis was to unravel the genetic basis of a particular adaptation: the thick-lipped phenotype observed in multiple cichlid lineages. Here, the aim was to generate a candidate gene list of loci that are putatively involved in the evolution of this particular adaptation. Therefore, the overall objective

of my thesis was to obtain further insights into the evolutionary genetics of naturally and sexually selected traits in cichlid fishes.

Outline of the thesis

My thesis is divided in two main parts: the first part (i.e., chapter 2 through 4) focuses on characterizing candidate gene families, while in the second part (i.e., chapter 5 and 6) genomic approaches are described.

In **chapter 2** I present the molecular characterization of the endothelin family of ligands and receptors. The genes of this family are putatively involved in the development and evolution of two key innovations in cichlids: the pharyngeal jaw apparatus of all cichlids and the egg-spots of haplochromines. These loci are thus candidate genes for evolutionary innovations. I found elevated rates of protein evolution in the two studied ligands compared to the receptors and relative high expression levels of the two studied developmental pathways in two evolutionary key innovations. These results suggest that the endothelin ligands and receptors are indeed involved in the morphogenesis of naturally and sexually selected traits in cichlids.

In **chapter 3** I discuss another gene family which is putatively involved in the same evolutionary novelties; the pharyngeal jaw apparatus and the anal fin egg-spots: the *distal-less homeobox* gene family. These loci are widely known for their crucial roles in the development (of components) of the nervous system, craniofacial skeleton and connective tissue, and in the formation of appendages and have been implicated with evolutionary novelties in other species such as; butterflies, other insects and vertebrates. They are thus, renowned candidate genes for morphological diversification across the animal kingdom, including cichlid fishes. In this chapter I show that the *dlx* repertoire of cichlid fishes is shaped by differential selection pressures and rates of evolution after gene duplication, and I try to relate this to their presumed crucial roles in the development of several morphological traits.

The results of a third gene family study, called the *Hivep* gene family, are highlighted in **chapter 4**. Here I combined molecular characterization of separate loci with an immune challenge to show the presumed immune-related roles of paralogs. The results revealed that exposure of the cichlid *Astatotilapia burtoni* to a vaccination with *Vibrio anguillarum* bacteria provoked an immune response and significant positive correlations between the expression levels of several *Hivep* paralogs and the lymphocyte/monocyte ratios. These results suggest a role for *Hivep* paralogs in the immunological parasite defense of *A. burtoni*. Furthermore, I hypothesize that immune genes might have been particular important during cichlid radiations, in which contact to novel pathogenic environments is likely to have left signs of adaptive protein evolution in genes, of which *Hivep* is a prime example.

Top-down, or whole genome, approaches can facilitate the generation of candidate gene lists and with the current revolution of sequencing techniques it becomes more and more feasible to perform such studies in non-model organisms. **Chapter 5** illustrates this development specifically in the field of molecular ecology. It is a

meeting review of the one-day symposium organized by the journal *Molecular Ecology* in 2012, which illustrates that Next Generation Sequencing approaches are extremely useful in many different subfields of molecular ecology.

One such technique, RNA Sequencing was used to determine the genetic basis of a convergent thick-lipped phenotype observed in both African and Nicaraguan species. These results, along with several other ecological analyses are presented in **chapter 6**. In this work, I was merely involved in the RNAseq and gene expression assays parts. A comparative Illumina RNA sequencing approach was used that involved thick and 'normal' lip tissue and a set of 141 differentially expressed candidate genes was identified. Six of these genes were further studied by means of comparative quantitative real-time PCR. Three strong candidates, *Actb*, *Cldn7* and *Copb* were identified. Interestingly, similar trends in gene expression were found between African and Central American thick-lipped species, indicating that an overlapping set of genes was independently recruited to build this particular phenotype in both lineages.

As closing chapter of this thesis, the work presented here is discussed in **chapter 7**.

References

Abzhanov A, Extavour CG, Groover A, *et al*. (2008) Are we there yet? Tracking the development of new model systems. *Trends in Genetics*, **24**, 353–360.

Barlow GW (2000) *The cichlid fishes: nature's grand experiment in evolution*. Perseus publishing, Cambridge.

Butler MA, Sawyer SA, Losos JB (2007) Sexual dimorphism and adaptive radiation in Anolis lizards. *Nature*, **447**, 202–205.

Chan YF, Marks ME, Jones FC, *et al*. (2010) Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science*, **327**, 302–305.

Colosimo PF, Hosemann KE, Balabhadra S, *et al*. (2005) Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. *Science*, **307**, 1928–1933.

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

Darwin C (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London.

Darwin C, Wallace A (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the proceedings of the Linnean Society of London. Zoology*, **3**, 45–62.

Darwin E (1818) *Zoonomia; or the laws of organic life*. Edward Earle, Philadelphia.

Day JJ, Cotton JA, Barraclough TG (2008) Tempo and mode of diversification of lake Tanganyika cichlid fishes. *PLoS ONE*, **3**, e1730.

Deutsch JC (2008) Colour diversification in Malawi cichlids: evidence for adaptation, reinforcement or sexual selection? *Biological Journal of the Linnean Society*, **62**, 1–14.

Egger B, Klaefiger Y, Theis A, Salzburger W (2011) A Sensory Bias Has Triggered the Evolution of Egg-Spots in Cichlid Fishes. *PLoS ONE*, **6**, e25601.

Fisher RA (1919) The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, **52**, 399–433.

Fryer G (1997) Biological implications of a suggested Late Pleistocene desiccation of Lake Victoria. *Hydrobiologia*, **354**, 177–182.

Fryer G, Iles TD (1972) *The cichlid fishes of the Great Lakes of Africa: their biology and Evolution*. Oliver & Boyd, Edinburgh.

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

Grant PR, Grant BR (2011) *How and why species multiply: The radiation of Darwin's finches*. Princeton University Press, Princeton, New Jersey.

Greenwood PH (1979) Towards a phyletic classification of the "genus" Haplochromis (Pisces, Cichlidae) and related taxa. Part 1. *Bulletin of the British Museum (Natural History) Zoology*, **35**, 265–322.

Haldane J (1932) The Time of Action of Genes, and Its Bearing on some Evolutionary Problems. *the american naturalist*, **66**, 5–24.

Hellig CJ, Kerschbaumer M, Sefc KM, Koblmüller S (2010) Allometric shape change of the lower pharyngeal jaw correlates with a dietary shift to piscivory in a cichlid fish. *Naturwissenschaften*, **97**, 663–672.

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

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. *Genetics*, **167**, 747–760.

Johnson T, Scholz C, Talbot M, *et al*. (1996) Late Pleistocene Desiccation of Lake Victoria and Rapid Evolution of Cichlid Fishes. *Science*, **273**, 1091–1093.

Jones FC, Grabherr MG, Chan YF, *et al*. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Knight ME, Turner GF (2004) Laboratory mating trials indicate incipient speciation by sexual selection among populations of the cichlid fish Pseudotropheus zebra from Lake Malawi. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 675–680.

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

Kornfield I, Smith PF (2000) African cichlid fishes: model systems for evolutionary biology. *Annual review of ecology and systematics*, 163–196.
Liem KF (1973) Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. *Systematic Zoology*, **22**, 425–441.
Losos JB (2009) *Lizards in an evolutionary tree: Ecology and adaptive radiation of Anoles*. University of California Press, Berkeley and Los Angeles.
Losos JB (2010) Adaptive Radiation, Ecological Opportunity, and Evolutionary Determinism. *the american naturalist*, **175**, 623–639.
Losos JB, Warheit KI, Schoener TW (1997) Adaptive differentiation following experimental island colonization in Anolis lizards. *Nature*, **387**, 70–73.
Mendel G (1866) Versuche über Pflanzenhybriden. *Verh. Naturforsch. Ver. Brünn*, **4**, 3–47.
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.
Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evolutionary Biology*, **11**, 116.
Muschick M, Indermaur A, Salzburger W (2012) Convergent Evolution within an Adaptive Radiation of Cichlid Fishes. *Current Biology*, **22**, 2362-2368.
Rosenblum EB (2006) Convergent evolution and divergent selection: Lizards at the White Sands ecotone. *the american naturalist*, **167**, 1–15.
Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**, 169–185.
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, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology*, **5**, 17.
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. *Systematic Biology*, **51**, 113–135.
Santos ME, Salzburger W (2012) How cichlids diversify. *Science*, **338**, 619–621.
Schluter D (2000) *The ecology of adaptive radiation* (PH Harvey, RM May, Eds,). oxford University Press, Oxford.
Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B*, **273**, 1987–1998.
Sefc KM (2011) Mating and Parental Care in Lake Tanganyika's Cichlids. *International Journal of Evolutionary Biology*, **2011**, 1–20.
Shapiro MD, Marks ME, Peichel CL, *et al*. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.
Simpson GG (1953) *The major features of evolution*. Columbia University Press, New York.
Sparks JS, Smith WL (2004) Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics*, **20**, 501–517.
Sparks JS, Smith WL (2005) Freshwater fishes, dispersal ability, and nonevidence: "Gondwana Life Rafts" to the Rescue. *Systematic Biology*, **54**, 158–165.
Steiner CC, Römpler H, Boettger LM, Schöneberg T, Hoekstra HE (2009) The Genetic Basis of Phenotypic Convergence in Beach Mice: Similar Pigment Patterns but Different Genes. *Molecular Biology and Evolution*, **26**, 35–45.
Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, **5**, e219.
Sturmbauer C, Husemann M, Danley PD (2011) Explosive speciation and adaptive radiation of East African cichlid fishes. In: *Biodiversity Hotspots* (eds Zachos FE, Habel JC), pp. 333–362. Springer Berlin Heidelberg, Berlin, Heidelberg.
Takahashi T, Koblmüller S (2011) The Adaptive Radiation of Cichlid Fish in Lake Tanganyika: A Morphological Perspective. *International Journal of Evolutionary Biology*, **2011**, 1–14.
Theis A, Salzburger W, Egger B (2012) The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. *PLoS ONE*, **7**, e29878.
Turner GF (2007) Adaptive radiation of cichlid fish. *Current Biology*, **17**, R827–31.
Turner GF, Burrows MT (1995) A model of sympatric speciation by sexual selection. *Proceedings of*

the Royal Society of London. Series B: Biological Sciences, **260**, 287–292.
Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL (2001) How many species of cichlid fishes are there in African lakes? In:, pp. 793–806.
Vignieri SN, Larson JG, Hoekstra HE (2010) The selective advantage of crypsis in mice. *Evolution*, **64**, 2153–2158.
Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
Wickler W (1962) "Egg-dummies" as natural releasers in mouth-breeding cichlids. *Nature*, **194**, 1092–1093.
Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Int. Congr. Genet.*, 356–366.

PART I

CHARACTERIZATION OF CANDIDATE GENE FAMILIES

CHAPTER 2

MOLECULAR CHARCTERIZATION OF TWO ENDOTHELIN PATHWAYS IN EAST AFRICAN CICHLID FISHES

ARTICLE

J Mol Evol (2011) 73:355–368
DOI 10.1007/s00239-012-9483-6

Molecular Characterization of Two Endothelin Pathways in East African Cichlid Fishes

Eveline T. Diepeveen · Walter Salzburger

Received: 7 September 2011 / Accepted: 4 January 2012 / Published online: 21 January 2012

Abstract The adaptive radiations of cichlid fishes in East Africa have been associated with the acquisition of evolutionary novelties as well as the ecological opportunities existing in the East African Great lakes. Two remarkable evolutionary innovations are the pharyngeal jaw apparatus, found in all cichlid species, and the anal fin egg-spots of mouthbrooding cichlids. Based on their conserved functions during the development of both the jaw apparatus and pigmentation, the endothelin ligands and receptors form a putative link between these naturally and sexually selected traits. Here we study the evolutionary history of four members of two endothelin pathways (Edn1/EdnrAa and Edn3b/EdnrB1a) to elucidate their possible roles during the evolution and development of key innovations in East African cichlids species. The analyses performed on partial sequences (ca. 6,000 bp per taxon) show that all four endothelin family members evolved under purifying selection, although both ligands are characterized by an accelerated rate of protein evolution in comparison to the receptors. In accordance with earlier findings, we show that the mature protein sequence of Edn1 and Edn3 are highly conserved, also in cichlids, whereas the preproendothelin parts are variable indicating relaxed selective constraints. In the receptors, nonsynonymous substitutions were mainly found in the ligand-binding domains suggesting functional divergence. Gene expression assays with Real-Time PCR indeed reveal that the two studied endothelin pathways are expressed in the cichlid pharyngeal jaw and in the haplochromine egg-spot (among other pigment-cell containing tissues), suggesting their involvement during morphogenesis of naturally and sexually selected traits in cichlids.

Keywords Endothelin receptor · Pigmentation · Pharyngeal jaw · Neural crest · Key innovation · Molecular evolution

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9483-6) contains supplementary material, which is available to authorized users.

E. T. Diepeveen (✉) · W. Salzburger (✉)
Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland
e-mail: eveline.diepeveen@unibas.ch

W. Salzburger
e-mail: walter.salzburger@unibas.ch

Background

The spectacular adaptive radiations of cichlid fishes provide a unique model system to study evolutionary processes. No other vertebrate family consists of such a great number of species (at least 3,000 species worldwide), such a degree of phenotypic diversity and such a rapid diversification rate. Above all, the species flocks of cichlids in the East African Great Lakes show the greatest variation in morphology, coloration, ecology, behavior, and reproductive biology (Barlow 2000; Coulter 1991; Fryer and Iles 1972; Keenleyside 1991; Kocher 2004; Kornfield and Smith 2000; Salzburger 2009; Seehausen 2006). Furthermore, the cichlids' occurrence in various geographic regions (Africa, India, Sri Lanka, and Central and South America) and the independent (and repeated) colonization of multiple African lakes (e.g., lakes Malawi, Tanganyika, and Victoria) form the basis for some outstanding cases of parallel evolution within and between lakes and regions (Duftner et al. 2007; Elmer et al. 2010; Kocher et al. 1993).

According to the ecological opportunity hypothesis, fast divergence of lineages can occur when available and underutilized niches are being filled (Schluter 2000). This

scenario seems very plausible for cichlids that colonized and subsequently radiated in large rift lakes with their diversity in habitats (e.g., surge habitat, shallow rocky shores, sandy ground, mud floor, benthic water) and food resources (e.g., algae, plants, small invertebrates, and other fish). New niches can also become accessible when novel traits, so-called evolutionary key innovations, evolve with which so far unaccessible niches can be exploited (Gavrilets and Losos 2009; Schluter 2000). The evolutionary success of cichlids in general, and the explosive radiation of the East African haplochromine lineage in particular, have been associated with the acquisition of such evolutionary innovations (Fryer and Iles 1972; Liem 1973; Salzburger et al. 2005). The pharyngeal jaw apparatus (Fig. 1a), for instance, is found in all cichlids and is thought to underlie diversification with respect to feeding ecology (Fryer and Iles 1972; Hunter 1998; Liem 1973). The diversity in cichlid pharyngeal jaw morphology (and the teeth on there) directly correlates with the diversity in diet, from algae to whole fishes (Liem 1973). Even though other groups of teleost fishes are characterized by the presence of pharyngeal structures that aid the passage of food into the esophagus, cichlid fishes show the greatest diversity in pharyngeal jaw morphology resulting in the efficient manipulation of a great variety of food. Thus, the pharyngeal jaw apparatus is thought to play an important role in niche specialization and, hence, speciation in cichlids (Liem 1973; Muschick et al. 2011; Salzburger 2009).

Sexual selection is the other major factor that is likely to have contributed to the cichlids' species-richness (see, e.g., Meyer 1993; Salzburger 2009; Seehausen et al. 1997; Turner & Burrows 1995), as indicated by the occurrence of multiple color morphs and sexual color dimorphism. This is particularly obvious in the most species-rich group of cichlids, the haplochromines. The haplochromines exclusively consists of maternal mouthbrooders, which protect their offspring from predation in their buccal cavities (Hert 1989; Mrowka 1987; Fig. 1b). The occurrence of egg-spots on the anal fins of male haplochromines and other C-lineage tribes has been proposed as yet another key innovation (Salzburger et al. 2005, 2007; Fig. 1c). Variation in size, number, and color of the egg-spots has been observed both between and within species (Salzburger et al. 2007). Egg-spots appear to play a prominent role in mating behavior (Wickler 1962), although their exact function remains elusive (Egger et al. 2011; Theis et al. 2012).

The above-mentioned evolutionary key innovations of cichlids share a developmental origin. Both the mouth apparatus (including the pharyngeal jaws) and pigmentation (body coloration and egg-spots) are derived from the neural crest (Albertson and Kocher 2006; Gans and Northcutt 1983; Hall 1999; Slack 2001), which is—by

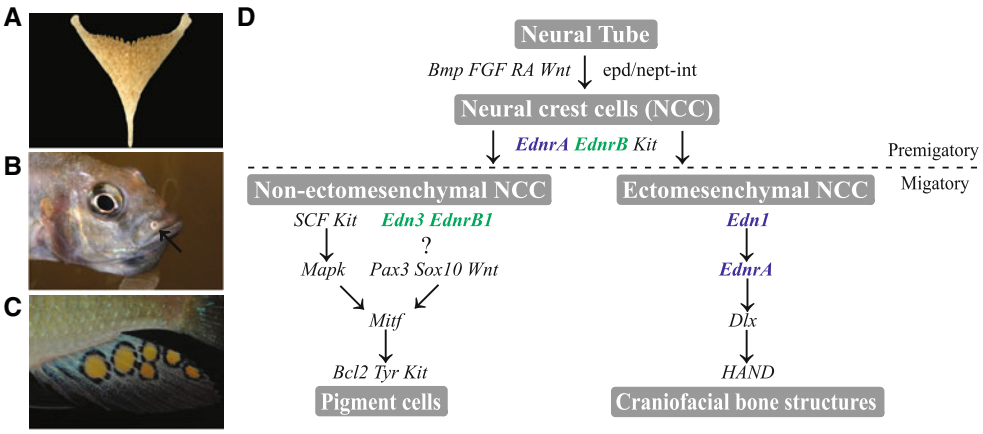


Fig. 1 Three putative key innovations and the genetic pathways involved in neural crest cell differentiation. **a** The lower pharyngeal jaw, which is a fusion of the fifth ceratobranchials, is responsible for grinding the food and is present in all cichlids. **b** A mouthbrooding haplochromine female. The female protects her eggs and later on the fry (as shown) from possible predation by holding them in her mouth. **c** The egg-spot on the anal fin of haplochromine males play an important role in breeding behavior. **d** The endothelin system is involved in the development of (pre)migratory neural crest cells and its derivatives in vertebrates. The *Edn3/EdnrB1* pathway (in *green*) is involved in pigment-cell development, while the *Edn1/EdnrAa* pathway (in *blue*) is involved in the lower jaw formation (among other craniofacial bone structures). Pathways depicted are simplified and not exhaustive. *Bcl2* B-cell leukemia/lymphoma 2, *Bmp* bone morphogenetic protein family, *Kit* kit receptor tyrosine kinase, *Dlx* distal-less homeobox, *Edn1/3* endothelin1/3, *EdnrA/B* endothelin receptor A/B, *epd/nept-int* epidermis/neuroepithelium-interaction, *FGF* fibroblast growth factor family, *HAND* heart and neural crest derivatives expressed transcript, *Mapk* mitogen-activated protein kinase, *Mitf* microphthalmia-associated transcription factor, *Pax3* paired box gene 3, *RA* retinoic acid, *SCF* stem cell factor, *Sox10* sex determining region Y-box 10, *Tyr* tyrosinase, *Wnt* wingless-type integration site family. Photographs taken by M. Muschick, A. Theis, and C. Heule. Pathways based on, e.g., Bonano et al. 2008; McGill et al. 2002; Parichy 2000; Pla and Larue 2003; Rawls et al. 2001; Slack 2001; Yanagisawa et al. 2003

itself—an evolutionary innovation of the vertebrate clade (e.g., Gans and Northcutt 1983; Shimeld and Holland 2000). Neural crest cells originate in the neural folds, and migrate away from the neural tube to form a wide variety of cell types across the whole body including neuron cells, pigment cells, craniofacial bones, Schwann cells, and smooth muscle cells (Chung et al. 2009; Slack 2001). The genetic and developmental basis of the neural crest-derived tissues and specifically those involved in evolutionary key innovations is, hence, of great importance for the study of cichlid evolution (Salzburger 2009).

Within the last decade several studies have focused on the genetic basis of naturally and sexually selected traits in cichlids (reviewed in Salzburger 2009). Albertson et al. (2005), for example, showed that *bmp4* is involved in craniofacial diversity and that it has the potential to change the mandible shape of cichlids. It had previously been shown that this gene underwent adaptive protein evolution in East African cichlid species (Terai et al. 2002b). More recently, Fraser et al. (2009) showed that similar genetic pathways underlie the tooth formation on both the oral and the pharyngeal jaws in cichlids. Seven Hox genes and several genes from a previously described dental gene network (Fraser et al. 2008) were shown to be expressed in developing pharyngeal jaw tissue in two Malawi cichlid species (Fraser et al. 2009). Apart from jaw-related genes, several candidate coloration genes have been studied in cichlids. Terai et al. (2003), for example, showed that cichlids express multiple alternatively spliced *hag* variants and that species-rich lineages feature more splice variants than species-poor ones. Also for this gene an accelerated rate of protein evolution was determined (Terai et al. 2002a). The *Pax7* locus appears to be involved in the formation of the orange blotch phenotype (Roberts et al. 2009). Finally, it has been shown that *csf1ra* is involved in the development of the colorful egg-spots on the male anal fin (Salzburger et al. 2007).

One specific family of candidate genes for neural crest-derived structures and, in particular, two prominent evolutionary innovations of cichlids (i.e., the pharyngeal jaw and coloration) is the endothelin (Edn) family. Endothelin 1 was discovered in 1988 as vasoconstrictor by Yanagisawa et al. (1988). Later, more members of this gene family were found and their functions annotated, so it is now evident that the endothelin system is involved in the formation of craniofacial structures, enteric ganglion neurons, pigmentation, and vasoconstriction in both mouse (Baynash et al. 1994; Clouthier et al. 1998; Hosoda et al. 1994; Kurihara et al. 1994; Ozeki et al. 2004; Yanagisawa et al. 1988) and zebrafish (Miller et al. 2000; Parichy et al. 2000). The active gene product of the endothelin genes is only 21 amino acids long and obtained via two steps of cleaving from the inactive precursors preproendothelin and big endothelin, respectively (Masaki 2004).

The Edn family arose in the vertebrate lineage from an ancestral signaling peptide and a G protein-coupled receptor. Subsequent expansion took place by three rounds of whole genome duplication (i.e., the two vertebrate-specific whole genome duplications and a fish-specific one), each followed by functional divergence and coevolution of the ligands and receptors (Braasch et al. 2009). Also, the range of interactions between the ligands and the receptors expanded with each round of genome duplication, since all endothelin receptors can—theoretically—bind all ligands, albeit with different affinities (note that this has not yet been examined in teleosts). Braasch et al. (2009) showed that functional changes in the receptors were mainly due to shifts of evolutionary rates and less to radical changes in amino acids. Furthermore, the sites where these changes were observed are mainly found in the extracellular loops or ligand-binding domains of the receptors and not in the intracellular domains where interaction with the G protein occurs. This indicates that with the expanding repertoire of the ligands, the receptors co-evolved by slightly changing the ligand–receptor binding affinity and selectivity. Since the third round of whole genome duplication took place in the lineage leading to the ray-finned fishes (Meyer and van de Peer 2005), the most abundant number of endothelins and receptors are found in today's teleost fishes. Six endothelin genes and five teleost endothelin receptor genes were identified in stickleback, while zebrafish, medaka, spotted green pufferfish, and torafugu are all characterized by a species-specific gene loss, resulting in a lower number of either receptor or ligand (Braasch et al. 2009).

The aim of this study is to elucidate the molecular evolutionary history of the endothelin family in East African cichlids and its possible role during the evolution and development of evolutionary key innovations. Therefore, two components of the endothelin core system (the ligand and receptor of two endothelin signaling pathways) were studied in a phylogenetically representative set of East African cichlids: the Edn1/EdnrA pathway, which is involved in the development of craniofacial structures (Miller et al. 2000; Nair et al. 2007; Pla and Larue 2003; Fig. 1d), and the Edn3/EdnrB pathway, which is involved in the development of pigment cells (Pla and Larue 2003; Fig. 1d). We determined partial DNA sequences of ligands and receptors in 26 cichlid species from Lake Tanganyika (total sequence length per species: ca. 6,000 bp) and performed analyses to determine the evolutionary history of these loci. We show that the endothelin family members evolved under purifying selection and that they are indeed expressed in pharyngeal jaw and egg-spot tissue.

Materials and methods

Tissue Sampling and DNA/RNA Extraction

In this study, we analyzed two components of the endothelin system in 26 species of cichlid fishes from Lake Tanganyika, East Africa, belonging to 14 different tribes, thus representing a great portion of the diversity of East African cichlids (Salzburger et al. 2002, 2005). The samples were collected during two consecutive expeditions to Lake Tanganyika in 2007 and 2008 (see Supplementary Table 1 for specimen information).

Genomic DNA was extracted from white muscle tissue and digested with Proteinase K followed by sodium chloride extraction and ethanol precipitation (Bruford et al. 1998). RNA was extracted from male adult *Astatotilapia burtoni* (laboratory strain) in order to analyze the expression pattern of the studied loci (see below). Animals were kept under standard conditions (12 h light, 12 h dark; 25°C). Twelve types of tissues were dissected (brain, liver, gonads, muscle, pharyngeal jaw, maxilla, mandible, skin, anal fin [both entire anal fin as egg-spot and non-egg-spot tissue independent], caudal fin, eyes, and whole juveniles were used as positive control) and ground using a mortar and a pestle, followed by an RNA extraction procedure with Trizol (Invitrogen). Hereafter, genomic DNA was removed from the samples by DNase treatment following the DNA Free protocol (Ambion Inc, Austin, TX). The RNA concentration was measured with a Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE) in order to use the recommended amount of RNA in the subsequent reverse transcriptase (RT) step (see below).

Reverse Transcription and Real-Time PCR

Reverse transcription of the cleaned-up RNA samples was performed with the High capacity RNA-to-cDNA kit (Applied Biosystems) following the manufacturer's protocol. The housekeeping gene *GAPDH* (*Glyceraldehyde-3-phosphat-dehydro-genase*) was used under standard polymerase chain reaction (PCR) conditions with Amplitaq DNA polymerase (see below) to test if the extraction and following steps were successful. Subsequent Real-Time PCRs were conducted on a StepOnePlus Real-Time PCR System (Applied Biosystems) following the manufacturer's protocol. Final cDNA and primer concentrations of 0.5 ng/μl and 900 nM, respectively, were used together with the SYBR Green master (Rox) dye (Roche, Basel, Switzerland). Comparative C_T experiments were used to quantify the level of expression with juvenile tissue as reference tissue and *Actin* (see Supplementary Table 2 for primer info) as endogenous control. Measurements of expression levels in target tissues were normalized using the endogenous control. Each reaction was replicated three times.

Loci, PCR Amplification, and Sequencing

Four nuclear loci were investigated here: *Endothelin 1* (*Edn1*), *Endothelin 3b* (*Edn3b*), *Endothelin receptor Aa* (*EdnrAa*), and *Endothelin receptor B1a* (*EdnrB1a*). These loci were chosen based on the availability of their teleost orthologous sequences in public databases, the quality of the subsequent alignments and the quality of the PCR products. Also we choose to study only one paralog per endothelin pathway. PCR primers for both the DNA and cDNA samples are listed in Supplementary Table 2 and were initially designed based on teleost orthologs (i.e., zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), Fugu (*Takifugu rubripes*), Tetraodon (*Tetraodon nigroviridis*) and stickleback (*Gasterosteus aculeatus*); see Braasch et al. 2009). Primers were tested on *A. burtoni* DNA. Resultant *A. burtoni* sequences were blasted using a translated nucleotide query. Based on these sequences an additional set of cichlid-specific primers was designed. Standard 12.5 μl PCR reactions were conducted on a 2720 thermal cycler and/or a Veriti 96-well thermal cycler (both Applied Biosystems) with 1× Taq buffer, 0.8 mM dNTP mix, 0.0125 units REDTaq DNA polymerase (all Sigma, Saint Louis, MO), 0.2 μM of each primer, 0.5 mM $MgCl_2$, and 20 ng of template DNA. For initially unsuccessful amplifications the following, modified, protocol was used: 1× Taq buffer, 0.25 units Amplitaq DNA polymerase (both Applied Biosystems), 0.8 mM dNTP mix (Sigma), 0.2 μM of each primer, 0.1 mM $MgCl_2$ and 20 ng of template DNA. Amplification consisted of an initial denaturation step at 94°C (2 min), 30–40 cycles of 94°C (45 s), 47–57°C (45 s), and 72°C (60–120 s) followed by a final extension at 72°C (10 min).

All PCR products were visualized with ethidium bromide on a 1.5% agarose gel. Hereafter, 2.5 μl amplified PCR-product was purified with ExoSAP-IT following the manufacturer's instructions (USB, Cleveland, OH). The subsequent sequencing reactions for both forward and reverse sequences separate were performed with 1 μl Terminator Ready Reaction Mix 3.1 (Applied Biosystems), 0.625 μM primer and the purified amplification products. The conditions of the cycle sequencing reaction were as followed: 94°C (1 min), 25 cycles of 94°C (10 s), 52°C (20 s), and 60°C (4 min).

The final purification step was performed with the BigDye X-Terminator system (Applied Biosystems). SAM solution (22.5 μl), X-Terminator beads (5 μl) and H_2O (14.5 μl) was added to each sample and mixed (2,000 rpm)

for 30 min, followed by a final spin down (1,500 rpm, 2 min).

The DNA fragments were visualized on a 3130*xl* capillary sequencer (Applied Biosystems). Sequences of the DNA samples were used in subsequent analyses (see below), while sequences from the cDNA samples were only used to confirm the respective locus. All sequences have been deposited into GenBank under accession numbers JQ309683–JQ309786 (see Supplementary Table 1).

Cloning

Since one PCR reaction for the *EdnrB1a* locus did not result in a clear signal in the sequencing reaction, PCR products were cloned with a TOPO TA cloning kit (Invitrogen) following the manufacturer's protocol. PCR reactions were performed directly on the clone products with the M13 primer pair that was supplied by the manufacturer. After confirmation of the proper insert size (i.e., again with visualization with ethidium bromide on a 1.5% agarose gel), PCR products were cleaned-up with a Gen Elute PCR clean up kit (Sigma) followed by the normal sequencing protocol (see above). A minimum of six colonies were sequenced for each sample and aberrant nucleotide positions due to PCR errors and only observed in one of these sequences, were removed before the analyses.

Molecular and Phylogenetic Analyses

Sequences were aligned and edited using Codon Code Aligner 3.5.6 (CodonCode Corporation, Dedham, MA). Exon/intron boundaries were determined by homology comparisons with five teleost orthologs (see Braasch et al. 2009).

First of all, we constructed phylogenetic trees consisting of both cichlid and other teleost orthologs to confirm the exact paralog of our sequences. Therefore, cichlid *Edn* and *Ednr* sequences were translated into protein sequences and aligned with the other teleost *Edn* and *Ednr* sequences (i.e., *D. rerio*, *O. latipes*, and *G. aculeatus*; see Braasch et al. 2009 for references) using TCoffee (Notredame et al. 2000; Poirot et al. 2003). ProtTest (Abascal et al. 2005; Drummond and Strimmer 2001; Guindon and Gascuel 2003) was used to select the most appropriate model of protein evolution, which was then used in the phylogenetic analysis on the RAxML server (Stamatakis et al. 2008).

Three different data sets were constructed from the cichlid endothelin sequences: a whole data set (wds), a protein-coding data set (cod), and a non-coding data set (noncod). The number of segregating sites (S) and the K_a/K_s values for the wds and the cod were conducted with the software package DnaSP 5.10 (Librado and Rozas 2009). For the cod data sets, sliding window analyses of the K_a/K_s ratio were also performed with the same program with a window size of 25 and an overlap of 10 bp. This comparison was performed between different groups defined on the basis of existing phylogenies (e.g., outgroup species, and the three large Tanganyikan cichlid clades haplochromines, lamprologines, and ectodines; see, e.g., Salzburger et al. 2002, 2005, 2007) and the remaining samples.

Phylogenetic analyses were performed for three data sets per locus: wds, cod, and noncod along with one concatenated data set. The concatenated data set includes the four endothelin family member data sets plus one additional non-endothelin, autosomal coding gene to increase the total length to 9,451 bp (Diepeveen and Salzburger unpublished data). Due to inconsistent results for the Cyprichromini *Edn1* sequence, this single sequence was omitted from the concatenated data set in an additional analysis. Likelihood ratio tests (LRTs) were carried out with jModeltest 0.1.1 (Guindon and Gascuel 2003; Posada 2008) and Bayesian information criterion and corrected Akaike information criteria were used to determine the best-fitting models of nucleotide substitutions. These models were used in the maximum likelihood (ML) search in PAUP* 4.0 (Swofford 2002) with *T. polylepis* and *O. tanganicae* as outgroup species (see Salzburger et al. 2002). To test the robustness of the resulting topologies, we conducted bootstrap analyses with 100 replicates. Bayesian Inference of phylogeny was conducted in MrBayes 3.2 with 10.500.000 generations in total and *T. polylepis* as outgroup species (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). To determine whether the Markov chain Monte Carlo analysis had run long enough, we applied the Are We There Yet (AWTY) system (Nylander et al. 2008).

To test if the studied loci evolved under non-neutral selection, we analyzed the synonymous (silent; K_s) and nonsynonymous (amino acid-changing; K_a) substitution rates of amino acid sequences within and between the 26 species. First, we ran the data on the Selecton Server (Doron-Faigenboim et al. 2005; Stern et al. 2007). LRTs were conducted between the M8 model (β and $\omega_s \geq 1$), which enables positive selection, and the M8a (β and $\omega_s = 1$; no positive selection) and M7 (only purifying selection) models. Furthermore, we applied a Bayesian approach to predict whether a site is undergoing positive selection, as implemented in Selecton.

Next, a branch test was performed using HyPhy (Kosakovsky Pond et al. 2005) to test whether specific branches showed signs of adaptive sequence evolution. Both the obtained ML phylogeny (see above) and an independent tree based on sequences of the mitochondrial *ND2* gene (Salzburger et al. 2002, 2005) were used for reconstructions of the d_n/d_s ratios on the branches of the

phylogeny. The Suzuki–Gojobori derived adaptation selection tool was then used ($P = 0.05$) to visualize the d_n/d_s ratios on a branch-scaled tree. This method also indicates sites evolving under non-neutral evolution.

The program Codeml from the PAML (Phylogenetic Analysis by ML) 4.2 package (Yang 1997, 2007) was subsequently used to test both site- and branch-specific adaptive evolution of coding sequences. Tests of positive selected sites were conducted by performing LRT of the following model comparisons: M1a (Nearly Neutral) with M2a (Positive Selection), M7 (β) with M8 (β & ω), and M8a (β & $\omega_s = 1$) with M8. The comparison between M0 (one-ratio) and M3 (discrete) was used as a test of variable ω among sites. The naïve empirical Bayes (NEB; Nielsen and Yang 1998; Yang et al. 2000) and the Bayes empirical Bayes (BEB; Yang et al. 2005) criteria were used to calculate the posterior probabilities for site classes and to identify sites under positive selection in case the LRT was significant. In addition to the site models, the branch model (*free-ratios* model with ω allowed to vary) was applied to detect positive selection acting on particular lineages (Yang 1998; Yang and Nielsen 1998).

Finally, we studied the observed nonsynonymous amino acid substitutions in more detail. For this, we compared our data with the information known about the protein domains of both human (*Homo sapiens*) and mouse (*Mus musculus*) sequences of the proteins studied here (source: The Universal Protein Resource database http://www.uniprot.org/).

Results

RNA Expression Pattern

The gene expression levels and patterns of the endothelin family members are depicted in Fig. 2 (Real-Time PCR-results) and Supplementary Fig. 1 (PCR-results). Overall the four endothelin family members are widely expressed with *Edn3b* expression not being detectable in three tissues (i.e., liver, muscle, and pharyngeal jaw). The ligand–receptor pair *Edn1* and *EdnrAa* are co-expressed in all craniofacial bone tissues (i.e., maxilla, mandible, and pharyngeal jaw) with a three-fold increase of *EdnrAa* expression in the pharyngeal jaw, although up to three-fold increases are also found in liver, muscle, maxilla, skin, and anal fin (Fig. 2a). The combination of *Edn3b* and *EdnrB1a* expression is found in the pigment containing tissues (i.e., skin and fin tissues), for which also the highest levels of *Edn3b* expression were detected (Fig. 2b). Furthermore, all genes that are expressed in the anal fin are also expressed in both egg-spot tissue and non-egg-spot tissue.

Sequencing Results and Orthology Assessment

The total length of the captured regions for *EdnrAa* was 963 bp (of which 246 bp were protein coding (cod)), 2091 bp (528 bp cod) for *EdnrB1a*, 2449 bp (540 bp cod) for *Edn1*, and 512 bp (195 bp cod) for *Edn3b* (see Table 1).

The ML analyses of the obtained cichlid sequences and the downloaded teleost sequences revealed that cichlid

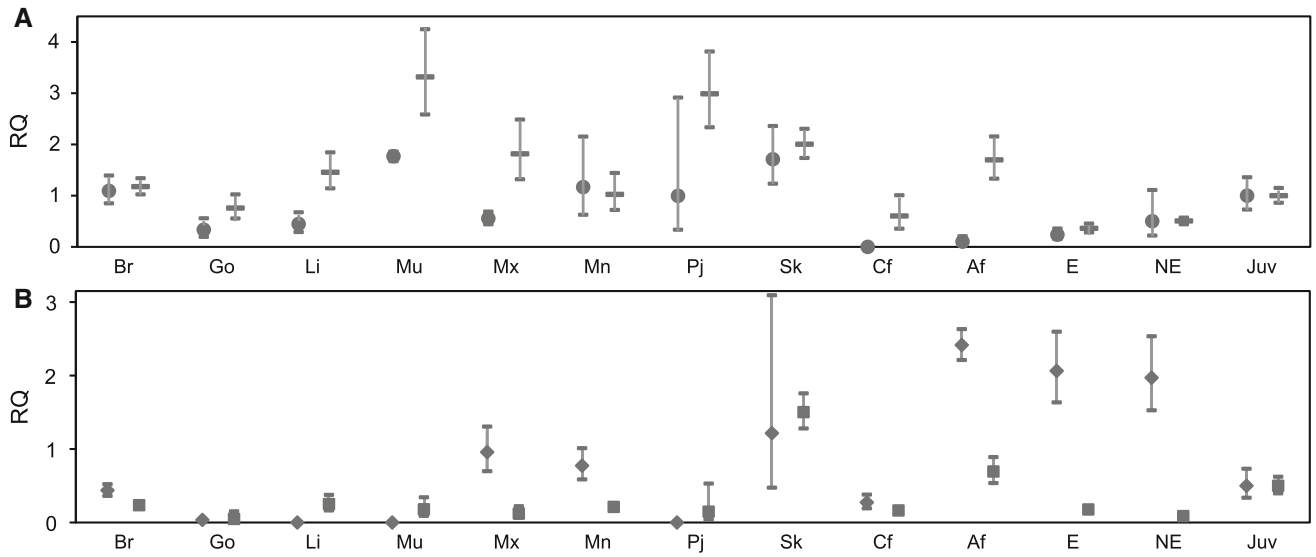


Fig. 2 Gene expression levels of the endothelin family members. **a** Relative quantitation (RQ) plot of the members of the *Edn1* (*circles*)/*EdnrAa* (*bars*) pathway. Gene expression was detected for *Edn1* and *EdnrAa* in all tissues except for *Edn1* in the caudal fin tissue. **b** Relative quantitation (RQ) plot for the *Edn3b* (*diamonds*)/*EdnrB1a* (*squares*). The two genes appear to be expressed in all tissues (except for *Edn3b* in the liver, muscle, and pharyngeal jaw tissues), albeit sometimes at low levels. For both analyses, juveniles were used as reference sample (Rf) and *Actin* as endogenous control. *Error bars* indicate the maximum and minimum expression levels determined by a 95% confidence level. *Af* anal fin, *Br* brain, *Cf* caudal fin, *E* egg-spot tissue, *Go* gonads, *Li* liver, *Mn* mandible, *Mu* muscle, *Mx* maxilla, *Ne* non-egg-spot tissue, *Pj* pharyngeal jaw, *Sk* skin

Table 1 Segregating sites and K_a/K_s values

Data set	Length[a]	S[b]	K_a/K_s[c]
Edn1	2449 (540)	16.1 (14.7)	0.2312
Edn3b	512 (195)	14.1 (12.9)	0.5700
EdnrAa	963 (246)	16.1 (9.4)	0.0880
EdnrB1a	2091 (528)	17.4 (3.3)	0.0000

[a] Total length of the sequenced region and protein coding regions (in parentheses) in base pairs

[b] Percentage of segregating sites for total sequenced region and protein coding regions (in parentheses)

[c] Nonsynonymous/synonymous substitution rate ratio for protein coding regions

EdnrAa, *EdnrB1a*, *Edn1*, and *Edn3b* each is more closely related to their teleost orthologs than to other teleost paralogs (Supplementary Fig. 2a, b). This indicates that the obtained cichlid sequences are indeed the presumed endothelin genes.

Phylogenetic Analysis

The wds consensus Bayesian phylogenies (BAY) with ML bootstrap values and BAY posterior probabilities are depicted in Supplementary Fig. 3a–e. In most data sets, both ML and BAY methods revealed a similar topology that is largely in agreement with available species trees (Salzburger et al. 2002, 2005, 2007). This topology consists of the Tylochromini (*T. polylepis*), Tilapiini (*O. tanganicae*), and the endemic Bathybatini, Boulengerochromini, and Trematocarini being placed in a basal position with respect to the substrate spawning Lamprologini, the Eretmodini and the 'C-lineage' (Cyphotilapiini, Cyprichromini, Ectodini, Limnochromini, Perissodini, and the Haplochromini/Tropheini clade; Clabaut et al. 2005; Day et al. 2008; Nishida 1991; Salzburger et al. 2002, 2005; Takahashi 2003).

Although, relationships between tribes are generally not recovered by the individual gene trees (e.g., Edn3b tree; Supplementary Fig. 3c), the data sets do contain an adequate amount of phylogenetic signal to group members of a tribe together. This is further confirmed by the moderate to high support values for the monophyly of the major lineages, while support was typically low for many interspecific relationships—reflecting a limited number of variable sites and in the case of Edn3b a small size of the data set (ca. 500 bp).

As in previous phylogenetic analyses, the position of *Eretmodus cyanostictus* remains unclear (Clabaut et al. 2005; Day et al. 2008; Salzburger et al. 2002); either unresolved as part of a major polytomy (Supplementary Fig. 3b–d) or even at a derived position within the C-lineage (Supplementary Fig. 3a). A similar observation was made for *Cyprichromis leptosoma*; a position as sister group to the perissodines was alternated with polytomous positions and even a position within the Lamprologines in the Edn1 phylogeny. Although this clustering was not supported by the cod and noncod analyses (data not shown), the same signal was retrieved when the wds was re-analyzed in parts of 800 bp, indicating that convergence via, e.g., lateral gene transfer or an hybridization event could have occurred.

Segregating Sites and K_a/K_s Values

The number of segregating sites (S) and the K_a/K_s values for each of the studied loci are depicted in Table 1. The lowest value for S was observed for the *EdnrB1a* cod data set, while the *EdnrB1a* wds data set was characterized by the highest value for this parameter. All the endothelin family members were characterized by K_a/K_s ratios <1 indicating purifying selection. Remarkable low K_a/K_s ratios were found for both endothelin receptors. When comparing the K_a/K_s ratios of ligands versus receptors we found significantly higher ratios in the ligands (unpaired T test, $P = 0.02$; Fig. 3a). Furthermore, the K_a/K_s ratios observed in the ligands is significantly higher than the mean K_a/K_s ratios observed by Baldo et al. (2011; Fig. 3b) in a pairwise comparison of 1,098 EST's between *Ophthalmotilapia ventralis* and *A. burtoni* (unpaired T test, $P < 0.01$).

Selection Pressure on Sites and Lineages

Even though the endothelin family members have conserved functions among all vertebrates, which would be a cause for selection against deleterious mutations, the co-opted functions of these genes during the development

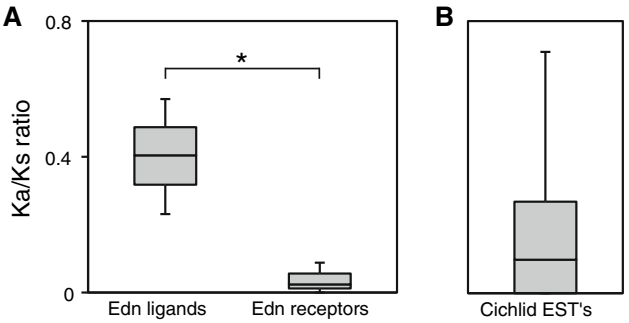


Fig. 3 K_a/K_s comparison between the endothelin ligands, endothelin receptors, and cichlid EST's. **a** The K_a/K_s values for the Edn ligands and Edn receptors. Higher values were found for the ligands than for the receptors ($P = 0.02$; unpaired T test; $N = 3$). **b** The K_a/K_s values for the 1,055 EST's derived from a genome wide approach based on a *O. ventralis*/*A. burtoni* comparison performed by Baldo et al. (2011). In this data set, 43 outliers were detected. Outlier values ranged between 0.74 and 2.66 and were omitted from the figure, but not from the analysis. Depicted are the mean, first and third quartile, and the minimum and maximum K_a/K_s values for each of the three groups

and evolution of two of the cichlids key innovations might be of such importance that this would have left signs in the genome by means of positive selection. If the latter is the case, differences in the strength of the selection pressure between lineages are to be expected, especially for the Edn3b/EdnrB1a pathway, since not all species possess egg-spots and/or colorful body patterns.

Table 2 summarizes the results of the various tests that were performed to infer sites and/or lineages under non-neutral evolution. Negatively selected sites were found in all loci. None of the LRT's of the PAML (BEB) and Selecton analyses for the endothelin family members were significant, suggesting that these loci evolved under purifying selection. Trends toward significance were found for *EdnrAa* and *Edn1* with Selecton; while with NEB in PAML 23 and 15 positively selected sites were identified for *Edn1* and *Edn3b*, respectively.

Positive selected branches were found with HyPhy for all loci except *EdnrAa* (Table 2). In *EdnrAa*, only one species (*C. macrops*) showed a slightly elevated K_a/K_s or d_n/d_s ratio ($d_n/d_s = 0.6$ (PAML) $-$ 0.9 (HyPhy)). Most species and clades showed d_n/d_s ratios around zero in *EdnrB1a* except for three Tropheini/Haplochromini species (*L. labiatus*, *S. diagramma*, *T. moori*), one Perissodini species (*P. straeleni*) and *C. frontosa,* for which elevated d_n/d_s ratios were detected. Nonetheless, other species of the Tropheini and Perissodini did not show elevated d_n/d_s ratios. In the branch-specific PAML analysis, no signs of positive selection were found in EdnrB1a.

Besides species-specific increases of K_a/K_s and d_n/d_s ratios the analysis of the *Edn1* data set also showed lineage-specific increases. All major branches towards the derived clades showed increased K_a/K_s and d_n/d_s ratios.

Both lineage and species-specific increases of K_a/K_s and d_n/d_s ratios were also found in *Edn3b*. Again several internal branches showed increased K_a/K_s and d_n/d_s ratios. The results for the internal branches were not unambiguous due to incongruent results of the separate analyses. The Perissodini were the only clade that consistently showed the same strong signal (i.e., positive selection) in all analyses. Furthermore, the *Edn3b* sequences of the following species showed signs of positive selection; *G. permaxillaris*, *N. pulcher*, *B. microlepis*, *T. nigrifrons*, *O. tanganicae* and the clade consisting of *S. diagramma* and *P. curvifrons*.

Table 2 Overview of selection pressure on sites and branches

Locus	Selecton/ HyPhy n.s.s.[a]	Selecton/ Hyphy p.s.s.[b]	PAML[c] NEB/BEB p.s.s.	HyPhy/ PAML p.s.b.[d]	Lineages
Edn1	33/5	– (4)/–	23/–	Yes/yes	All
Edn3b	14/3	–/–	15/–	Yes/yes	All
EdnrAa	16/3	–(1)/–	–/–	–/–	–
EdnrB1a	37/8	–/–	–/–	Yes/–	Cypho, Periss, Troph[e]

[a] Negative selected sites ($P \leq 0.05$)

[b] Positive selected sites ($P \leq 0.05$); values in brackets: Selecton $P \leq 0.10$

[c] Models M2a and M8; NEB ($P \leq 0.01$); BEB ($P \leq 0.05$); for *Edn1* only model M2a

[d] Positive selected branches

[e] Cyphotilapiini, Perissodini, Tropheini

Sliding Window Analysis

The results of the sliding window analyses are depicted in Fig. 4. For EdnrAa there is only one region found with elevated omega values (K_a/K_s) in all analyzed groups. Two regions of weakly elevated omega values were observed for EdnrB1a. It seems that the Haplochromini have the highest omega value at the first peak, while the second peak is found in more groups (i.e., the two outgroup species, Haplochromini, Lamprologini, and Perissodini). Nonsynonymous substitutions were found in both the transmembrane and the extracellular and cytoplasmic domains of both the endothelin receptors (see also Supplementary Table 3).

For the Edn1 and Edn3b data sets, several peaks were identified in the preproendothelin for which all groups surveyed show elevated omega values for at least one of these peaks. In all groups, we have found species or lineage-specific nonsynonymous substitutions that coincide with these peaks. For both endothelins, the highest scores are found for the outgroup species *T. polylepis*. Furthermore, the results for the Perissodini show two strong peaks, caused by two lineage-specific substitutions. One remarkable nonsynonymous substitution is found in *E. cyanostictus*; a transversion led to a Threonine–Serine substitution at the second amino acid position of the functional protein. A Serine at this position is also observed in the Edn1, 2, and 4 proteins.

Discussion

In this study, we investigated the molecular evolution of two endothelin ligand–receptor pairs (Edn1–EdnrAa and Edn3b–EdnrB1a) in East African cichlid fishes. We performed a variety of phylogenetic and molecular analyses in a set of 26 species and assessed the expression of the genes under study in twelve different tissues of a reference species. Here we discuss our results with respect to the observed gene expression pattern, selection pressure, and their functional and evolutionary implications.

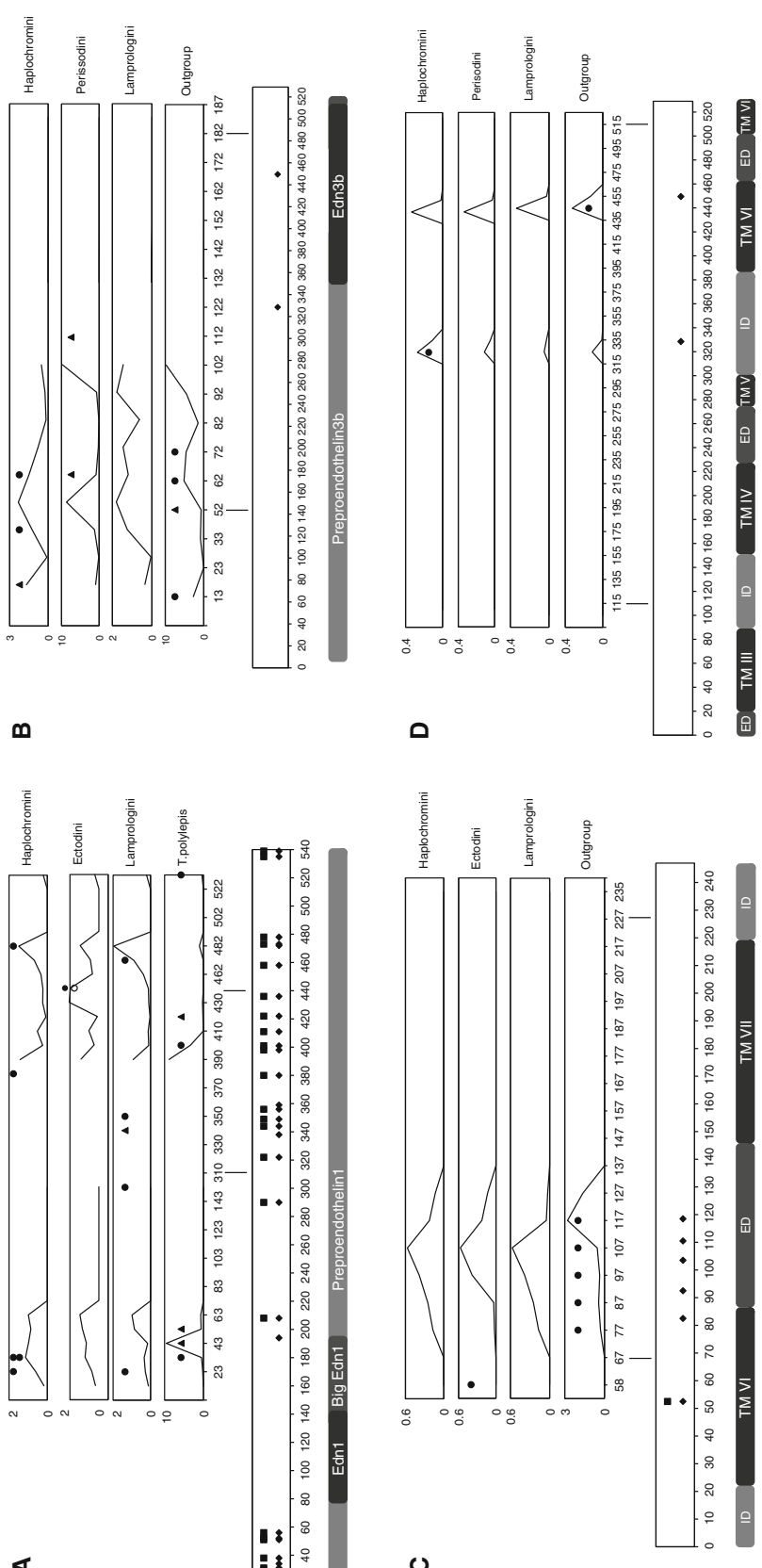


Fig. 4 Sliding window plots of the endothelin ligands or receptors Species-specific nonsynonymous substitutions are displayed by *filled circles* in the sliding window plots, while *filled triangles* depict lineage-specific nonsynonymous substitutions. Comparisons were performed between the members of the specified group and the remaining samples. Window sizes were 25 bp with 10 bp overlap. *Bars* show corresponding regions between the sliding windows and the overview of putative positive selected sites (*filled squares*) and the nonsynonymous substitutions (*filled diamonds*). Functional domains are depicted as *gray*-tone blocks. *ID* intracellular domain, *TM* transmembrane domain, *ED* extracellular domain. **a** Elevated K_a/K_s ratio, positive selected sites (i.e., PAML NEB, $P \leq 0.01$) and nonsynonymous sites are mainly found in the preproendothelin 1, with one nonsynonymous substitution found in big endothelin 1. The sequenced part of the signaling protein is depicted by the *filled circle symbol*. **b** One nonsynonymous substitution was found in the mature Edn3b protein. The positive selected sites were found with the NEB ($P \leq 0.01$). **c** Nonsynonymous substitutions were found in the transmembrane and extracellular parts of EdnrAa. The positive selected site was retrieved with Selecton ($P \leq 0.10$). **d** The intracellular and transmembrane parts of EdnrB1a hold only two nonsynonymous substitutions

Gene Expression Assays

Our gene expression assays reveal that the members of the two studied endothelin pathways are expressed in a broad variety of tissues in the cichlid *A. burtoni* (Fig. 2, Supplementary Fig. 1), which is in concordance with the wide range of functions known for the endothelin family members. Most studied and described functions of the endothelin system are related to the blood pressure regulation functions of *Edn1* in adult vertebrates (e.g., Yanagisawa et al. 1988) and the functions during early development of neural crest derivatives (e.g., Miller et al. 2000; Parichy et al. 2000). Yet, the endothelin system is not limited to vasoconstriction or early development, since several neural crest derivatives such as pigmentation and the (dentition on the cichlids') pharyngeal jaws are produced and maintained throughout adulthood. Here we show that the two endothelin pathways under study are differently expressed in the pharyngeal jaw apparatus and anal fin egg-spots. We find *Edn1* and *EdnrAa* co-expression in the pharyngeal jaw, while *Edn3b* and *EdnrB1a* are co-expressed in the anal fin and egg-spot tissue (Fig. 2, Supplementary Fig. 1). This is in agreement with previously described endothelin functions: *EdnrB1*/*EdnrB2* duplicates and *Edn3* are known to be involved in pigmentation in a range of vertebrates including zebrafish (e.g., Parichy et al. 2000; Pla and Larue 2003); whereas a correlation between *Edn1*, *EdnrA* duplicates and the pharyngeal bone development has been uncovered in zebrafish (Kimmel et al. 2003; Pla and Larue 2003).

Deciphering the Selection Pressure Acting Upon the Endothelin Family Members in Cichlids

Patterns of DNA sequence variation are important indicators of the mode of selection acting upon a locus. In particular, the ratio of non-synonymous substitutions per non-synonymous site (K_a) to synonymous substitutions per synonymous site (K_s) is considered a powerful tool in the study of adaptive evolution (Hughes 1997). Under neutrality synonymous and non-synonymous substitutions are assumed to occur at equal rates, resulting in a ratio close to one. Purifying selection will remove deleterious mutations, thus, leading to a decrease of the K_a/K_s ratio, while positive selection will favor advantageous mutations, which is characterized by a K_a/K_s ratio exceeding one. Theory predicts that both positive and purifying selection can occur after gen(om)e duplication events. Although the fate of most duplicated genes is nonfunctionalization (e.g., Ohno 1970; Zhang 2003), models of adaptive evolution by gene duplication (Hughes 1999) predict that soon after a duplication event positive selection acts on mutations that lead to new or enhanced functions (i.e., neofunctionalization). Subsequently, once having evolved, these functions are maintained by purifying selection acting at the gene level. Neofunctionalization is, however, not the only outcome that leaves a signature of purifying selection in the genes after a duplication event. Selection against deleterious mutations could also occur subsequent to subfunctionalization. Subfunctionalization is known to have occurred in the zebrafish *EdnrA* paralogs, which have partially redundant functions during lower jaw formation (Nair et al. 2007).

Here we show that the endothelin ligand and receptor family members evolved under purifying selection. Interestingly, substantial variation between loci was observed in the K_a/K_s comparisons. While we detected very low K_a/K_s values for the two endothelin receptors, the two endothelin ligands showed elevated rates (Table 1), even when compared to genome wide comparisons between two mouthbrooding species from Lake Tanganyika (Fig. 3). The accelerated rate of protein evolution observed in the ligands is similar to what has been observed for *bmp4* by Terai et al. (2002b). In both type of genes, the variation accounting for the elevated rate of protein evolution is not observed in the functional protein domain, but rather in the pro-domain/preproendothelin parts. Therefore, the mature proteins are likely to be functional, although posttranslational regulation or cleavage efficiency could have been affected (see next section; Terai et al. 2002b).

Implications on the Functional Level of Endothelins and Their Receptors in Cichlids

As mentioned before, the extracellular section and loops of the endothelin receptors are used for ligand binding, while interaction with the G protein takes place at the intracellular loops (Masaki et al. 1999; Orry and Wallace 2000). The extracellular part can be further divided into an "address" (ligand–receptor selectivity) and a "message" (ligand–receptor binding) part (Masaki et al. 1999; Sakamoto et al. 1993). In both endothelin receptors under study, almost all non-synonymous changes (detected in *T. polylepis*, *C. macrops*, and *O. tanganicae*) were confined to the "address" region. The additional substitution in EdnrB1a (specific for *A. burtoni*) concerns the cytoplasmic part (G protein interaction). Thus, most changes in the endothelin receptors of cichlids are found in the parts responsible for ligand–receptor interaction, supporting the co-evolution theory of Braasch et al. (2009).

The majority of the changes in the ligands affected the preproendothelin part with only two exceptions: one substitution occurred in the big endothelin 1 of *N. furcifer* and one in the Edn3b mature protein part of *E. cyanostictus*. This last substitution is found at the second amino acid of the mature protein, a site that is also variable among other

endothelins (Braasch et al. 2009). A serine at this position is commonly observed in the Edn1, 2, and 4 proteins among teleost and other vertebrates, but—so far—not in any of the Edn3 duplicates.

Because of the particular way of protein maturation (preproendothelin > big endothelin > endothelin), it might be speculated that the observed mutations in the preproendothelin part could affect posttranslational regulation. At least for *endothelin 1*, it is known that the regulation is ultimately controlled by the rate of transcription. *Cis*-acting elements both in the promoter and further upstream, epigenetic mechanisms and several other types of regulatory interactions (e.g., calcium levels, extracellular hypertonicity) do play an important role, too (reviewed in Stow et al. 2011). It was further shown that several regulatory interactions take place directly on the *Edn1* mRNA (e.g., through GAPDH, microRNA), which can either be stimulating or repressing (Stow et al. 2011).

The observed changes in the preproendothelin parts could—in principle—also affect the cleaving process. We found that the pairs of basic amino acids, where the actual cleavage takes place, stay unaffected in cichlids. However, substitutions in the preproendothelin part could still affect the efficiency of cleavage and with that the concentration of the mature protein (the mature protein itself stays unaffected and would function normally). Definite conclusions about the functional effects of the nonsynonymous changes can, however, only be made on the basis of functional assays (e.g., Hoekstra et al. 2006).

Evolutionary Implications

The pharyngeal jaw apparatus and anal fin egg-spots are two putative evolutionary key innovations of cichlids. A key innovation is a novel trait in evolution that enables the exploitation of new niches and, hence, could initiate adaptive radiation (Gavrilets and Losos 2009; Schluter 2000; Simpson 1953). In this way, the trait is directly responsible for the evolutionary success of a radiating group. Little is known, however, about how the relationship between trait evolution and organismal diversification is reflected in the genes and genetic pathways underlying such innovations. Often it is assumed that genes underlying (ecologically) important traits show signs of non-neutral evolution. In cichlids, for example, the downstream color gene *csf1ra*, which is expressed in egg-spots, appears to have evolved adaptively in the direct ancestor of the egg-spot bearing haplochromines (Salzburger et al. 2007). An accelerated protein evolution has also been found in *bmp4*, which is involved in (pharyngeal) bone formation (Fraser et al. 2009; Terai et al. 2002b). It is unclear how common such signatures of positive selection are in the genes underlying evolutionary innovations. Obviously, many more genes are likely to be involved in the evolution of a novel trait than the single/few positively selected genes in the above-mentioned examples. Also, in cases where genes underlying a novel trait are part of a conservative pathway that is co-opted, one would expect a signature of purifying rather than positive selection. This scenario could have happened with the endothelin family members. For instance, when the conserved pathway involved in the development of neural crest-derived tissues was co-opted during the evolution of key innovations in cichlid fishes.

Speciation and organismal diversification can be greatly facilitated by the concurrent action of natural and sexual selection. This is particularly the case when an ecologically important, hence naturally selected trait, is at the same time the target of mate choice. In sticklebacks, for example, it has been shown that body size, which is influenced by the foraging habitat, also acts as trigger in mate preference (Hatfield and Schluter 1999; Nagel and Schluter 1998). Feulner et al. (2009) suggested that the electric organ discharge of weakly electric fish is a magic trait serving for both electrolocation (foraging) and mate recognition (mate choice). Another single trait that is both under disruptive natural selection and acts as a cue for assortative mating is the color patterns of the coral reef fishes of the genus *Hypoplectrus* (Puebla et al. 2007).

Just as natural and sexual selection are both affecting a single trait in these examples, the endothelin family as a whole might be affected by both natural and sexual selection. The endothelin family members, thus, form a link between a naturally selected trait (mouth morphology, including the pharyngeal jaws) and a sexually selected one (coloration)—via their underlying genes and genetic pathways. This makes the endothelin gene family a strong candidate for a more thorough functional, developmental, and evolutionary examination in cichlid fishes.

Conclusions

In this work, we studied the molecular evolution of two ligand-receptor pairs of the endothelin system (Edn1/EdnrAa and Edn3b/EdnrB1a) in a representative set of 26 East African cichlid species. Sequences of the endothelin family members show signs of purifying selection in cichlids, a common force observed after gen(om)e duplication events. The two ligands show an accelerated rate of protein evolution in comparison to the receptors, most likely caused by the highly variable preproendothelin parts. No evidence was found for one of the endothelin pathways to have evolved under positive selection in one or more lineages. While the amino acids of the mature endothelins are rather conserved in cichlids, the variable preproendothelin parts could influence posttranslational regulation.

We also found several changes in the endothelin receptors that could possibly affect the ligand-receptor binding functions. Future functional studies should test these hypotheses. Expression pattern analyses confirm that Edn1/EdnrAa are co-expressed at a relatively high expression level in the pharyngeal jaw tissue, whereas Edn3b/EdnrB1a at relatively high levels in skin and anal fin tissue (including the anal fin egg-spots). Because of their anticipated function during the morphogenesis of the (pharyngeal) jaw apparatus and coloration and pigmentation, the endothelin gene family members form an important link between these naturally and sexually selected traits in cichlids.

Acknowledgments We would like to thank Brigitte Aeschbach and Nicolas Boileau for help and advice on the lab work, Michael Matschiner for advice on the phylogenetic analyses, Emilia Santos for providing the "egg-spots" and "non-egg-spots" RNA samples and Laura Baldo for providing the K_a/K_s values from the cichlid EST comparison. We are very grateful to Ingo Braasch and the anonymous reviewers for their valuable comments and suggestions on earlier version of this manuscript. This study was supported by the European Research Council (Starting Grant "INTERGENADAPT" to WS) and the Swiss National Science Foundation (Grant 3100A0_122458 to WS).

Conflict of interest The authors declare that they have no conflict of interest.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

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

Albertson RC, Streelman JT, Kocher TD, Yelick PC (2005) Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc Natl Acad Sci USA 102:16287–16292

Baldo L, Santos EM, Salzburger S (2011) Comparative transcriptomics of eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. Genome Biol Evol 3:443–455

Barlow GW (2000) The cichlid fishes: nature's grand experiment in evolution. Perseus publishing, Cambridge

Baynash AG, Hosoda K, Giaid A, Richardson JA, Emoto N, Hammer RE, Yanagisawa M (1994) Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. Cell 79:1277–1285

Bonano M, Tríbulo C, De Calisto J, Marchant L, Sánchez SS, Mayor R, Aybar MJ (2008) A new role for the endothelin-1/endothelin-A receptor signaling during early neural crest specification. Dev Biol 323:114–129

Braasch I, Volff J-N, Schartl M (2009) The endothelin system: evolution of vertebrate-specific ligand–receptor interactions by three rounds of genome duplication. Mol Biol Evol 26:783–799

Bruford MW, Hanotte O, Brookfield JFY, Burke T (1998) Multilocus and single-locus DNA fingerprinting. In: Hoelzel AR (ed) Molecular genetic analysis of populations: a practical approach. Oxford University Press, Oxford, pp 287–336

Chung I-H, Yamaza T, Zhao H, Chuong P-H, SHi S, Chai Y (2009) Stem cell property of postmigratory cranial neural crest cells and their utility in alveolar bone regeneration and tooth development. Stem Cells 27:866–877

Clabaut C, Salzburger W, Meyer A (2005) Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. J Mol Evol 61:666–681

Clouthier DE, Hosoda K, Richardson JA, Williams SC, Yanagisawa H, Kuwaki T, Kumada M, Hammer RE, Yanagisawa M (1998) Cranial and cardiac neural crest defects in endothelin-A receptor-deficient mice. Development 125:813–824

Coulter GW (1991) Lake Tanganyika and its life. British Museum (Natural History) and Oxford University Press, Oxford

Day JJ, Cotton JA, Barraclough TG (2008) Tempo and mode of diversification of Lake Tanganyika cichlid fishes. PLOS One 3:e1730. doi:10.1371/journal.pone.0001730

Doron-Faigenboim A, Stern A, Bacharach E, Pupko T (2005) Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics 21:2101–2103

Drummond A, Strimmer K (2001) PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics 17:662–663

Duftner N, Sefc KM, Koblmüller S, Salzburger W, Taborsky M, Sturmbauer C (2007) Parallel evolution of facial stripe patterns in the *Neolamprologus brichardi/pulcher* species complex endemic to Lake Tanganyika. Mol Phylogenet Evol 45:706–715

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

Elmer KR, Kusche H, Lehtonen TK, Meyer A (2010) 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 365:1763–1782

Feulner PGD, Plath M, Engelmann J, Kirschbaum F, Tiedemann R (2009) Electrifying love: electric fish use species-specific discharge for mate recognition. Biol Lett 5:225–228

Fraser GJ, Bloomquist RF, Streelman JT (2008) A periodic pattern generator for dental diversity. BMC Biol 6:32. doi:10.1186/1741-7007-6-32

Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman TJ (2009) An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol 7:e31. doi:10.1371/journal.pbio.1000031

Fryer G, Iles TD (1972) The cichlid fishes of the Great Lakes of Africa: their biology and Evolution. Oliver & Boyd, Edinburgh

Gans C, Northcutt RG (1983) Neural crest and the origin of vertebrates: a new head. Science 220:268–273

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

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum-likelihood. Syst Biol 52:696–704

Hall BK (1999) The neural crest in development and evolution. Springer-Verlag, New York

Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873

Hert E (1989) The function of egg-spots in an African mouth-brooding cichlid fish. Anim Behav 37:726–732

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

Hosoda K, Hammer RE, Richardson JA, Baynash AG, Cheung JC, Giaid A, Yanagisawa M (1994) Targeted and natural (piebald-lethal) mutations of endothelin-B receptor gene produce

megacolon associated with spotted coat color in mice. Cell 79:1267–1276
Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755
Hughes AL (1997) Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. Mol Biol Evol 14:1–5
Hughes AL (1999) Adaptive evolution of genes and genomes. Oxford University Press, Oxford
Hunter JP (1998) Key innovations and the ecology of macroevolution. Trends Ecol Evol 13:31–36
Keenleyside MHA (1991) Cichlid fishes: behaviour, ecology and evolution. Chapman & Hall, London
Kimmel CB, Ullmann B, Walker M, Miller CT, Crump JG (2003) Endothelin 1-mediated regulation of pharyngeal bone development in zebrafish. Development 130:1339–1351
Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nat 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
Kornfield I, Smith PF (2000) African cichlid fishes: model systems for evolutionary biology. Annu Rev Ecol Syst 31:163–196
Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679
Kurihara Y, Kurihara H, Suzuki H, Kodama T, Maemura K, Nagai R, Oda H, Kuwaki T, Cao W-H, Kamada N, Jishage K, Ouchi Y, Azuma S, Toyoda Y, Ishikawa T, Kumada M, Yazaki Y (1994) Elevated blood-pressure and craniofacial abnormalities in mice deficient in endothelin-1. Nature 368:703–710
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452
Liem KF (1973) Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. Syst Biol 22:425–441
Masaki T (2004) Historical review: endothelin. Trends Pharmacol Sci 25:219–224
Masaki T, Ninomiya H, Sakamoto A, Okamoto Y (1999) Structural basis of the function of endothelin receptor. Mol Cell Biochem 190:153–156
McGill GG, Horstmann M, Widlund HR, Du J, Motyckova G, Nishimura E, Lin Y-L, Ramaswamy S, Avery W, Ding H-F, Jordan S, Jackson I, Korsmeyer S, Golub T, Fisher D (2002) Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability. Cell 109:707–718
Meyer A (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. Trends Ecol Evol 8:279–284
Meyer A, van de Peer Y (2005) From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). Bioessays 27:937–945
Miller CT, Schilling TF, Lee KH, Parker J, Kimmel CB (2000) Sucker encodes a zebrafish endothelin-1 required for ventral pharyngeal arch development. Development 127:3815–3828
Mrowka W (1987) Egg stealing in a mouthbrooding cichlid fish. Anim Behav 35:923–925
Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol Biol 11:116. doi:10.1186/1471-2148-11-116
Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. Evolution 52:209–218
Nair S, Li W, Cornell R, Schilling TF (2007) Requirements for endothelin type-A receptors and endothelin-1 signaling in the facial ectoderm for the patterning of skeletogenic neural crest cells in zebrafish. Development 134:335–345
Nielsen R, Yang ZH (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936
Nishida M (1991) Lake Tanganyika as an evolutionary reservoir of old lineages of East-African cichlid fishes: inferences from allozyme data. Cell Mol Life Sci 47:974–979
Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217
Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583
Ohno S (1970) Evolution by gene duplication. Springer Verlag, New York
Orry AJW, Wallace BA (2000) Modeling and docking the endothelin G-protein-coupled receptor. Biophys J 79:3083–3094
Ozeki H, Kurihara Y, Tonami K, Watatani S, Kurihara H (2004) Endothelin-1 regulates the dorsoventral branchial arch patterning in mice. Mech Dev 121:387–395
Parichy DM, Mellgren EM, Rawls JF, Lopes SS, Kelsh RN, Johnson SL (2000) Mutational analysis of the endothelin receptor b1 (rose) during neural crest and pigment pattern development in the zebrafish *Danio rerio*. Dev Biol 227:294–306
Pla P, Larue L (2003) Involvement of endothelin receptors in normal and pathological development of neural crest cells. Int J Dev Biol 47:315–325
Poirot O, O'Toole E, Notredame C (2003) Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. Nucleic Acids Res 31:3503–3506
Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25:1253–1256
Puebla O, Bermingham E, Guichard F, Whiteman E (2007) Colour pattern as a single trait driving speciation in Hypoplectrus coral reef fishes? Proc R Soc Lond B 274:1265–1271
Rawls JF, Mellgren EM, Johnson SL (2001) How the zebrafish gets its stripes. Dev Biol 240:301–314
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
Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574
Sakamoto A, Yanagisawa M, Sakurai T, Nakao K, Toyo-oka T, Yano M, Masaki T (1993) The ligand-receptor interactions of the endothelin systems are mediated by distinct "message" and "address" domains. J Cardiovasc Pharmacol 22:S113–S116
Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol Ecol 18:169–185
Salzburger W, Meyer A, Baric S, Veheyen 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. doi:10.1186/1471-2148-5-17
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. doi:10.1186/1741-7007-5-51
Schluter D (2000) The ecology of adaptive radiation. Oxford University Press, Oxford
Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. Proc R Soc Lond B 273:1987–1998
Seehausen O, Alphen v JJM, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. Science 277:1808–1811

Shimeld SM, Holland PWH (2000) Vertebrate innovations. Proc Natl Acad Sci USA 97:4449–4452

Simpson GG (1953) The major features of evolution. Columbia University Press, New York

Slack JMW (2001) Essential developmental biology. Blackwell Publishing, Oxford

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web-servers. Syst Biol 57:758–771

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res 35:W506–W511

Stow LR, Jacobs ME, Wingo CS, Cain BD (2011) Endothelin-1 gene regulation. FASEB J 25:16–28

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). VERSION 4. Sinauer Associates, Sunderland

Takahashi T (2003) Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). Ichthyol Res 50:367–382

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, 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

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

Theis A, Salzburger W, Egger B (2012) The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni*. PloS ONE 7(1): e29878. doi:10.1371/journal.pone.0029878

Turner GF, Burrows MT (1995) A model of sympatric speciation by sexual selection. Proc Biol Sci Lond B 260:287–292

Wickler W (1962) Egg-dummies as natural releasers in mouth-breeding cichlids. Nature 194:1092–1093

Yanagisawa M, Kurihara H, Kimura S, Tomobe Y, Kobayashi M, Mitsui Y, Yazaki Y, Goto K, Masaki T (1988) A novel potent vasoconstrictor peptide produced by vascular endothelial cells. Nature 332:411–415

Yanagisawa H, Clouthier DE, Richardson JA, Charite J, Olson EN (2003) Targeted deletion of a branchial arch-specific enhancer reveals a role of *dHAND* in craniofacial development. Development 130:1069–1078

Yang ZH (1997) PAML, A program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556

Yang ZH (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

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

Yang ZH, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Yang ZH, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

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

Zhang JZ (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18:292–298

CHAPTER 2

MOLECULAR CHARCTERIZATION OF TWO ENDOTHELIN PATHWAYS IN EAST AFRICAN CICHLID FISHES

SUPPORTING MATERIAL

Supplementary Table 1

Specimen information and GenBank accession numbers

Species	Tribe	GenBank accession numbers			
		Edn1	Edn3b	EdnrAa	EdnrB1a
Bathybates graueri	Bathybatini	JQ309744	JQ309717	JQ309693	JQ309774
Boulengerochromis microlepis	Boulengerochromini	JQ309751	JQ309722	JQ309708	JQ309763
Cyphotilapia frontosa	Cyphotilapiini	JQ309752	JQ309713	JQ309700	JQ309780
Cyprichromis leptosoma	Cyprichromini	JQ309742	JQ309711	JQ309697	JQ309773
Callochromis macrops	Ectodini	JQ309743	JQ309733	JQ309685	JQ309764
Cyathopharynx furcifer	Ectodini	JQ309753	JQ309723	JQ309705	JQ309765
Eretmodus cyanostictus	Eretmodini	JQ309754	JQ309724	JQ309684	JQ309784
Astatotilapia burtoni	Haplochromini	JQ309755	JQ309725	JQ309706	JQ309783
Ctenochromis horei	Haplochromini	JQ309756	JQ309726	JQ309701	JQ309781
Altolamprologus fasciatus	Lamprologini	JQ309758	JQ309728	JQ309696	JQ309785
Lepidiolamprologus elongatus	Lamprologini	JQ309757	JQ309727	JQ309695	JQ309782
Neolamprologus furcifer	Lamprologini	JQ309745	JQ309718	JQ309707	JQ309775
Neolamprologus pulcher	Lamprologini	JQ309760	JQ309719	JQ309686	JQ309776
Variabilichromis moorii	Lamprologini	JQ309735	JQ309734	JQ309691	JQ309771
Gnathochromis permaxillaris	Limnochromini	JQ309746	JQ309712	JQ309687	JQ309777
Limnochromis staneri	Limnochromini	JQ309759	JQ309710	JQ309692	JQ309762
Perissodus microlepis	Perissodini	JQ309747	JQ309732	JQ309688	JQ309768
Plecodus straeleni	Perissodini	JQ309736	JQ309709	JQ309698	JQ309769
Oreochromis tanganicae	Tilapiini	JQ309748	JQ309729	JQ309689	JQ309778
Trematocara nigrifrons	Trematocarini	JQ309749	JQ309720	JQ309690	JQ309761
Lobochilotes labiatus	Tropheini	JQ309737	JQ309714	JQ309703	JQ309766
Petrochromis famula	Tropheini	JQ309750	JQ309721	JQ309694	JQ309779
Pseudosimochromis curvifrons	Tropheini	JQ309738	JQ309715	JQ309704	JQ309770
Simochromis diagramma	Tropheini	JQ309739	JQ309731	JQ309702	JQ309767
Tropheus moori	Tropheini	JQ309740	JQ309716	JQ309699	JQ309786
Tylochromis polylepis	Tylochromini	JQ309741	JQ309730	JQ309683	JQ309772

Supplementary Table 2

Primer names and Sequences

Locus	Forward primer	5' - 3' Sequence	Reverse Primer	5' - 3' Sequence
EdnrAa	EdnrAa_C_4F_21[a]	TCAGTGAGCACCTCAAACAG	EdnrAa_C_6R_34	ATTGATGCCGAAGTAATCCA
	EdnrAa_C_5F27	TTCCCTCGTCCTCATCTTTG	EdnrAa_C_7R_36[a,b]	TAACACCAGCAGCACAAGCA
	EdnrAa_5F34[b]	GTCCTCATCTTTGCCCTCTG		
EdnrB1a	EdnrB1aC_4F_14	CAAAGGACTGGTGGCTGTTC	EdnrB1aC_6R_68[a,b]	CGATGGGGTTGATACAGGAG
	EdnrB1a_2F20[a]	CATTCGGGGTGACTCTGTG	EdnrB1a_3R96	GCAGCAAGCAGATCCTCAA
	EdnrB1a_5F3[b]	GAGGGAGGTGGCGAAGACAG		
Edn1	Edn1_C1F_61	TACGTTTGGATTTCCGTGTT	Edn1_C3R_31	AGCGTTGCCCAATCCGTA
	Edn1_4F98[b]	CAGCAGACACGGACAGGATT	Edn1_5R176[a]	CCCTTCCTGCCTCTGTATGA
	Edn1_3F104	ACGACAACACCTGCCTGAAC	Edn1_5R28	GCCGTTTATCACTGTTGTTCTT
	Edn1_2F36[a]	GCCTCCACCGTCAAAGAG	Edn1_C5R_87[b]	TGCTGTCTCACCCTCCACTT
Edn3b	Edn3b_F66[a]	GGACCACATGAACTGCACAC	Edn3b_R83	GGGGGTGTTGATCCAGATG
	Edn3b_F142	TCAGGCGTTTGCTTTATTTTG	Edn3b_R545[a,b]	GTGTCTCTGCCCCCTCGT
	Edn3b_F40I[b]	GCGTCTACTACTGCCACTT		
GAPDH	GAPDH_Cich1_F[a]	GCATCATCCGTGAGGTTACTCC	GAPDH_Cich1_R[a]	TGTCAGCTTACCATTGAGTTCAGG
Actin	Actin_Ab_Fw[b]	GGCCCAGAGCAAGAGAGCTAT	Actin_Ab_Rev[b]	GATGCCAGATCTCTTCTCCATGTCATCC

[a-b] Primers used in the expression pattern analyses with PCR[a] and Real-Time PCR[b]

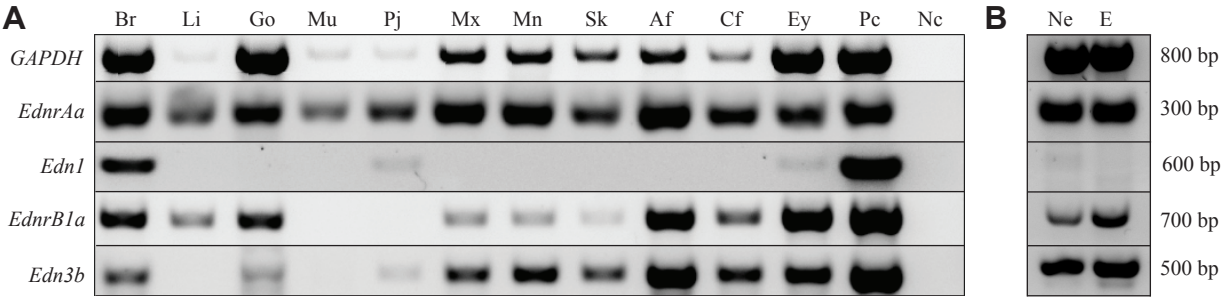
Supplementary Table 3

Overview of nonsynonymous substitutions

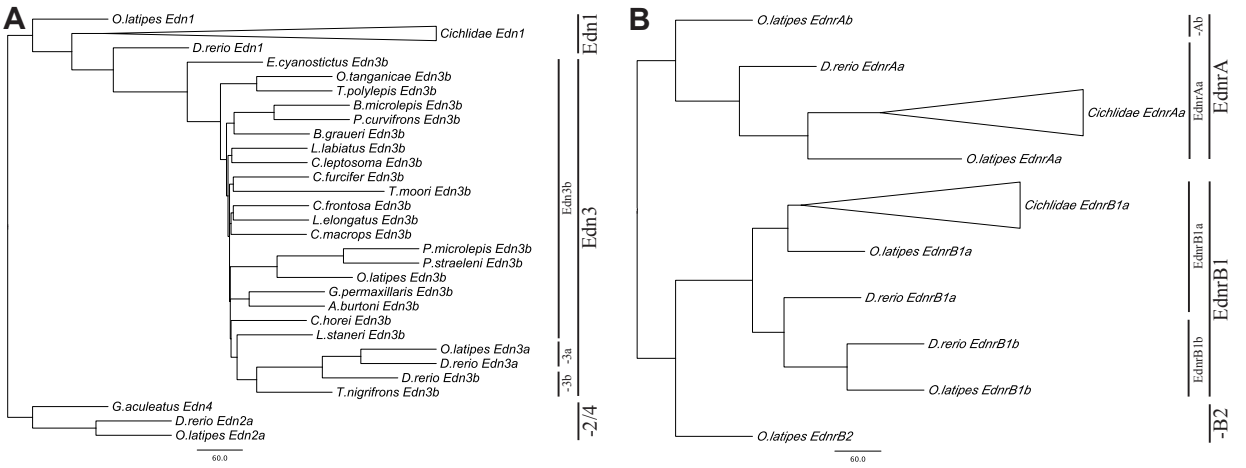
Locus	Haplo		Ecto		Peri		Cypri		Cyph		Limn		Lamp		Eret		Bathy		Boule		Trem		Tylo		Tilap		Total charge	Total polarity
EdnrAa	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	1	2	1
EdnrB1a	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-
Edn1	-	5	-	1	1[a]	1	1[a]	-	-	-	1	1	1	4	-	2	-	1	-	1	-	5	3[b]	4	3[b]	1	14	7
Edn3b	1	2	-	-	2	-	-	-	-	-	-	1	-	1	-	1	-	1	-	1	-	2	1[b]	3	1[b]	2	4	8

Lineage specific (left column) and species-specific (right column) nonsynonymous substitutions per locus for the Haplochromines, Ectodines, Perissodines, Cyprichromines, Limnochromines, Lamprologines, Eretmodines, Bathybatines, Boulengerochromines, Trematocarines, Tylochromines and Tilapiini's. The number of nonsynonymous substitutions causing a change in charge and polarity are shown in the last two columns

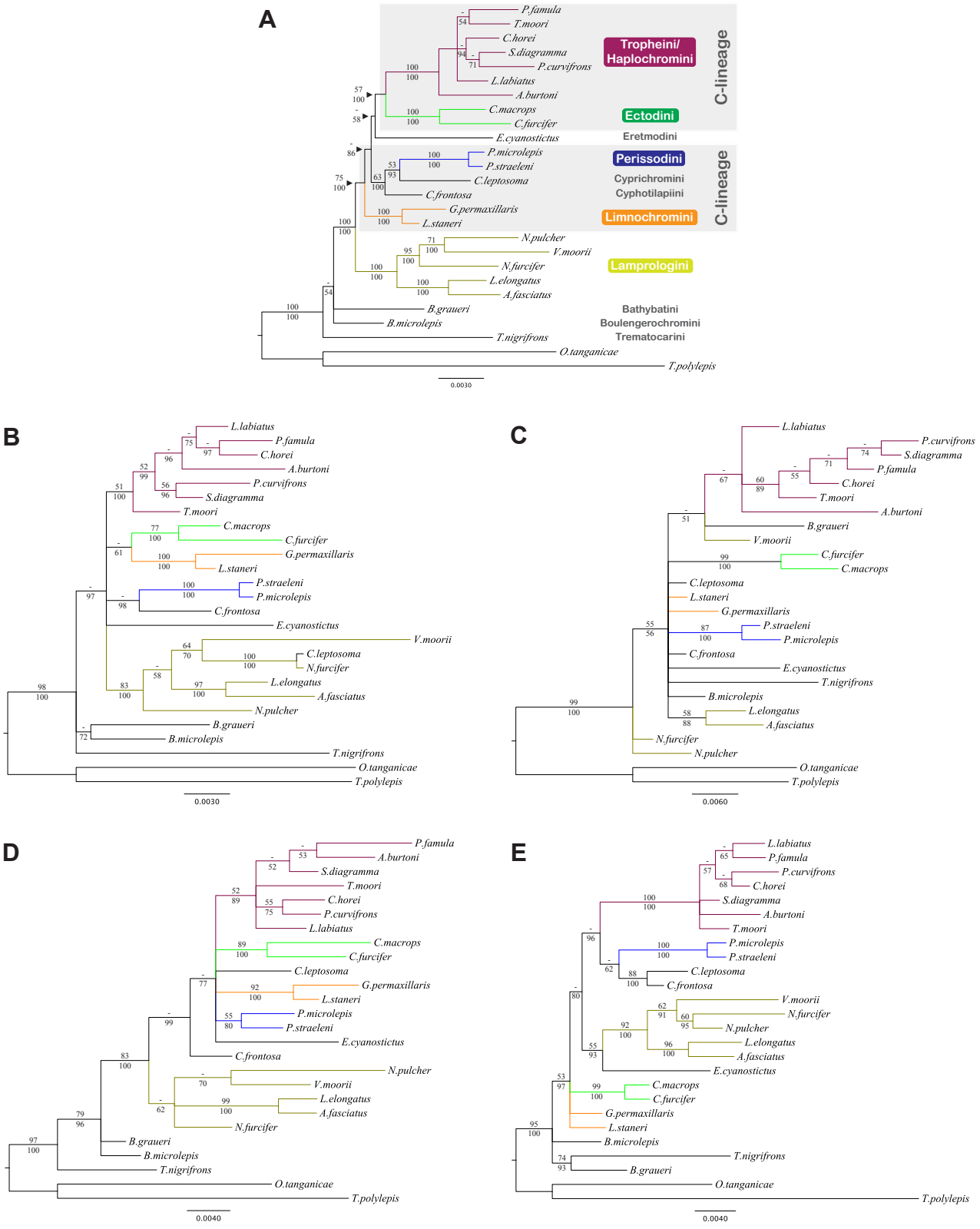
[a] One nonsynonymous substitution is shared by the Cyprichromini and one of the Perissodini
[b] Nonsynonymous substitutions are shared by the two outgroup clades Tylochromini and Tilapiini



Supplementary Figure S4. Expression patterns of studied loci. For each PCR 20 ng of template cDNA was used and 5 µl PCR product was loaded on a 1,5% agarose gel. **a** The PCR products of the amplificated GAPDH (~800 bp), EdnrAa (~300 bp), Edn1 (~600 bp), EdnrB1a (~700 bp) and Edn3b (~500 bp) loci for the following tissues; brain (Br), liver (Li), gonads (Go), muscle (Mu), pharyngeal jaw (Pj), maxilla (Mx), mandible (Mn), skin (Sk), anal fin (Af), caudal fin (Cf), eye (Ey), whole juveniles as positive control (Pc) and H2O as negative control (Nc). **b** The PCR products for the non-egg spot tissue (Ne) and egg spot tissue (E).



Supplementary Figure S5. Consensus phylogenies of the ML analyses of the endothelin family members with their teleost homologs. **a** Phylogenetic relationships of cichlid endothelin proteins (223 AA) and their teleost orthologs using the JTT+I+G model. The cichlid Edn1 sequences (collapsed and depicted as triangle) cluster together with other teleost Edn1. Similarly, the cichlid Edn3b sequences cluster together with the teleost Edn3b sequences. **b** Phylogenetic relationships of cichlid endothelin receptor proteins (463 AA) and their teleost orthologs using the Blosum62+G model. Cichlid EdnrAa sequences cluster together with the teleost EdnrAa sequences, while cichlid EdnrB1a cluster together with teleost EdnrB1a sequences.



Supplementary Figure S6. Bayesian phylogenetic hypotheses of the wds data set of all studied loci. Five major cichlid lineages are retrieved in most analyses and indicated by the following color scheme: asparagus green=Lamprologini, blue=Perissodini, orange=Limnochromini, green=Ectodini, maroon=Tropheini and Haplochromini. The trees were rooted with T. polylepis (Tylochromini) and O.tanganicae (Tilapiini). Branch lengths for T. polylepis were shortened (50%) in all phylogenies. ML bootstrap values and BAY posterior probabilities

above 50% are indicated respectively above and under the branch. **a** Phylogeny based on the concatenated data set (i.e., the four Endothelin family members plus one additional non-Endothelin, autosomal coding gene; total length: 9451 bp) using the GTR+I+G model. Tribes belonging to the C-lineage (Cyprichromini, Cyphotilapiini, Perissodini, Limnochromini, Ectodini, Tropheini and Haplochromini) are highlighted in grey. Phylogenetic relationships between clades are largely in agreement with available species trees (Salzburger et al. 2002; 2005; 2007). **b** Phylogeny based on Edn1 (2449 bp) using the HKY+G model showing a seven-tribe polytomy for the derived clades. C. leptosoma is clustering within the lamprologines. **c** Phylogeny based on Edn3b (512 bp) using the HKY+G model. The tree is mostly unresolved, with two Lamprologini species at the base of a major polytomy. Several tribes are identified (i.e., Perissodini, Ectodini and Tropheini and Haplochromini). **d** Phylogeny based on EdnrAa (963 bp) using the HKY+G model showing a six-tribe polytomy for the derived clades, while the five major tribes are recovered. **e** Phylogeny based on EdnrB1a (2091 bp) using the GTR+G model. A basal position was found for the ectodines, the lamprologines and perissodines were found at a more derived position. Note that the Tropheini are the Tanganyikan representatives of the Haplochromini (Salzburger et al. 2005).

CHAPTER 3

SEQUENCE ANALYSES OF THE *DISTAL-LESS HOMEOBOX* GENE FAMILY IN EAST AFRICAN CICHLID FISHES REVEAL SIGNATURES OF POSITIVE SELECTION

ARTICLE

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Sequence analyses of the *distal-less homeobox* gene family in East African cichlid fishes reveal signatures of positive selection

Eveline T Diepeveen*, Fabienne D Kim and Walter Salzburger

Abstract

Background: Gen(om)e duplication events are hypothesized as key mechanisms underlying the origin of phenotypic diversity and evolutionary innovation. The diverse and species-rich lineage of teleost fishes is a renowned example of this scenario, because of the fish-specific genome duplication. Gene families, generated by this and other gene duplication events, have been previously found to play a role in the evolution and development of innovations in cichlid fishes - a prime model system to study the genetic basis of rapid speciation, adaptation and evolutionary innovation. The *distal-less homeobox* genes are particularly interesting candidate genes for evolutionary novelties, such as the pharyngeal jaw apparatus and the anal fin egg-spots. Here we study the *dlx* repertoire in 23 East African cichlid fishes to determine the rate of evolution and the signatures of selection pressure.

Results: Four intact *dlx* clusters were retrieved from cichlid draft genomes. Phylogenetic analyses of these eight *dlx* loci in ten teleost species, followed by an in-depth analysis of 23 East African cichlid species, show that there is disparity in the rates of evolution of the *dlx* paralogs. *Dlx3a* and *dlx4b* are the fastest evolving *dlx* genes, while *dlx1a* and *dlx6a* evolved more slowly. Subsequent analyses of the nonsynonymous-synonymous substitution rate ratios indicate that *dlx3b*, *dlx4a* and *dlx5a* evolved under purifying selection, while signs of positive selection were found for *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b*.

Conclusions: Our results indicate that the *dlx* repertoire of teleost fishes and cichlid fishes in particular, is shaped by differential selection pressures and rates of evolution after gene duplication. Although the divergence of the *dlx* paralogs are putative signs of new or altered functions, comparisons with available expression patterns indicate that the three *dlx* loci under strong purifying selection, *dlx3b*, *dlx4a* and *dlx5a*, are transcribed at high levels in the cichlids' pharyngeal jaw and anal fin. The *dlx* paralogs emerge as excellent candidate genes for the development of evolutionary innovations in cichlids, although further functional analyses are necessary to elucidate their respective contribution.

Keywords: *Distal-less homeobox* gene, Molecular evolution, Cichlid fishes, Teleost fishes, Positive selection, Differential selection, Gene duplication, dN/dS

Background

Teleost fishes (Teleostei) are among the most diverse lineages on Earth and with nearly 30,000 species the most species-rich vertebrate group. This is in stark contrast to the more basal non-teleost ray-finned fishes that are characterized by small numbers of species. A causal explanation for this discrepancy in speciation rates between the derived Teleostei and the non-teleost ray-finned fishes might be the fish-specific genome duplication (FSGD) that occurred in the ancestor of modern teleosts ([1-4] and references therein). It has been hypothesized that the FSGD has laid down the genetic conditions necessary for the evolution of phenotypic diversity [5], although the exact causes of diversification of such a large clade are likely to be more complex and most probably also include other factors [6].

The Hox gene clusters, which evolved through both tandem and whole genome duplications, represent illustrative examples for the contribution of duplicated genes to morphological evolution across the animal kingdom (see e.g., [7-9]). Together with other homeotic genes,

* Correspondence: evelinetdiepeveen@gmail.com
Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

Hox genes play a crucial role in the development of the multicellular body plan (e.g., anterior-posterior patterning; [10]). Furthermore, Hox genes are known to be involved in the development of evolutionary novelties, such as walking limbs and the wings of insects [11-15]. It has been shown that different mechanisms such as *cis*-regulatory evolution, changes in protein function and post-transcriptional regulation of the Hox genes contribute to morphological diversification (reviewed in e.g., [8,15,16]).

East African cichlid fishes show a remarkable level of phenotypic diversity between closely related species and constitute the most diverse adaptive radiations known [17-21]. Although several smaller radiations of cichlid fishes exist outside of Africa (e.g., in Central and South America), an astonishingly high number of cichlid species (close to 1900 species [22]) evolved in and around lakes Malawi, Victoria and Tanganyika in the last few million to several thousand years [23,24]. The various cichlid species differ in body shape, coloration, reproductive biology and mouth morphology [25-27] - traits which are thought to, at least partly, underlie the evolutionary success of cichlid fishes [18,27,28]. Furthermore, several morphological innovations are unique to cichlids or specific lineages thereof. The highly modified and morphological diverse pharyngeal jaw apparatus, for example, correlates with the diversity in foraging strategies exploited by the different cichlid species [27-29]. The occurrence of several color morphs within species, sexual color dimorphism and anal fin egg-spots are three characteristic features of the extremely species-rich and mouthbrooding haplochromine lineage [30].

As a result of their great phenotypic diversity and high number of species, cichlid fishes provide an ideal set up to examine the genetic basis of rapid speciation, evolutionary innovations and adaptation [21,31-37]. An important strategy is the study of so-called candidate genes, i.e., genes with known functions in development in other organisms such as zebrafish. For example, it has been shown that *csf1ra*, which was identified as xanthophore marker in zebrafish [38,39] is involved in the morphogenesis of the egg-spots of haplochromine cichlids [31]. Furthermore, species-specific jaw shapes of different cichlid species correlate with differences in early *bmp4* expression patterns, a gene which has also the potential to change the mandibular morphology in zebrafish [40]. Many of these candidate genes belong to larger gene families such as the endothelin family of ligands and receptors that are putatively involved in the morphogenesis of the pharyngeal jaw apparatus and pigmentation [36], and the above mentioned Hox gene clusters [41].

Recently, Renz *et al.* [35] characterized seven *distal-less homeobox* (*dlx*) genes and examined their expression patterns in the developing pharyngeal arches and/or pharyngeal teeth of the haplochromine cichlid *Astatotilapia burtoni*. The vertebrate *dlx* genes are widely known for their crucial roles in the development (of components) of the nervous system, craniofacial skeleton and connective tissue and in the formation of appendages [reviewed in 42]. These functions seem to be conserved across a wide range of animal taxa. For example, the vertebrate *dlx* genes are homologs of, and share several functions with, the single *Distal-less* (*dll*) gene of *Drosophila* [42]. Within vertebrates, the expression patterns of dlx homologs is similar in early development [35,42-45]. At the same time, *dlx* genes have been implicated with evolutionary novelties such as the eyespots in various butterfly species [46-48], the insect antenna [49,50] and the vertebrate craniofacial bones [51].

Phylogenetic analyses and the chromosomal arrangements of the vertebrate *dlx* genes suggest that the extant *dlx* repertoire has evolved by an initial tandem duplication, followed by two rounds of whole genome duplication in the lineage towards vertebrates and a third one in the lineage towards teleost fishes, the FSGD [35,44,45,52]. These duplication events resulted in multiple so-called *dlx* clusters, in which two *dlx* genes are located in a tail-to-tail arrangement on the respective chromosome. Linked *dlx* genes are transcribed coincidently due to shared *cis*-regulating elements in the intergenic regions [35,42,43]. Four of these *dlx* clusters have been identified in teleost fish; *dlx1a-dlx2a*, *dlx3a-dlx4a*, *dlx3b-dlx4b* and *dlx5a-dlx6a* [43,44]. Seven of these *dlx* genes have been identified in the cichlid *A. burtoni*, where they are expressed in tissues that make up putative evolutionary innovations [35].

Here, we analyzed the *dlx* repertoire and diversity in detail in a phylogenetically representative set of 23 East African cichlid species in order to study the molecular evolution of this prominent developmental gene family. To this end, we first performed phylogenetic comparisons of the dlx proteins including the sixty amino acids long homeobox domain in a range of teleost fishes in combination with blast searches of these sequences against the draft genomes of four cichlid species. Teleost and cichlid-specific phylogenies were examined to compare the rate of evolution of both between and within *dlx* gene trees. Several studies have shown that loci putatively involved in evolutionary innovations are characterized by adaptive protein evolution in cichlids [31,36,53]. Therefore, all loci were screened for elevated rates of protein evolution by means of d_N/d_S analyses. Our analyses indicate the presence of *dlx3a* in cichlids and that the *dlx* repertoire of cichlid fishes is shaped by differential selection pressures and rates of evolution, with signs of positive selection on specific sites in *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b*.

Methods

Dlx protein sequence comparison in teleost fishes

The sequences of nine dlx proteins (i.e., dlx1a, dlx2a, dlx2b, dlx3a, dlx3b, dlx4a, dlx4b, dlx5a and dlx6a) of seven teleost

species (i.e., zebrafish (*Danio rerio*), Atlantic cod (*Gradus morhua*), three-spined stickleback (*Gasterosteus aculeatus*), spotted green pufferfish (*Tetraodon nigroviridis*), Japanese pufferfish (*Takifugu rubripes*), Japanese medaka (*Oryzias latipes*) and Nile Tilapia (*Oreochromis niloticus*)) were obtained from Ensemble (release 68, July 2012; see Additional file 1 for accession numbers). *Dlx2b* was excluded from all further analyses, due to its lineage-specific loss in percomorphs, to which all studied species belong except *D. rerio* and *G. morhua* (see [35]). Sequences were aligned with Tcoffee [54,55], ambiguous sites were removed and tblastx searches were performed to determine dlx protein sequences in the draft cichlid genomes of *Astatotilapia burtoni*, *Neolamprologus brichardi* and *Pundamilia nyererei* (BROAD Institute, unpublished data; see Additional file 1 for scaffold numbers). To determine the rate of evolution for each of the dlx proteins, phylogenetic analyses were performed in PAUP* 4.0 [56] under parsimony settings and the number of amino acid changes was obtained. *D. rerio* or *G. morhua* was used as outgroup species and bootstrap analyses with 100 replicates were conducted to test the robustness of the obtained topologies. Next, the sixty amino acids long homeobox domain was extracted from the sequences and aligned to the homeobox domain of the single *Distal-less* (*Dll*) gene of *Drosophila melanogaster* [Ensemble: FBgn0000157] in Geneious 5.6 [57] for closer inspection of the conservation of the domain and to identify gene-specific substitutions.

Cichlid samples and genomic DNA sequencing

White muscle and/or fin clip samples were collected during fieldwork in Zambia in 2007 and 2008 using a standard operating procedure described in [29]. In total 23 Lake Tanganyikan cichlid species were included in this study (Additional file 2). Genomic DNA was extracted following a standard Proteinase K protocol [58]. Cichlid-specific PCR primers were designed based on available and/or draft genomic and transcriptomic cichlid sequences, which were identified by tblastx searches of publicly available *dlx* sequences from other teleost species (see Additional file 1 for species and accession numbers). This was done for eight *dlx* loci: *dlx1a*, *dlx2a*, *dlx3a*, *dlx3b*, *dlx4a*, *dlx4b*, *dlx5a* and *dlx6a* (see Additional file 3 for primer sequences). Standard PCR reactions, purification steps and sequencing reactions were set up and performed as described elsewhere [36]. PCR products of the partially sequenced loci were visualized with GelRed (Biotium) on a 1.5% agarose gel and sequenced on a 3130*xl* capillary sequencer (Applied Biosystems). Partial sequences were aligned and visually inspected using Codon Code Aligner 3.7.1 (CodonCode Corporation, Dedham, MA). Exon/intron boundaries were determined by homology comparisons with the sequences from the other teleost species. All generated cichlid *dlx* sequences have been deposited into GenBank [GenBank: KC285366-KC285546] (Additional file 2).

Phylogenetic analyses of cichlid samples

Individual gene trees were constructed using maximum likelihood in PAUP* 4.0 [56] and Bayesian Inference in MrBayes 3.2 [59,60]. The best-fitting model of nucleotide substitution was determined with the corrected Akaike information criteria and likelihood ratio tests conducted in jModeltest 0.1.1 [61,62]. Bootstrap analyses with 100 replicates were performed in PAUP* and MrBayes was run for 10.500.000 generations. *Oreochromis tanganicae* was used as outgroup (see e.g., [63]). Phylogenetic analysis of a concatenated dataset of 9.2 kb was performed as described above in PAUP* to generate a common input tree file (treeBASE submission 14433) for the subsequent analyses.

Coding sequence data of the 23 cichlid species (treeBASE submission 14433) was assessed with both site- and branch-site models as implemented in the program Codeml of the software package PAML (Phylogenetic Analysis by Maximum Likelihood) 4.3 [64,65]. The following parameters were estimated for all eight *dlx* datasets under different models: the nonsynonymous/synonymous substitution rate ratio, ω, the proportion of sites assigned to an ω category, $p_{0,1,2}$, and the p and q parameters of the β distribution. Tests of positively selected sites were conducted by performing Likelihood Ratio Tests (LRT) of the following model comparisons: M1a (Nearly Neutral) with M2a (Positive Selection), M7 (β) with M8 (β & $\omega_s \geq 1$), and M8a (β & $\omega_s = 1$) with M8. The comparison between M0 (one-ratio) and M3 (discrete) was used as a test of variable ω among sites. The naïve empirical Bayes (NEB; [66,67]) and the Bayes empirical Bayes (BEB; [68]) criteria were used to calculate the posterior probabilities for site classes and the BEB was used to identify sites under positive selection when the LRT was significant. To test whether the *dlx* genes evolved under non-neutral evolution in specific lineages a LRT between the null model ($\omega_s = 1$) and the alternative model ($\omega_s \geq 1$) was performed in the branch-site analyses. Branches of interest, or so-called foreground branches, were chosen based on the results of the phylogenetic analyses and branch tests performed in Hyphy ([69], following [36]).

Additional tests of positive selection on the partial *dlx* sequences were performed with the Sitewise Likelihood Ratio estimation of selection program (SLR; [70]) v1.3. The common input tree file was used (see above) and the significance level was set to 95%.

Amino acid substitutions were screened for possible effect on protein function with the program SIFT (Sorting Intolerant from Tolerant; [71].

Results

Dlx protein sequence comparison in teleost fishes

The tblastx searches of the teleost dlx proteins resulted in the retrieval of eight *dlx* genes in all four cichlid species. Furthermore, the genomic locations of these *dlx* loci (Additional file 1) indicate that four *dlx* clusters are present in the cichlid lineage: *dlx1a-dlx2a*; *dlx3a-dlx4a*; *dlx3b-dlx4b* and *dlx5a-dlx6a*. All other teleost species examined contain this full set of genes, except zebrafish, in which *dlx3a* could not be located, and medaka, in which *dlx4b* is missing, as previously noted [35,44,45]. Interestingly, in contrast to Renz *et al.* [35] we do find evidence for the existence of *dlx3a* in cichlids, including *A. burtoni* (Figure 1, Additional file 4).

The sixty amino acid long homeobox domain of the eight teleost dlx proteins are highly conserved among teleost fish and even between teleosts and the single Dll protein of *D. melanogaster* (Additional file 4). Despite the high level of conservation, several locus-specific amino acid substitutions are present in the paralogs, making it possible to distinguish between individual dlx homeobox domains.

Phylogenetic analyses of the dlx protein sequences were performed to examine the rate of evolution of the dlx paralogs in teleost fishes. The overall and relative longest trees were found for *dlx4b* and *dlx3a*, while for *dlx1a* and *dlx6a* the shortest tree lengths were observed (Figure 1 and Table 1). Typically the longest branches were observed in the two basal species *D. rerio* and *G. morhua*. Interestingly, relatively long branch lengths for the branch towards the four cichlid species were observed for *dlx3a* and *dlx6a*, indicting elevated rates of molecular evolution. The opposite scenario was observed in the overall more conserved dlx1a and dlx5a proteins. To study these effects in more detail cichlid specific gene trees were constructed.

The rate of *dlx* gene evolution in East African cichlid fishes

To reconstruct the molecular evolutionary history of the *dlx* homologs in East African cichlid species, we determined the rate of evolution and the signatures of selection

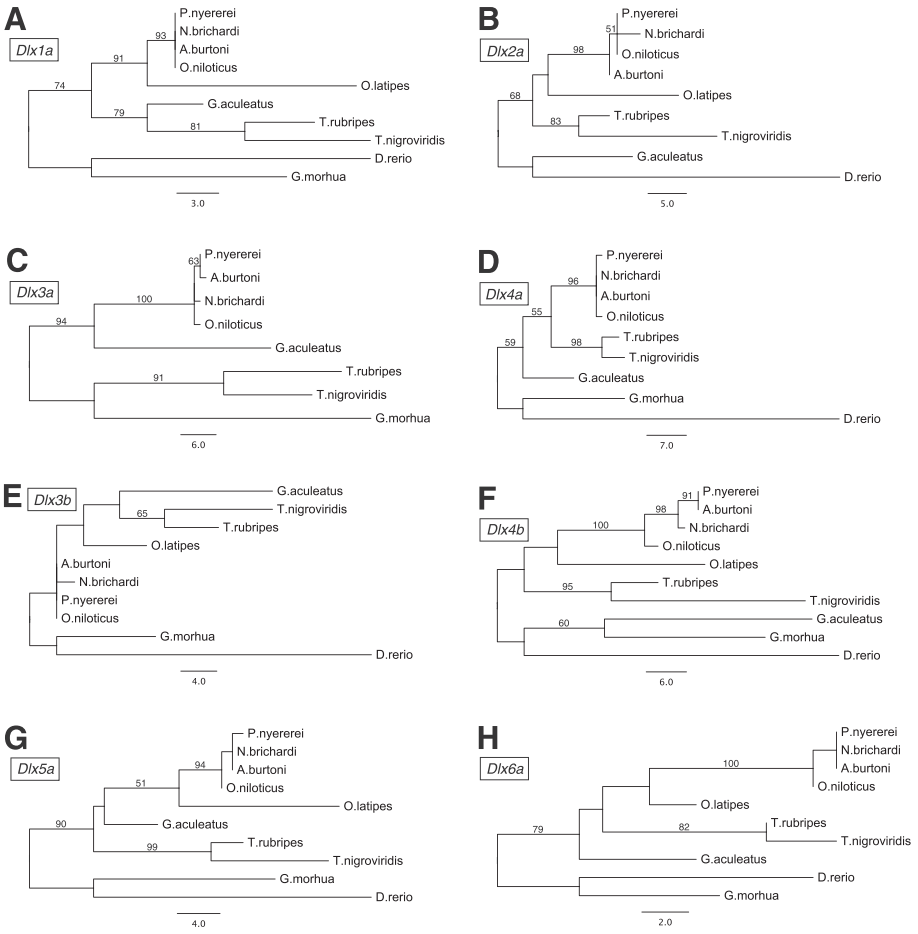


Figure 1 Maximum Likelihood phylogenetic hypotheses for the eight dlx paralogs in teleost fishes. (A) Dlx1a (254 amino acids (aa)). **(B)** Dlx2a (276 aa). **(C)** Dlx3a (307 aa). **(D)** Dlx4a (259 aa). **(E)** Dlx3b (283 aa). **(F)** Dlx4b (257 aa). **(G)** Dlx5a (285 aa). **(H)** Dlx6a (247 aa). Bootstrap probabilities (PAUP*) above 50% are shown.

Table 1 Overall and relative tree lengths of teleost protein phylogenies

dlx paralog	Length (# amino acids)	Tree length (absolute)	Tree length (relative)
dlx1a	254	93	0.366
dlx2a	276	122	0.442
dlx3a	307	179	0.583
dlx3b	283	108	0.382
dlx4a	259	123	0.475
dlx4b	257	222	0.864
dlx5a	285	125	0.439
dlx6a	247	56	0.227

pressure in a phylogenetically representative set of 23 species. The gene trees of the obtained partial cichlid *dlx* sequences resulted in various polytomies (Additional file 5), probably due to the limited size of some of the datasets (minimum of 0.7 kb). Although for each gene tree specific branches were observed with relative long branches, there is not a particular species or clade that has evolved under faster rates of evolution in all of the *dlx* loci examined. Interestingly, three branches have relative long branch lengths in multiple topologies: the branch towards the Lamprologini (*dlx2a*, *dlx4a* and *dlx5a*), *C. leptosoma* (*dlx3b*, *dlx4a* and *dlx5a*) and *C. furcifer* (*dlx1a* and *dlx6a*). The relative tree lengths (Additional file 5 and Table 2) of these gene trees reveal similar results as the teleost protein trees, with *dlx4b* and *dlx3a* evolving fastest and *dlx1a* and *dlx6a* evolving more slowly.

Observed signatures of selection pressure in cichlid *dlx* loci

To investigate signatures of selection pressure in the *dlx* loci, we performed detailed analyses of the d_N/d_S ratios. Maximum likelihood parameter estimations for ω, $p_{0,1,2}$ and p and q under different evolutionary models can be found in Table 3 for all eight *dlx* loci. Estimations of ω under the M0 model suggest that the *dlx* genes evolved under purifying selection with ω ranging from 0.0001 (*dlx5a*) to 0.457 (*dlx2a*). A small proportion of sites, 0.00001-24.2%, was estimated to have evolved neutrally ($\omega = 1$) under the M1a model. By using models that allow ω to vary among sites, 0.7-12.3% of sites was detected with $\omega > 1$ in *dlx1a*, *dlx2a*, *dlx3a*, *dlx4b* and *dlx6a*. Overall, most sites are estimated to have evolved under purifying selection, with highest proportions found in *dlx3b*, *dlx4a* and *dlx5a*.

Table 2 Overall and relative tree lengths of cichlid *dlx* gene trees

dlx paralog	Length (# base pairs)	Tree length (absolute)	Tree length (relative) 10^{-4}
dlx1a	737	0.036	0.483
dlx2a	1371	0.094	0.684
dlx3a	666	0.061	0.910
dlx3b	1972	0.120	0.609
dlx4a	1166	0.104	0.892
dlx4b	722	0.068	0.937
dlx5a	1538	0.093	0.607
dlx6a	1710	0.093	0.542

Likelihood ratio tests of the subsequent model comparisons (Table 4) resulted in the rejection of the null models in only the following comparisons per loci: *dlx1a* (M8a-M8), *dlx2a* (all four comparisons), *dlx3a* (M0-M3; M8a-M8) and *dlx4b* (all four comparisons). Positively selected sites were detected with the BEB in *dlx2a* (5 sites), *dlx3a* (1 site) and *dlx4b* (3 sites; see Table 4, Figure 2). The less constraining analyses with the NEB resulted in two more putative positively selected sites in *dlx1a* (1) and *dlx2a* (1; Figure 2). Fewer positively selected sites were identified by the SLR analyses for *dlx2a* (position: 36; significance: 99%), *dlx3a* (37, 157; 99%, 95%) and *dlx4b* (145; 99%).

None of the performed LTRs of the branch-site analyses were significant ($1 \geq p \geq 0.20$) indicating that although the ω ratios do vary among sites (see above), the ω ratios do not vary significantly among lineages.

Amino acid substitutions and their predicted effect on function

Next, the individual amino acid substitutions were examined in more detail. The total protein length and the number of amino acid substitutions per locus are shown in Table 5 (see also Figure 2 and Additional file 6). A relative large number of substitutions was observed in *dlx2a* (13), *dlx3a* (16) and *dlx4b* (10), while in *dlx5a* no substitution was found. Most of the amino acid substitutions are species-specific (i.e., observed in a single species), although lineage-specific substitutions were observed for the lamprologines (*dlx2a*, *dlx3a*, *dlx4b*), ectodines (*dlx2a*) and haplochromines (*dlx2a*, *dlx3a*). None of the observed amino acid substitutions have a predicted effect on the protein functions (see Table 5), although two substitutions were observed in the homeobox domain of *dlx2a* (Figure 2).

Selection regimes on the *dlx* clusters

It is known that the paired members of each of the four *dlx* clusters (Additional file 4) are transcribed concurrently [35,42-45]. To characterize if the members of the same *dlx* cluster evolved at similar rates and under similar selection regimes, we had a closer inspection of these paired genes. First, the teleost dlx protein and cichlid gene trees show that overall and relative tree lengths (or the rate of evolution) differ between the two genes within a cluster. Loci with the highest (*dlx3a*: 0.583/0.910 and *dlx4b*: 0.864/0.937) or the smallest (*dlx1a*: 0.366/0.483 and *dlx6a*: 0.227/0.542) tree lengths are never observed

Table 3 Site model parameter estimates generated by the CodeML analyses for the eight *dlx* paralogs

Locus	Parameter estimates under different models						
	M0 (one ratio)	M1a (neutral)	M2a (selection)	M3 (discrete)	M7 (ß)	M8 (ß & ω)	M8a (ß & $\omega_s = 1$)
dlx1a	$\omega = 0.111$	$p_0 = 0.940$, $\omega_0 = 0$	$p_0 = 0.980$, $\omega_0 = 0$	$p_0 = 0.640$, $\omega_0 = 0$	$p = 0.005$, $q = 0.049$	$p = 0.005$, $q = 22.651$	$p = 1.333$, $q = 24.671$
		$p_1 = 0.060$, $\omega_1 = 1$	$p_1 = 0$, $\omega_1 = 1$	$p_1 = 0.340$, $\omega_1 = 0$		$p_0 = 0.980$	$p_0 = 0.936$
			$p_2 = 0.020$, $\omega_2 = 12.580$	**$p_2 = 0.020$, $\omega_2 = 12.580$**		**$p_1 = 0.020$, $\omega = 12.580$**	$p_1 = 0.065$, $\omega = 1$
dlx2a	$\omega = 0.457$	$p_0 = 0.833$, $\omega_0 = 0$	$p_0 = 0.879$, $\omega_0 = 0$	$p_0 = 0$, $\omega_0 = 0$	$p = 0.005$, $q = 0.011$	$p = 0.005$, $q = 0.046$	$p = 0.005$, $q = 12.618$
		$p_1 = 0.167$, $\omega_1 = 1$	$p_1 = 0.110$, $\omega_1 = 1$	$p_1 = 0.986$, $\omega_1 = 0.081$		$p_0 = 0.989$	$p_0 = 0.833$
			$p_2 = 0.010$, $\omega_2 = 18.903$	**$p_2 = 0.014$, $\omega_2 = 16.629$**		**$p_1 = 0.011$, $\omega = 18.053$**	$p_1 = 0.167$, $\omega = 1$
dlx3a	$\omega = 0.320$	$p_0 = 0.758$, $\omega_0 = 0$	$p_0 = 0.921$, $\omega_0 = 0$	$p_0 = 0.877$, $\omega_0 = 0$	$p = 0.005$, $q = 0.017$	$p = 0.015$, $q = 0.088$	$p = 0.005$, $q = 13.826$
		$p_1 = 0.242$, $\omega_1 = 1$	$p_1 = 0.034$, $\omega_1 = 1$	**$p_1 = 0.120$, $\omega_1 = 2.099$**		$p_0 = 0.032$	$p_0 = 0.758$
			$p_2 = 0.045$, $\omega_2 = 4.451$	**$p_2 = 0.003$, $\omega_2 = 11.657$**		**$p_1 = 0.032$, $\omega = 5.096$**	$p_1 = 0.242$, $\omega = 1$
dlx3b	$\omega = 0.047$	$p_0 = 1$, $\omega_0 = 0.047$	$p_0 = 1$, $\omega_0 = 0.047$	$p_0 = 0.243$, $\omega_0 = 0.047$	$p = 4.896$, $q = 99.00$	$p = 4.895$, $q = 99.00$	$p = 4.895$, $q = 99.00$
		$p_1 = 1E\text{-}5$, $\omega_1 = 1$	$p_1 = 0$, $\omega_1 = 1$	$p_1 = 0.262$, $\omega_1 = 0.047$		$p_0 = 1$	$p_0 = 1$
			$p_2 = 0$, $\omega_2 = 1$	$p_2 = 0.495$, $\omega_2 = 0.047$		$p_1 = 1E\text{-}5$, $\omega = 1$	$p_1 = 1E\text{-}5$, $\omega = 1$
dlx4a	$\omega = 0.050$	$p_0 = 1$, $\omega_0 = 0.050$	$p_0 = 1$, $\omega_0 = 0.050$	$p_0 = 0.216$, $\omega_0 = 0.050$	$p = 5.267$, $q = 99.00$	$p = 5.266$, $q = 99.00$	$p = 0.137$, $q = 2.305$
		$p_1 = 1E\text{-}5$, $\omega_1 = 1$	$p_1 = 0$, $\omega_1 = 1$	$p_1 = 0.384$, $\omega_1 = 0.050$		$p_0 = 1$	$p_0 = 1$
			$p_2 = 0$, $\omega_2 = 1$	$p_2 = 0.400$, $\omega_2 = 0.050$		$p_1 = 1E\text{-}5$, $\omega = 1$	$p_1 = 1E\text{-}5$, $\omega = 1$
dlx4b	$\omega = 0.259$	$p_0 = 0.872$, $\omega_0 = 0$	$p_0 = 0.993$, $\omega_0 = 0.090$	$p_0 = 0$, $\omega_0 = 0$	$p = 0.005$, $q = 0.042$	$p = 9.871$, $q = 99.00$	$p = 0.005$, $q = 30.817$
		$p_1 = 0.128$, $\omega_1 = 1$	$p_1 = 0$, $\omega_1 = 1$	$p_1 = 0.993$, $\omega_1 = 0.090$		$p_0 = 0.993$	$p_0 = 0.872$
			$p_2 = 0.007$, $\omega_2 = 12.858$	**$p_2 = 0.007$, $\omega_2 = 12.858$**		**$p_1 = 0.007$, $\omega = 12.859$**	$p_1 = 0.128$, $\omega = 1$
dlx5a	$\omega = 1E\text{-}4$	$p_0 = 1$, $\omega_0 = 0$	$p_0 = 1$, $\omega_0 = 0$	$p_0 = 0.424$, $\omega_0 = 0$	$p = 0.005$, $q = 1.294$	$p = 0.005$, $q = 99.00$	$p = 0.005$, $q = 44.274$
		$p_1 = 1E\text{-}5$, $\omega_1 = 1$	$p_1 = 0$, $\omega_1 = 1$	$p_1 = 0.420$, $\omega_1 = 0$		$p_0 = 1$	$p_0 = 1$
			$p_2 = 0$, $\omega_2 = 1$	$p_2 = 0.156$, $\omega_2 = 0$		$p_1 = 1E\text{-}5$, $\omega = 1$	$p_1 = 1E\text{-}5$, $\omega = 1$
dlx6a	$\omega = 0.029$	$p_0 = 0.967$, $\omega_0 = 0$	$p_0 = 0.992$, $\omega_0 = 0$	$p_0 = 0.899$, $\omega_0 = 0$	$p = 0.066$, $q = 1.696$	$p = 0.005$, $q = 2.390$	$p = 0.005$, $q = 1.672$
		$p_1 = 0.033$, $\omega_1 = 1$	$p_1 = 0$, $\omega_1 = 1$	$p_1 = 0.093$, $\omega_1 = 0$		$p_0 = 0.992$	$p_0 = 0.967$
			$p_2 = 0.008$, $\omega_2 = 6.336$	**$p_2 = 0.008$, $\omega_2 = 6.336$**		**$p_1 = 0.008$, $\omega = 6.336$**	$p_1 = 0.033$, $\omega = 1$

Note: $p_{0\text{-}2}$ are the proportions of sites assigned to an ω category or to a beta distribution with p and q as parameters. ω ratios greater than one and their corresponding proportions are depicted in bold.

Table 4 Likelihood ratio test (LTR) statistics of site model comparisons for *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b*

Locus	Test	LRT (2Δ*l*)	p	Selected sites (BEB)
dlx1a	M0 vs M3	8.416	0.077	-
	M1a vs M2a	3.396	ns	-
	M7 vs M8	3.680	ns	-
	M8a vs M8	5.084	0.012	-
dlx2a	M0 vs M3	70.438	<0.001	-
	M1a vs M2a	39.198	<0.001	**36A**, **48 T**, *255I*
	M7 vs M8	43.093	<0.001	**36A**, **48 T**, *195A*, *254A*, *255I*
	M8a vs M8	39.168	<0.001	See M7 vs M8 comparison
dlx3a	M0 vs M3	12.605	0.013	-
	M1a vs M2a	3.858	ns	-
	M7 vs M8	4.258	ns	-
	M8a vs M8	3.872	0.025	*37S*
dlx4b	M0 vs M3	39.110	<0.001	-
	M1a vs M2a	16.940	<0.001	*48Q*, *135A*, **145 T**
	M7 vs M8	17.367	<0.001	*48Q*, *135A*, **145 T**
	M8a vs M8	16.931	<0.001	See M7 vs M8 comparison

LRT values, p-values and positively selected sites identified by the BEB ($p < 0.01$ (in bold) and $p < 0.05$ (in italic); CodeML) are shown.

within the same cluster. Furthermore, the mode of selection seems to differ between members of the same dlx clusters as well. While strong purifying selection was observed for *dlx3b*, *dlx4a* and *dlx5a*, their paired cluster members *dlx4b, dlx3a* and *dlx6a* show sign of elevated ω-values. A notable exception to this observation is the *dlx1a-dlx2a* cluster. For both genes a proportion of sites was found with elevated ω-values (note that the proportion is considerably bigger for *dlx2a*). These observations indicate that although clusters are transcribed concurrently, selection seems to act on the individual gene level rather than on the level of the *dlx* gene clusters. Also the observed patterns are not in concordance with the two groups of homeobox domains that emerged from the initial tandem duplication (see [52] and Additional file 4).

Discussion

In this work, we present a detailed evolutionary characterization of the *dlx* gene repertoire in East African cichlid fishes. Previously, Renz *et al.* [35] studied the embryonic expression patterns of *dlx* genes in cichlids and showed that they are expressed in e.g., the developing jaw apparatus and anal fin, tissues that contribute to two putative evolutionary innovations: the pharyngeal jaw and the egg-spots on the anal fin of the cichlid *A. burtoni*. Here, we study the molecular evolution of *dlx* genes in a representative set of 23 East African cichlid species. We performed comparative phylogenetic analyses and detailed screens of nonsynonymous-synonymous substitution rate ratios to determine the selective pressure acting upon these candidate genes for evolutionary novelties in cichlid fishes.

Dlx3a did not get lost in the cichlid lineage

Our phylogenetic analyses of dlx proteins extends previous analyses (e.g., [35]) by the inclusion of cod [72] and four different cichlid species (i.e., *O. niloticus*, *N. brichardi*, *A. burtoni* and *P. nyererei*; BROAD Institute). Although our results agree with most of the available hypotheses on the evolutionary loss of *dlx* genes in specific teleost lineages (i.e., *dlx3a* in zebrafish and *dlx4a* in medaka), we did detect *dlx3a* in cichlids and thus refute the cichlid-specific gene loss hypothesis of *dlx3a* put forward by Renz *et al.* [35]. Not only were we able to locate this gene in all four cichlid genomes examined (Additional file 1), we also gathered partial gene sequences for this locus in all 23 cichlid species included (Additional files 4 and 6). Furthermore, in-house tblastx searches of this newly identified paralog against preliminary cichlid EST libraries (BROAD Institute, unpublished data) resulted in multiple hits, providing proof of its expression in – at least - *Astatotilapia burtoni*, *Oreochromis niloticus* and *Metriaclima zebra*.

Selection on *dlx* paralogs in relation to gene duplication events

Gene-wide estimates of the d_N/d_S ratios indicate that all loci evolved under purifying selection ($\omega < 1$), indicating strong selection against deleterious mutations, commonly observed in functional proteins. Additional analyses of individual codons indicate that the sequenced regions of *dlx3b*, *dlx4a* and *dlx5a* evolved under purifying selection, while positive selection acting on specific codons was detected for a small proportion of sites (i.e., up to 12%) for *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b* (i.e., a smaller number of positively selected sites was found with more stringent SLR analyses for *dlx2a*, *dlx3a* and *dlx4b*). Plausible reasons for the excess of nonsynonymous mutations in these loci are either lowered functional constraints or directional selection, as Sumiyama and colleagues suggested for *Dlx7* in mouse [73]. Different modes of selection are thus found to have acted on the *dlx* paralogs in cichlids after the genome duplication events.

Differential selection after gen(om)e duplication is a commonly observed phenomenon and is associated with the fate of the gene duplicates i.e., non-, sub- or neofunctionalization. Sub- and neofunctionalization are adaptive processes by which either spatial or temporal partitioning of the ancestral function or the evolution of complete new functions take place [5,74-76]. While ancestral functions can be maintained by retaining the protein sequences and preventing deleterious mutations through purifying selection, relaxed selection on the other duplicate can lead to the introduction of mutations and subsequent divergence

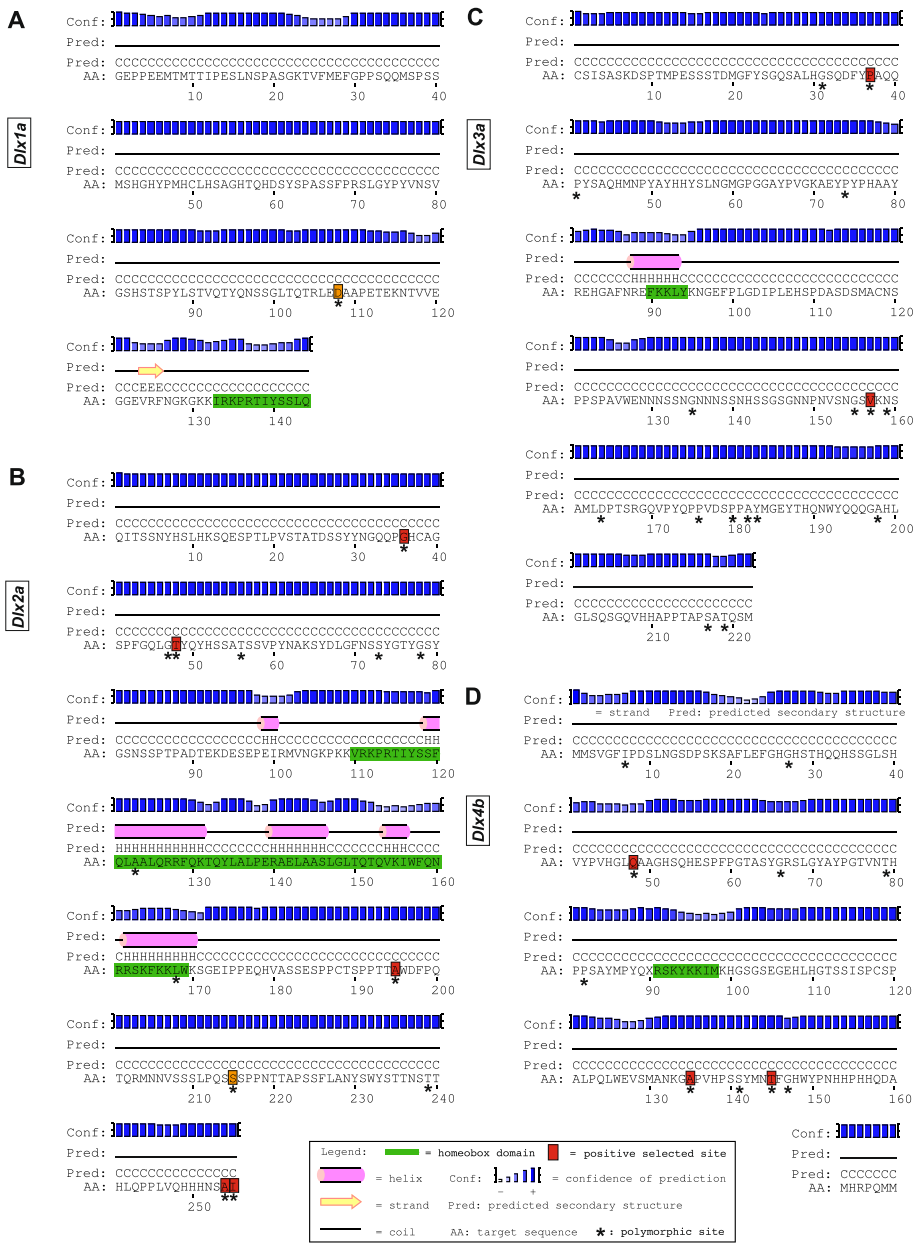


Figure 2 Secondary structure and positively selected sites for four partially sequenced *Astatotilapia burtoni* Dlx proteins. Secondary structure predictions were obtained from the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/). Positively selected sites identified by the site model analyses (CodeML) and the SLR analyses are highlighted in red (BEB and/or SLR) or orange (NEB) boxes. **(A)** Dlx1a. **(B)** Dlx2a. **(C)** Dlx3a. **(D)** Dlx4b.

[5,75,76]. Most of these changes are deleterious and are followed by the loss of the gene over time (i.e., nonfunctionalization). On rare occasions the mutations can lead to an altered function of the protein (i.e., neofunctionalization; change within the protein) or altered expression pattern (subfunctionalization; change in regulatory regions), which can be characterized by elevated ω values and the maintenance of the mutations results in divergence of the two duplicates.

Many studies have focused on duplicated genes in relation to divergence of duplicates (see e.g., [77-80] and references therein). An interesting case of subfunctionalization was described in leaf-eating Colobine monkeys, in which the pancreatic ribonuclease gene (*RNASE1*), necessary to digest its specialized diet, was duplicated [81,82]. Although the two gene-products are used in the same process (i.e., digestion of bacterial RNA), the duplicate gene shows many substitutions, while the ancestral locus

Table 5 Amino acid substitution and their predicted effect on function for the eight cichlid *dlx* loci

Locus	L	S	E
dlx1a	144	1	Tolerated
dlx2a	255	13	Tolerated
dlx3a	222	16	Tolerated
dlx3b	160	4	Tolerated
dlx4a	96	1	Tolerated
dlx4b	167	10	Tolerated
dlx5a	271	0	-
dlx6a	236	2	Tolerated

Length of obtained protein sequence (L), the number of amino acid substitutions (S) and the predicted effect of the substitutions on the function generated by the SIFT analyses (E) are shown.

did not change [81]. Similar patterns of heterogeneity in amino acid substitutions or differential selection were also observed by Dermitzakis and Clark [83] between duplicates of several developmental gene families (e.g., *Notch*, *Bmp* and *Hox9*) in mouse and human. Interestingly, differential selection regimes acting on paralogs were also found in the murine *Dlx3-Dlx7* cluster, with *Dlx7* evolving more rapidly than *Dlx3* [73]. Our results of differential selection acting on the cichlid *dlx* paralogs are thus comparable to previously studied cases of duplicated genes. We even detect a similar pattern as Sumiyama *et al.* [73], with *dlx4b* evolving more rapidly than *dlx3b* (i.e., relative tree length 0.937 vs 0.609).

The adaptive protein evolution as observed in *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b* together with the evolutionary history of the gene family, could thus be a sign of possible new or altered functions of these *dlx* paralogs in cichlids. Although we did not observe amino acid substitutions with predicted apparent effect on the protein function in our partial sequences, other mechanisms, such as *cis*-regulatory evolution might have altered the expression patterns after gene duplication. Gene expression analyses in cichlid and zebrafish indicate that clusters are often transcribed concurrently and that the *dlx* duplicates exhibit overlapping expression patterns in particular during the development of brain and pharyngeal arches [35,44,45]. This co-expression of the *dlx* clusters is controlled through intergenic *cis*-regulatory regions [35,42,43]. While mutations in these regions are expected to affect the expression of both paralogs, changes in the coding regions of the *dlx* loci are likely to affect the individual *dlx* locus' function, which could lead to neofunctionalization.

Selection pressure on *dlx* paralogs in relation to evolutionary innovations

We found an interesting pattern comparing our d_N/d_S results with the expression patterns found by Renz *et al.* [35] in relation to evolutionary novelties in cichlids. In the developing pharyngeal teeth and the anal fin *dlx3b, dlx4a* (not in anal fin) and *dlx5a*, the exact loci for which we found strong patterns of purifying selection, are expressed at high levels. Although this observation seems to contradict other cases in which candidate genes showed accelerated rates of protein evolution (see [31,53,84]), they do not stand alone (see e.g., [36]). It has been shown that minor changes in the complex genetic pathways underlying the development of morphological structures can lead to the evolution of novelties (see e.g., [85]). Furthermore, many cases of morphological adaptation are driven by *cis*-regulatory evolution (reviewed in [86]). Several intergenic *cis*-regulatory elements have been identified in the *dlx* clusters in *A. burtoni* by Renz *et al.* [35], but the functional characterization in cichlids is yet to be performed. It is thus possible that only a small fraction of genes involved in the evolutionary novelties in cichlids show signs of adaptive evolution and that the three *dlx* loci were co-opted for their ancestral functions.

According to Renz *et al.* [35], the five *dlx* genes for which we found signatures of positive selection, are either not expressed at all or at low levels during pharyngeal teeth and anal fin development in the cichlid *A. burtoni*. Low levels of gene expression were observed for *dlx2a* in the developing pharyngeal teeth in cichlids [35], while higher *dlx2a* expression levels were observed in other teleost species [33,44,45]. *Dlx4b* and *dlx6a* expression has previously been shown in the developing pharyngeal teeth of zebrafish and/or medaka [44,45], but has not been observed in cichlids (yet). Furthermore, multiple *dlx* genes, including loci with signatures of positive selection, appear to be expressed in the developing anal fin tissue at time points coinciding with egg-spot development in *A. burtoni* (E. Santos, personal communications). Therefore, it is likely that several *dlx* paralogs, for which we found signs of positive selection, are involved in the development of evolutionary innovations in cichlids, in contrast to the initial findings of Renz *et al.* [35]. Future detailed and extended functional analyses should be conducted to elucidate their role in the development of these evolutionary important traits in cichlid fishes.

Conclusions

In this study, we provide an in depth molecular evolutionary analysis of the *dlx* gene repertoire in teleost fishes. We located and generated partial sequences for *dlx3a* in 23 East African cichlid species, refuting the hypothesis of Renz *et al.* [35] that *dlx3a* got lost in the cichlid lineage. Phylogenetic analyses of the teleost *dlx* gene repertoire show that substantial differences exist in the rate of evolution among teleost *dlx* paralogs. In addition, analyses of the nonsynonymous-synonymous

substitution rates of the cichlid *dlx* paralogs revealed strong differences in the selection pressure acting upon *dlx* paralogs and cluster members. Although differential selection pressure after gene duplication is a putative sign of new or altered functions, we observed a link between the *dlx* loci under strong purifying selection, in particular, and high expression levels in two cichlids' novelties; the pharyngeal jaw and anal fin. This indicates that other mechanisms than adaptive protein evolution are likely to be involved in the co-option of these genes. Furthermore, several (preliminary) studies found that at least three other *dlx* paralogs, for which we found signs of positive selection, are actually expressed in the developing pharyngeal teeth and/or haplochromine anal fin. Hence, the *dlx* paralogs appear as candidate genes for the development of evolutionary innovations in cichlids, although further functional analyses should elucidate the role of positive selection therein.

Availability of supporting data

The datasets supporting the results of this article are publicly available in the GenBank repository under accession numbers: KC285366-KC285546 and in the treeBASE repository under submission number 14433, http://purl.org/phylo/treebase/phylows/study/TB2:S14433.

Additional files

Additional file 1: Accession numbers and/or genomic location of the teleost *dlx* sequences.

Additional file 2: Specimen information and GenBank Accession numbers.

Additional file 3: Primer information and primer sequences.

Additional file 4: Protein comparison of the teleost dlx homeobox domains. Depicted are the amino acid sequences of the homeobox domains for each of the four teleost clusters: dlx1a-dlx2a, dlx4a-dlx3a, dlx4b-dlx3b and dlx6a-dlx5a in comparison with the single Dll homeobox sequence (here depicted in duplo) of *Drosophila melanogaster*. Sequences can be divided in two groups; dlx1a, dlx4a, dlx4b and dlx6a versus dlx2a, dlx3a, dlx3b and dlx5a. The two sixty amino acid long homeobox domains of each cluster are depicted in separate boxes. The top graph displays the mean pairwise identity of all sequences (i.e., green = 100% identity and brown ≥ 30% identity). Numbers represent the amino acid position within the homeobox.

Additional file 5: Maximum likelihood gene trees based on 23 cichlid species for the eight *dlx* loci. Bootstrap values (PAUP*) and Bayesian posterior probabilities (MrBayes) above 50% are shown respectively above and below the branches. A color key for the ten studied cichlid lineages is given in the box below the figure. (**a**) *Dlx1a* (737 base pairs (bp); TPM3uf model). Two major polytomies were recovered. The lamprologines cluster together with the Boulengerochromini, Bathybatini and the Cyphotilapiini. *A. burtoni* is found at the base with *O. tanganicae*. (**b**) *Dlx2a* (1371 bp; HKY + I model). Polytomous tree with all members of the lineages Lamprologines, Ectodines, Haplochromines and Limnochromines recovered as monophyletic clades. (**c**) *Dlx3a* (666 bp; HKY model). Polytomous tree, with only the Lamprologines recovered as monophyletic clade. (**d**) *Dlx4a* (1166 bp; TPM3uf + I + G). Polytomous relationships were observed between multiple lineages, although most lineages are monophyletic except the Haplochromines (**e**) *Dlx3b* (1972 bp; GTR + I + G). Moderately resolved tree. (**f**) *Dlx4b* (722 bp; TPM3uf). Mostly polytomous relationships between species, except the Limnochromini and most members of the Lamprologines. (**g**) *Dlx5a* (1538 bp; TIM2 + G). Basal polytomy divides ingroup species except *G. permaxillaris*, in two big clades. (**h**) *Dlx6a* (1710 bp; TIM3 + G). Limnochromines, Lamprologines and Haplochromines recovered as monophyletic clades, although the relationships between lineages are largely polytomous.

Additional file 6: Four partially sequenced cichlid Dlx proteins. Depicted are the amino acid sequences of *Astatotilapia burtoni* (**a**, **c**, **d**) and *Ctenochromis horei* (**b**). Secondary structure predictions were obtained from the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/). (**a**) Dlx3b. (**b**) Dlx4a. (**c**) Dlx5a. (**d**) Dlx6a.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

ETD, FDK and WS conceived the study. FDK generated the data. ETD and FDK analyzed the data. ETD and WS wrote the paper. All authors read and approved the final manuscript.

Authors' information

ETD is a PhD student and FDK a master student in the group of WS. WS is a Professor of Zoology and Evolutionary Biology at the University of Basel. The research of his team focuses on the genetic basis of adaptation, evolutionary innovation and animal diversification of mainly the exceptionally diverse cichlid fishes.

Acknowledgements

We would like to express our gratitude to past and current members of the Salzburger lab for their contribution to sampling during fieldwork; to Brigitte Aeschbach and Nicolas Boileau for assistance during lab work; to Britta Meyer for advice on the phylogenetic analyses; to Emília Santos for help on designing the study and comments on earlier drafts of this manuscript; and to Richard Kluin for grammatical advice. We would also like to thank the BROAD institute for sharing unpublished cichlid genome sequence data with the community. The valuable suggestions of two anonymous reviewers greatly helped improving this manuscript. This study was supported by the Freiwillige Akademische Gesellschaft Basel (Dissertation support grant to ETD), European Research Council (Starting Grant "INTERGENADAPT" to WS) and the Swiss National Science Foundation (Grant 3100A0_122458 to WS).

Received: 9 May 2013 Accepted: 5 July 2013
Published: 17 July 2013

References

1. Meyer A, Van de Peer Y: **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD).** *Bioessays* 2005, **27:**937–945.
2. Volff JN: **Genome evolution and biodiversity in teleost fish.** *Heredity* 2005, **94:**280–294.
3. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y: **Genome duplication, a trait shared by 22,000 species of ray-finned fish.** *Genome Res* 2003, **13:**382–390.
4. Taylor JS, Van de Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Phil Trans R Soc Lond B* 2001, **356:**1661–1679.
5. Ohno S: *Evolution by gene duplication.* New York: Springer Verlag; 1970.
6. Santini F, Harmon LJ, Carnevale G, Alfaro ME: **Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes.** *BMC Evol Biol* 2009, **9:**194.
7. De Rosa R, Grenier JK, Andreevas T, Cook CE, Adoutte A, Akam M, Carroll SB, Balavoine G: **Hox genes in brachiopods and priapulids and protostome evolution.** *Nature* 1999, **399:**772–776.
8. Lemons D, McGinnis W: **Genomic evolution of Hox gene clusters.** *Science* 2006, **313:**1918–1922.
9. Ruddle FH, Bartels JL, Bentley KL, Kappen C, Murtha MT, Pendleton JW: **Evolution of Hox genes.** *Annu Rev Genet* 1994, **28:**423–442.
10. Gehring WJ, Hiromi Y: **Homeotic genes and the homeobox.** *Annu Rev Genet* 1986, **20:**147–173.

11. Cohn MJ, Tickle C: **Limbs: a model for pattern formation within the vertebrate body plan.** *Trends Genet* 1996, **12:**253–257.
12. Zakany J, Duboule D: **The role of Hox genes during vertebrate limb development.** *Curr Opin Genet Dev* 2007, **17:**359–366.
13. Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selegue J, Carroll SB: ***Ultrabithorax* function in butterfly wings and the evolution of insect wing patterns.** *Curr Biol* 1999, **9:**109–115.
14. Warren RW, Nagy L, Selegue J, Gates J, Carroll SB: **Evolution of homeotic gene regulation and function in flies and butterflies.** *Nature* 1994, **372:**458–461.
15. Pick L, Heffer A: **Hox gene evolution: multiple mechanisms contributing to evolutionary novelties.** *Ann N Y Acad Sci* 2012, **1256:**15–32.
16. Pearson JC, Lemons D, McGinnis W: **Modulating Hox gene functions during animal body patterning.** *Nat Rev Genet* 2005, **6:**893–904.
17. Seehausen O: **African cichlid fish: a model system in adaptive radiation research.** *Proc R Soc B* 2006, **273:**1987–1998.
18. Salzburger W: **The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes.** *Mol Ecol* 2009, **18:**169–185.
19. Kornfield I, Smith PF: **African cichlid fishes: model systems for evolutionary biology.** *Annu Rev Ecol Syst* 2000, **31:**163–196.
20. Kocher TD: **Adaptive evolution and explosive speciation: the cichlid fish model.** *Nat Rev Genet* 2004, **5:**288–298.
21. Santos ME, Salzburger W: **How cichlids diversify.** *Science* 2012, **338:**619–621.
22. Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL: **How many species of cichlid fishes are there in African lakes?** *Mol Ecol* 2001, **10:**793–806.
23. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF: **Age of cichlids: new dates for ancient lake fish radiations.** *Mol Biol Evol* 2007, **24:**1269–1282.
24. Verheyen E, Salzburger W, Snoeks J, Meyer A: **Origin of the superflock of cichlid fishes from Lake Victoria, East Africa.** *Science* 2003, **300:**325–329.
25. Barlow GW: *The cichlid fishes: nature's grand experiment in evolution.* Cambridge: Perseus publishing; 2000.
26. Coulter GW: *Lake Tanganyika and its life.* Oxford: British Museum (Natural History) and Oxford University Press; 1991.
27. Fryer G, Iles TD: *The cichlid fishes of the Great Lakes of Africa: their biology and Evolution.* Edinburgh: Oliver & Boyd; 1972:1–334.
28. Liem KF: **Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws.** *Systematic Zoology* 1973, **22:**425–441.
29. Muschick M, Indermaur A, Salzburger W: **Convergent evolution within an adaptive radiation of cichlid fishes.** *Curr Biol* 2012, **22:**2362–2368.
30. Salzburger W, Mack T, Verheyen E, Meyer A: **Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes.** *BMC Evol Biol* 2005, **5:**17.
31. Salzburger W, Braasch I, Meyer A: **Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes.** *BMC Biol* 2007, **5:**51.
32. Roberts RB, Ser JR, Kocher TD: **Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes.** *Science* 2009, **326:**998–1001.
33. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Streelman JT: **An ancient gene network is co-opted for teeth on old and new jaws.** *Plos Biol* 2009, **7:**e31.
34. Fraser GJ, Bloomquist RF, Streelman JT: **A periodic pattern generator for dental diversity.** *BMC Biol* 2008, **6:**32.
35. Renz AJ, Gunter HM, Fischer JM, Qiu H, Meyer A, Kuraku S: **Ancestral and derived attributes of the dlx gene repertoire, cluster structure and expression patterns in an African cichlid fish.** *EvoDevo* 2011, **2:**1.
36. Diepeveen ET, Salzburger W: **Molecular characterization of two endothelin pathways in East African cichlid fishes.** *J Mol Evol* 2011, **73:**355–368.
37. Kuraku S, Meyer A: **Genomic analysis of cichlid fish 'natural mutants'.** *Curr Opin Genet Dev* 2008, **18:**551–558.
38. Parichy DM, Ransom DG, Paw B, Zon LI, Johnson SL: **An orthologue of the *kit*-related gene *fms* is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, *Danio rerio*.** *Development* 2000, **127:**3031–3044.
39. Parichy DM, Turner JM: **Temporal and cellular requirements for Fms signaling during zebrafish adult pigment pattern development.** *Development* 2003, **130:**817–833.
40. Albertson RC, Streelman JT, Kocher TD, Yelick PC: **Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies.** *P Natl Acad Sci Usa* 2005, **102:**16287–16292.
41. Hoegg S, Boore JL, Kuehl JV, Meyer A: **Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*.** *BMC Genomics* 2007, **8:**317.
42. Panganiban G, Rubenstein J: **Developmental functions of the *Distal-less*/Dlx homeobox genes.** *Development* 2002, **129:**4371–4386.
43. Ellies DL, Stock DW, Hatch G, Giroux G, Weiss KM, Ekker M: **Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish *dlx* genes.** *Genomics* 1997, **45:**580–590.
44. Borday-Birraux V, Van der Heyden C, Debiais-Thibaud M, Verreijdt L, Stock DW, Huysseune A, Sire J-Y: **Expression of Dlx genes during the development of the zebrafish pharyngeal dentition: evolutionary implications.** *Evol Dev* 2006, **8:**130–141.
45. Debiais-Thibaud M, Germon I, Laurenti P, Casane D, Borday-Birraux V: **Low divergence in *Dlx* gene expression between dentitions of the medaka (*Oryzias latipes*) versus high level of expression shuffling in osteichtyans.** *Evol Dev* 2008, **10:**464–476.
46. Beldade P, Brakefield PM, Long AD: **Contribution of *Distal-less* to quantitative variation in butterfly eyespots.** *Nature* 2002, **415:**315–318.
47. Reed RD, Serfas MS: **Butterfly wing pattern evolution is associated with changes in a Notch/Distal-less temporal pattern formation process.** *Curr Biol* 2004, **14:**1159–1166.
48. Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A, French V, Carroll SB: **Development, plasticity and evolution of butterfly eyespot patterns.** *Nature* 1996, **384:**236–242.
49. Sunkel CE, Whittle J: ***Brista*: A gene involved in the specification and differentiation of distal cephalic and thoracic structures in *Drosophila melanogaster*.** *Roux's Arch Dev Biol* 1987, **196:**124–132.
50. Dong P, Chu J, Panganiban G: **Coexpression of the homeobox genes Distal-less and homothorax determines Drosophila antennal identity.** *Development* 2000, **127:**209–216.
51. Gordon CT, Brinas I, Rodda FA, Bendall AJ, Farlie PG: **Role of *Dlx* genes in craniofacial morphogenesis: *Dlx2* influences skeletal patterning by inducing ectomesenchymal aggregation in ovo.** *Evol Dev* 2010, **12:**459–473.
52. Stock DW, Ellies DL, Zhao ZY, Ekker M, Ruddle FH, Weiss KM: **The evolution of the vertebrate Dlx gene family.** *P Natl Acad Sci Usa* 1996, **93:**10858–10863.
53. Terai Y, Morikawa N, Okada N: **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* 2002, **19:**1628–1632.
54. Notredame C, Higgins DG, Heringa J: **T-Coffee: a novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302:**205–217.
55. Poirot O, O'Toole E: **Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments.** *Nucleic Acids Res* 2003, **31:**3503–3506.
56. Swofford DL: *Phylogenetic Analysis Using Parsimony (*and Other Methods).* Sunderland: Sinauer Associates; 2003.
57. *Geneious version 5.6 created by Biomatters.* Available from http://www.geneious.com/.
58. Bruford MW, Hanotte O, Brookfield J, Burke T: **Multilocus and single-locus DNA fingerprinting**. In *Molecular genetic analysis of populations: a practical approach.* Edited by Hoelzel AR. Oxford: Oxford University Press; 1998:287–336.
59. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2001, **17:**754–755.
60. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572–1574.
61. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum-likelihood.** *Syst Biol* 2003, **52:**696–704.
62. Posada D: **jModelTest: phylogenetic model averaging.** *Mol Biol Evol* 2008, **25:**1253–1256.
63. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C: **Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas.** *Systematic Biology* 2002, **51:**113–135.
64. Yang Z: **PAML, A program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13:**555–556.
65. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24:**1586–1591.
66. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148:**929–936.

67. Yang Z, Nielsen R, Goldman N: **Codon- substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155:**431–449.
68. Yang Z, Wong W, Nielsen R: **Bayes empirical Bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22:**1107–1118.
69. Kosakovsky Pond SL, Frost S, Muse SV: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005, **22:**676–679.
70. Massingham T, Goldman N: **Detecting amino acid sites under positive selection and purifying selection.** *Genetics* 2005, **169:**1753–1762.
71. Ng P, Henikoff ST: **SIFT: predicting amino acid changes that affect protein functions.** *Nucleic Acids Research* 2003, **31:**3812–3814.
72. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel J-H, Aken B, Andersen Ø, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjøen T, Kuhl H, *et al*: **The genome sequence of Atlantic cod reveals a unique immune system.** *Nature* 2011, **477:**207–210.
73. Sumiyama K, Irvine SQ, Stock DW, Weiss KM, Kawasaki K, Shimizu N, Shashikant CS, Miller W, Ruddle FH: **Genomic structure and functional control of the *Dlx3-7* bigene cluster.** *P Natl Acad Sci Usa* 2002, **99:**780–785.
74. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151:**1531–1545.
75. Sidow A: **Gen(om)e duplications in the evolution of early vertebrates.** *Curr Opin Genet Dev* 1996, **6:**715–722.
76. Ohta T: **Simulating evolution by gene duplication.** *Genetics* 1987, **115:**207–213.
77. Marotta M, Piontkivska H, Tanaka H: **Molecular trajectories leading to the alternative fates of duplicate genes.** *PLoS ONE* 2012, **7:**e38958.
78. Conrad B, Antonarakis SE: **Gene duplication: a drive for phenotypic diversity and cause of human disease.** *Annu Rev Genom Human Genet* 2007, **8:**17–35.
79. Prince VE, Pickett FB: **Splitting pairs: the diverging fates of duplicated genes.** *Nature Reviews Genetics* 2002, **3:**827–837.
80. Braasch I, Volff JN, Schartl M: **The endothelin system: evolution of vertebrate-specific ligand-receptor interactions by three rounds of genome duplication.** *Mol Biol Evol* 2009, **26:**783–799.
81. Zhang J, Zhang Y-P, Rosenberg HF: **Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey.** *Nat Genet* 2002, **30:**411–415.
82. Zhang J: **Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys.** *Nat Genet* 2006, **38:**819–823.
83. Dermitzakis ET, Clark AG: **Differential selection after duplication in mammalian developmental genes.** *Mol Biol Evol* 2001, **18:**557–562.
84. Terai Y, Morikawa N, Kawakami K, Okada N: **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* 2002, **19:**574–578.
85. Wagner GP, Lynch VJ: **Evolutionary novelties.** *Current Biology* 2010, **20:**R48–R52.
86. Prud'homme B, Gompel N, Carroll SB: **Emerging principles of regulatory evolution.** *P Natl Acad Sci Usa* 2007, **104:**8605–8612.

doi:10.1186/1471-2148-13-153
Cite this article as: Diepeveen *et al.*: Sequence analyses of the *distal-less homeobox* gene family in East African cichlid fishes reveal signatures of positive selection. *BMC Evolutionary Biology* 2013 **13**:153.

CHAPTER 3

SEQUENCE ANALYSES OF THE *DISTAL-LESS HOMEOBOX* GENE FAMILY IN EAST AFRICAN CICHLID FISHES REVEAL SIGNATURES OF POSITIVE SELECTION

SUPPORTING MATERIAL

Additional File S1

Ensemble/EMBL accession numbers or genomic location of teleost *dlx* sequences

Species	Ensemble/GenBank accession numbers and preliminary genome information			
	dlx1a	***dlx2a***	***dlx3a***	***dlx3b***
Danio rerio	ENSDARG00000013125	ENSDARG00000079964	-	ENSDARG00000014626
Gradus morhua	ENSGMOG00000002250	ENSGMOG00000002233	ENSGMOG00000018564	ENSGMOG00000013282
Gasterosteus aculeatus	ENSGACG00000004931	ENSGACG00000004914	ENSGACG00000005117	ENSGACG00000009888
Tetraodon nigroviridis	ENSTNIG00000016938	ENSTNIG00000016939	ENSTNIG00000011165	ENSTNIG00000012856
Takifugu rubripes	ENSTRUG00000017741	ENSTRUG00000017732	ENSTRUG00000007563	ENSTRUG00000016555
Oryzias latipes	ENSORLG00000017364	ENSORLG00000017372	scaffold784[3]	ENSORLG00000004116
Oreochromis niloticus	ENSONIG00000008725	ENSONIG00000008722	ENSONIG00000006437	ENSONIG00000019897
Astatotilapia burtoni[1]	scaffold_41	scaffold_41	scaffold_363	scaffold_149
Astatotilapia burtoni[2]	FN667596	FN667597	-	FN667598
Neolamprologus brichardi[1]	scaffold_13	scaffold_13	scaffold_134	scaffold_76
Pundamilia nyererei[1]	scaffold_8	scaffold_8	scaffold_294	scaffold_4

Additional File 1 cont.

Species	Ensemble/GenBank accession numbers and preliminary genome information			
	dlx4a	***dlx4b***	***dlx5a***	***dlx6a***
Danio rerio	ENSDARG00000011956	ENSDARG00000071560	ENSDARG00000042296	ENSDARG00000042291
Gradus morhua	ENSGMOG00000018574	ENSGMOG00000013275	ENSGMOG00000013639	ENSGMOG00000013631
Gasterosteus aculeatus	ENSGACG00000005120	ENSGACG00000009881	ENSGACG00000006792	ENSGACG00000006780
Tetraodon nigrovirdis	ENSTNIG00000011166	ENSTNIG00000012857	ENSTNIG00000017761	ENSTNIG00000017762
Takifugu rubripes	ENSTRUG00000007722	ENSTRUG00000016558	ENSTRUG00000000270	ENSTRUG00000007566
Oryzias latipes	-	ENSORLG00000004106	ENSORLG00000004561	ENSORLG00000004542
Oreochromis niloticus	ENSONIG00000006433	ENSONIG00000019896	ENSONIG00000010755	ENSONIG00000010756
Astatotilapia burtoni[1]	scaffold_363	scaffold_149	scaffold_97	scaffold_97
Astatotilapia burtoni[2]	FN667600	FN667599	FN667601	FN667602
Neolamprologus brichardi[1]	scaffold_134	scaffold_76	scaffold_12	scaffold_12
Pundamilia nyererei[1]	scaffold_294	scaffold_4	scaffold_29	scaffold_29

[1] v1 assembly of the preliminary whole genome sequences by the Cichlid Genome Consortium (cichlid.umd.edu/CGCindex.html)

[2] cDNA sequences by Renz *et al*

[3] Identified by Debiais-Thibaud *et al*. (2008); uncharacterized gene/protein could not be retrieved in Ensemble and uniprot

Additional File S2

Specimen information and GenBank accession numbers

Species	Tribe	GenBank accession numbers			
		Dlx1a	**Dlx2a**	**Dlx3a**	**Dlx3b**
		Dlx4a	**Dlx4b**	**Dlx5a**	**Dlx6a**
Bathybates graueri	Bathybatini	KC285376	KC285403	KC285413	KC285454
		KC285458	KC285481	KC285513	KC285526
Boulengerochromis microlepis	Boulengerochromini	KC285381	KC285404	KC285414	KC285440
		KC285459	KC285482	KC285518	KC285527
Cyphotilapia frontosa	Cyphotilapiini	KC285382	KC285397	KC285415	KC285441
		KC285462	KC285484	KC285519	KC285528
Cyprichromis leptosoma	Cyprichromini	KC285383	KC285398	KC285418	KC285442
		KC285463	KC285485	KC285520	KC285529
Callochromis macrops	Ectodini	KC285384	KC285399	KC285419	KC285443
		KC285479	KC285493	KC285521	-
Cyathopharynx furcifer	Ectodini	KC285385	KC285400	KC285416	KC285444
		KC285461	KC285483	KC285503	KC285530
Astatotilapia burtoni	Haplochromini	KC285386	KC285401	KC285411	KC285445
		KC285478	KC285480	KC285522	KC285531
Ctenochromis horei	Haplochromini	KC285387	KC285402	KC285417	KC285446
		KC285460	KC285496	KC285523	KC285532
Altolamprologus fasciatus	Lamprologini	KC285373	KC285393	KC285412	KC285437
		KC285457	KC285502	KC285510	KC285540
Lepidiolamprologus elongatus	Lamprologini	KC285374	KC285394	KC285422	KC285438
		KC285465	KC285489	KC285511	KC285541
Neolamprologus furcifer	Lamprologini	KC285388	KC285406	KC285425	KC285450
		KC285468	KC285490	KC285524	KC285533
Neolamprologus pulcher	Lamprologini	KC285366	KC285407	KC285426	KC285451
		KC285469	KC285491	KC285504	KC285534
Variabilichromis moorii	Lamprologini	KC285375	KC285395	KC285433	KC285453
		KC285477	KC285492	KC285512	KC285542
Gnathochromis permaxillaris	Limnochromini	KC285377	KC285410	KC285420	KC285455
		KC285464	KC285501	KC285514	KC285543
Limnochromis staneri	Limnochromini	KC285367	KC285389	KC285424	KC285434
		KC285466	KC285486	KC285505	KC285535
Perissodus microlepis	Perissodini	KC285368	-	KC285430	KC285452
		KC285471	KC285487	KC285506	KC285536
Plecodus straeleni	Perissodini	KC285369	KC285408	KC285431	KC285447
		KC285473	KC285488	KC285507	-
Oreochromis tanganicae	Tilapiini	KC285378	KC285409	KC285427	KC285456
		KC285470	KC285494	KC285515	KC285544
Limnotilapia dardenii	Tropheini	KC285372	KC285392	KC285421	KC285436
		KC285476	KC285495	KC285509	KC285539
Lobochilotes labiatus	Tropheini	KC285370	KC285390	KC285423	KC285449
		KC285467	KC285497	KC285525	KC285537
Petrochromis famula	Tropheini	KC285379	KC285405	KC285429	KC285448
		KC285472	KC285498	KC285516	KC285545
Pseudosimochromis curvifrons	Tropheini	KC285380	KC285396	KC285428	KC285439
		KC285474	KC285499	KC285517	KC285546
Tropheus moori	Tropheini	KC285371	KC285391	KC285432	KC285435
		KC285475	KC285500	KC285508	KC285538

Additional File S3

Primer Sequences

Locus	Forward sequence 5' - 3'	Reverse sequence 5' - 3'
dlx1a	CCGTCCTTCACTGAGATCAT	CTCCGGTAGCGCCAAATAC
dlx2a	CTCTCCCCCAGTTCCAAGAT	TTTCTGGCTCGCTCTCATCT
	TCGAACCAGATTACCTCAA	TACACCCGACGTTTTCTCGT
dlx3a	GTTCCATCTCGGCTTCCA	ACCTCYCTGTTGAATGCT
	GTCGCTCCAAGTTTAAGAAG	AGTAGACGGCTCCCATACTC
dlx3b	TGCGAGTATTTTGACCGATCT	CGAGGTATTGTGCCTTCTGG
	AGGTTCCCTTGGAGCACAG	ACAGTGCAGTCGTTCCGTTT
dlx4a	TTACAACAAACCGGGAATCG	GGACTGTGGTAGGGTGGGTA
	GGACTGTGGTAGGGTGGGTA	CCTCCTCATACAAATGAAAAGC
dlx4b	CTGTTCGTGATGTAAAGACG	GCTGTTGCTGTGATTGCTGT
	GCGCTCCAAGTACAAAAAGA	TTTTATTCTCGCACGGCATT
dlx5a	ACGCGCTCTCTCTCGAAGTA	GAGGTGAGGATGGTGAATGG
	GAGGTGAGGATGGTGAATGG	GACTGCCATCAAATACATTGC
dlx6a	AGTGCTCTCAAGGCAGAAAA	TCCGGTAAAGCGAGGTACTG
	AGCAGCAGAAGACGACAGTG	ACCTTGCTTTCGTCGTCAGT

Additional File S4
Input tree file and sequence files (fasta format) for the PAML analyses

Input tree file (based on a concatenated dataset of 9.2 kb):

((((((((((Npul:0.00148,(Vmoo:0.003181,Nfur:0.003007):0.000445):0.002076,(Afas:0.003877,Lelo:0.002103):0.002711):0.004143,(((Bgra:0.005906,Otan:0.014106):0.000638,Bmic:0.002503):0.001119,Cfro:0.002355):0.000394):0.000325,(Lsta:0.004174,Gper:0.002544):0.001405):0.000644,((Pmic:0.001325,Pstr:0.000914):0.002134,Clep:0.00461):0.001363):0.000718,(Cmac:0.003115,Cfur:0.003039):0.002852):0.001739,Abur:0.002978):0.001436,Tmoo:0.002356):0.000786,Llab:0.001519):0.000563,Pfam:0.001172):0.000257,Chor:0.002166):0.000139,Ldar:0.000754,Pcur:0.002183)

Dlx1a input file:

>dlx1a_Neopul
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Limsta
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Permic
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACACAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTTGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Plesta
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTTGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Loblab
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Tromoo
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Tylpol
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Altfas
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGAGGCAGCGCCAGAAACAG
AGAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Lepelo
GGGGAACCGCCAGA--
AGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCCTTCCTCCATGTCTCACG
GACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATACGTAAACTCCGTCGGCA
GCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGAGGCAGCGCCAGAAACAGAGAAAAACACAGTGGT
GGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Varmoo
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Batgra
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Gnaper
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Oretan
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGTTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Petfam
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Psecur
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Boumic
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA

GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Cypfro
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Cyplep
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAATGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Calmac
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Cyafur
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Astbur
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAAGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Ctehor
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA
>dlx1a_Neofur
GGGGAACCGCCAGAAGAGATGACAATGACTACGATTCCAGAAAGTCTTAATAGCCCTGCCTCAGGAAAAACCGTTTTCATGGAGTTTGGACCCCCGAGTCAACAAATGTCGCC
TTCCTCCATGTCTCACGGACACTATCCCATGCACTGTTTACATTCTGCAGGTCACACACAGCACGACAGCTACAGTCCAGCCTCGTCGTTCCCCAGATCCTTGGGTTATCCATAC
GTAAACTCCGTCGGCAGCCATTCCACCAGTCCATATCTCAGCACAGTACAGACTTACCAAAACAGCTCGGGACTCACGCAGACACGATTAGAGGATGCAGCGCCAGAAACAGA
GAAAAACACAGTGGTGGAAGGCGGAGAGGTGCGCTTCAACGGCAAAGGGAAAAAGATTAGAAAACCAAGGACTATATATTCGAGTTTACAA

Dlx2a input file:

>dlx2a_Limsta
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAATCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGCGCACTG
CGCCGGGTCACCGTTTGGACAACTGAGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGCGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAAGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Loblab
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCTCTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Tromoo
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCTCTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCGCTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Tylpol
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCTCTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGGTGCGCTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Altfas
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGCGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCACCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Lepelo
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGCGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAG----------
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCACCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Varmoo
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGCGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGTCCTA
>dlx2a_Psecur
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG

CGCCGGGTCACCGTTTGGACAACTGGGCTCTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCGCTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Cypfro
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Cyplep
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCGCGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGCTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Calmac
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCAACGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAGATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGTTCTAGTTTACCTCAGAGCAACAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACCACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Cyafur
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGTTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Astbur
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCGCTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTACCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Ctehor
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCTCTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCGCTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA
>dlx2a_Batgra
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACCACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGC----------
>dlx2a_Boumic
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCT----------
>dlx2a_Petfam
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCTCTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCGCTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCT----------
>dlx2a_Neofur
CAGATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGCGCACTG
CGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTAC
ATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACACAGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCT
CGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAA
ACACAGGTTAAAATCTGGTTCCAAAACCGGCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGTCCTA
>dlx2a_Neopul
--
GATTACCTCAAGCAATTACCACAGCTTGCACAAATCGCAGGAGTCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGCGCACTGCG
CCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAGCAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTACAT
ACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAG----------

ACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCAC
GTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTT
GGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGTCCATA
>dlx2a_Plesta

TCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTGCGCCGGGTCACCGTTTGGACAACTGGGCACTTACCAGTACCACAG
CAGCGCCACGAGTTCTGTGCCATATAACGCGAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTACATACGGTTCTTATGGCTCCAACTCATCGCCAACTCCCGCAGACAC

AGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTCAGGAAACCTCGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGG
TTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAAACACAGGTTAAAATCTGGTTCCAAAACCGTCGCTCCAAGTTCAAG
AAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCACGTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAAT
GAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTTGGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTG
CAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA

>dlx2a_Oretan

TCCCCGACTCTGCCGGTTTCCACGGCGACGGACAGCAGCTATTACAACGGCCAGCAGCCTGGGCACTGCGCCGGGTCACCGTTTGGACAGCTGGGCACTTACCAGTACCACAG
CAGCGCCACGAGTTCTGTGCCATATAACGCAAAGTCATACGACCTCGGTTTCAACTCATCGTATGGTACATACGGTTCTTATGGCTCTAACTCATCGCCAACTCCCGCAGACAC
AGAGAAAGATGAGAGCGAGCCAGAAATCCGGATGGTTAATGGAAAACCAAAGAAGGTTAGGAAACCTCGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGAGG
TTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGAGCTGGCAGCGTCGCTGGGCCTTACGCAAACACAGGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAA
GAAGCTGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATGTTGCTTCCAGTGAATCTCCCCCGTGCACGTCTCCACCAACTACCGCCTGGGATTTTCCACAGACTCAAAGAAT
GAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTCCAAACACGACTGCGCCTTCGTCGTTTTTGGCTAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTG
CAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA

>dlx2a_Gnaper

GGAAAACCAAAGAAGGTCAGGAAACCTCGAACCATTTACTCCAGCTTCCAACTGGCTGCACTTCAACGGCGGTTTCAAAAGACTCAGTATTTGGCTCTACCAGAACGGGCCGA
GCTGGCAGCGTCGCTGGGCCTTACGCAAACACAAGTTAAAATCTGGTTCCAAAACCGCCGCTCCAAGTTCAAGAAGATGTGGAAAAGTGGAGAAATCCCCCCAGAACAACATG
TTGCTTCCAGTGAATCTCCCCCGTGCACGTCTCCACCAACTACCGCCTGGGACTTTCCACAGACTCAAAGAATGAACAATGTCAGCTCTAGTTTACCTCAGAGCAGCAGCCCTC
CAAACACGACTGCGCCTTCGTCGTTTTTGGCAAACTACTCCTGGTACTCAACTACGAACTCTGCCACGCATCTGCAGCCTCCTCTGGTTCAGCATCACCACAACTCCGCCATA

Dlx3a input file:

>dlx3a_A.burtoni

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCTGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTCAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_A.fasciatus

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAAGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAATACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_B.graueri

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACAGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCARCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTATCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGTGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACACA
CCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_B.microlepis

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACAGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGTCTATATGGGGGAGTACACACA
CCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTTGGCCACACAGAGTATG

>dlx3a_C.frontosa

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_C.furcifer

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGACGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGCAGCAATGCTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_C.horei

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACTCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCGCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_C.leptosoma

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAATACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTACACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCATCGGCCACACAGAGTATG

>dlx3a_C.macrops

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCTGCAC
AGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCCC
CATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCGA
CTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAAC
GTCAGCAACGGTAGCGTCAAAAGTAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACACA
CCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_G.permaxillaris

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTGTATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_L.dardenii

CACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCACAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACT
ACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACTCTTACCCCCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAA

GCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCGACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAA
CAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAACGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGT
AGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACACACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCA
GGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_L.elongatus

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCTGCAC
AGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCCC
CATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCGA
CTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAAC
GTCAGCAATGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAATACACACAC
CAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_L.labiatus

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAATACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_L.staneri

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCC---
ACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCACAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTA
CAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCCCCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAG
CTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCGACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAAC
AACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAACGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTA
GGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACACACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAG
GCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_N.furcifer

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACAATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACAGTAGCATCAAAAATAGTGCAATGCTGAATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAATACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_N.pulcher

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAATACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_O.tanganicae

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCGTACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGGGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCGGACGCCAGCG
ATTCCATGGCTTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGCAGCAACGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCATCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCAGGTGGATTCTTCACCCGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCGCCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_P.curvifrons

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACTCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_P.famula

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_P.microlepis

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_P.straeleni

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_T.moorii

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACCATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACCCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
CCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAGTACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

>dlx3a_V.moorii

TGTTCCATCTCGGCTTCCAAGGATTCCCCGACAATGCCGGAGTCCTCCTCCACAGATATGGGCTTCTATAGCGGCCAGAGCGCGCTCCACGGCTCGCAGGATTTCTACTCGGCA
CAGCAGCCGTACTCCGCTCAGCACATGAACCCGTACGCATACCACCACTACAGCCTGAACGGAATGGGTCCCGGCGGTGCCTACCCCGTTGGGAAGGCAGAGTACCCTTACCC
TCATGCAGCATACAGGGAGCACGGAGCATTCAACAGAGAGTTTAAGAAGCTCTACAAGAACGGAGAGTTTCCGCTCGGTGATATTCCTCTTGAACACAGTCCAGACGCCAGCG
ACTCCATGGCCTGCAACTCTCCTCCATCCCCAGCTGTGTGGGAAAACAACAACAGTAGCAATGGTAACAATAACAGCAGCAACCACAGCAGCGGGAGCGGCAACAACCCAAA
CGTCAGCAACGGTAGCGTCAAAAATAGTGCAATGCTGGATCCTACCAGTAGGGGGCAGGTTCCCTATCAGCCTCCGGTGGATTCTTCACCTGCCTATATGGGGGAATACACAC
ACCAGAACTGGTACCAACAGCAGGGTGCACACTTAGGTCTCTCACAGTCAGGCCAAGTGCACCACGCACCACCAACTGCTCCGTCGGCCACACAGAGTATG

Dlx3b input file:

>dlx3b_Altfas

GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT

>dlx3b_Astbur
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTATGCCACCAAATCTGAATACCCATACACGGTGAAAGAGGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Boumic
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCACCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Calmac
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCACCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Ctehor
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Cyafur
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTATCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCACCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Cypfro
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Cyplep
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGGGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGGTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACAATCTACTCTTCCCCGCCATCGCCAGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Lepelo
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Limsta
GACTCCCCTACTCTGCCCGAGTCATCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCACCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Loblab
GACTCCCCTACTCTGCCCGAATCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGC----------
----
TCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Neopul
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGA----------
----
TCCCCGCCATCGCCCGCAGTCTGGGACAACAACAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Petfam
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATC------
TCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Plesta
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAACCGAAAAAGATTCGCAAGCCGAGGACGATCTA----
TCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Psecur
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACAGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Tromoo
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Tylpol
GACTCCCCTACTCTGCCCGAGTCGTCCGTGACGGACATGGGCTACTACAGTGGACAGACGGCCCACGGCCATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAA
CTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATACCCATACACGGTGAAAGAAGAGCCAGAGCCGGAGGTTCGTATGG
TAAATGGAAAGCCGAAAAAGATTCGCAAGCCGAGGACGATCTACTCTTCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCT
CAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGATTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGT
CCCGCAGCAAAGCGTGGGAGCTGTTTAT
>dlx3b_Neofur

CATCATGAATATTATCAGAGTCAGCCGTACGGGCAGCCCATGAACTCTTACCATCACCAGTTTAATCTGAACGGAATGGGAGCTGCTGGAGCGTACGCCACCAAATCTGAATA
CCCATACACGGTGAAAGAAGAG----------
TCCCCGCCATCGCCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT

>dlx3b_Batgra

TCCCCGCCATCGCCCGCGGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT

>dlx3b_Gnaper

TCCCCGCCATCACCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT

>dlx3b_Oretan

TCCCCGCCGTCGCCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAGCACCCGGGAACCGTTCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT

>dlx3b_Permic

TCCCCGCCATCGCCCGCTGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT

>dlx3b_Varmoo

TCCCCGCCATCGCCCGCAGTCTGGGACAACAGCAGCACCCCTCAGAACACCCCGATCAGCAGACCTCAGGTGCCGCAGCCGACGCACAGTTCATCGCCGCCGTACCTGGAGGA
TTATAATAACCACTGGTACCAGCAGGGATCACACCTACAACACCCGGGAGCCGTGCACCACCCAGTCCCGCAGCAAAGCGTGGGAGCTGTTTAT

Dlx4a input file:

>dlx4a_Altfas
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGGACTCAGATGATG

>dlx4a_Batgra
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGATTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAACA
ATTTTGGACACTGGTATCCCGGACACCAGCAGGAGCCCATGCCAAGAACTCAGATGATG

>dlx4a_Boumic
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCTGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Ctehor
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Cyafur
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Cypfro
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Cyplep
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACTTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCTTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAACA
ATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Gnaper
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Lepelo
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGGACTCAGATGATG

>dlx4a_Limsta
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Loblab
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Neofur
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Neopul
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGGACTCAGATGATG

>dlx4a_Oretan
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGGCCCTCGCAGCACGATGGGCCCTTCTCCTCCGGTGCTTCCTCTTACGGTCGCCCGCTGGTGAACA
ATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Permic
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Petfam
ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG

CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Plestr

ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Psecur

ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTTGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Tromoo

ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Tylpol

ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

>dlx4a_Varmoo

ATGACTATGAGCTCCATATCAGACACTTTAGTAACGTCCGATCCGTCCAAATCTGCGTTTTTGGAGTTCGGTGGACACAGCTACCCTGGACACCCTGGACATCAACAACCTTCG
CCAGGCTTATCCCACAACCATTATCCGGTCCACGGACTGCACGCTGTCGGACCCTCGCAGCACGATGGGCCCTTCTCCTCCGGGGCTTCCTCTTACGGTCGCCCGCTGGTGAAC
AATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGGACTCAGATGATG

>dlx4a_Astbur

TGAACAATTTTGGACACTGGTATCCCGGACACCAGCAGGAACCCATGCCAAGAACTCAGATGATG

Dlx4b input file:

>dlx4b_Astbur

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACA-
CGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCATCTCTCCCTGCTCGCCTGCATTGCCCCAGCTATGGGAGGT
CTCTATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCATTGGTATCCGAATCATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGAT
GATG

>dlx4b_Altfas

CGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCATCTCTCCCTGCTCGCCCGCATTGCCACAGCTATGGGAGGT
CTCCATGGCGAACAAAGGAGGCCCGGTCCATCCGAGCAGTTACATGAACAGTTTCGGTCACTGGTATCCGAATCATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGA
TGATG

>dlx4b_Batgra

ATGATGTCTGTGGGTTTTATCCCTGACAGTTTGAATGGATCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCACGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACAC----------

>dlx4b_Boumic

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCACGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACAGTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Cyafur

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCAGTCCGCTTACATGCCCTACCAACA-
CGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCATCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGT
CTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTATATGAACAGTTTCGGTCACTGGTATCCGAATCATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGAT
GATG

>dlx4b_Cypfro

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Cyplep

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGATACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Limsta

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Permic

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCATTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Plestr

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCATTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Lepelo

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAACACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCACAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGGCCCGGTCCATCCGAGCAGTTACATGAACAGTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Neofur

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAACACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCACAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGGCCCGGTCCATCCGAGCAGTTACATGAACAGTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Neopul

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCACAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGGCCCGGTCCATCCGAGCAGTTACATGAACAGTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Varmoo

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCATGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACGCTC----------

>dlx4b_Calmac

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTC----------

>dlx4b_Oretan

ATGATGTCTGTGGGTTTTATGCCTGACAGTCTTAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGCCACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCACGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTCTGGGCTACGCCTACCCAGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACAGTTTCAGTCACTGGTATCCGAATC
ACCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Tylpol

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACACTTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCATTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Ctehor

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCATTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Loblab

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATAGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCTATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAGTTACATGAACACTTTCGGTCATTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

>dlx4b_Petfam

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTC----------

>dlx4b_Psecur

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACA----------

>dlx4b_Tromoo

ATGATGTCTGTGGGTTTTATCCCTGACAGTTTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTC----------

>dlx4b_Gnaper

ATGATGTCTGTGGGTTTTATCCCTGACAGTCTGAATGGCTCAGATCCCTCCAAGTCGGCCTTTCTAGAATTCGGACACGGACACTCGACGCACCAGCAGCACTCCTCCGGACTCT
CTCACGTTTATCCGGTTCACGGCTTGCAAGCTGCTGGCCATTCCCAGCACGAAAGCCCTTTTCCTGGTACCGCGTCCTATGGCCGCTCTTTGGGCTACGCCTACCCCGGCACGGT
GAACACTCATCCCCCGTCCGCTTACATGCCCTACCAACACCGCTCCAAGTACAAAAAGATCATGAAGCATGGCAGCGGATCAGAGGGAGAACATCTCCACGGCACCAGCTCCA
TCTCTCCCTGCTCGCCCGCATTGCCCCAGCTATGGGAGGTCTCCATGGCGAACAAAGGAGCCCCGGTCCATCCGAGCAATTACATGAACACTTTCGGTCACTGGTATCCGAATC
ATCACCCTCATCATCAGGACGCGATGCACAGGCCTCAGATGATG

Dlx5a input file:

>dlx5a_Cyafur

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCTT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCACTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCTAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGG-
CATGGTACACCTCCGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATTCAGCATTCGTTGGCTCTTGGAGCGGGGACGTTATAT

>dlx5a_Neopul

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCCGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACATCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Limsta

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCCGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Permic

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACATAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Plesta

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACATAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Tromoo

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Tylpol

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACACAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Altfas

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Lepelo

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Varmoo

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCTGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACATCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Batgra

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAACCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Gnaper

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Oretan

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTATTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCCTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAGCTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCAGCCGTATGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Petfam

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACACAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT

>dlx5a_Psecur

GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Boumic
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAATTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Cypfro
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Cyplep
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAGCCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACATAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACCACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGAAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Calmac
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCTT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCACTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTCC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCTAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Astbur
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAATGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Ctehor
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACTTCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Neofur
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGCCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAGTATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCGCAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGCGCTG
CCGGAGAGAGCCGAACTGGCCGCCTCGCTGGGGCTAACGCAAACACAGGTGAAAATTTGGTTCCAAAACAGAAGATCAAAGATGAAGAAGATCATGAAAAACGGCGAGCTTC
CCCCGGAACACAGCCCCAGCTCCAGCGACCCCATGGCCTGCAACTCTCCACAGTCCCCGGCCGTCTGGGACACACAGGGTCCCTCCAGGCCGCACAGTCTACAGCCACAAAGC
ATCAACACGACGGCTTCTAACTTTTTGGAAAACACGGGGTCATGGTACACCTCAGCCGGCAGCTCCATGACATCCCACCTGCAGACCCCTAACTCAATACAGCATTCGTTGGCT
CTTGGAGCGGGGACGTTATAT
>dlx5a_Loblab
GCAGACTTTGAGAGCTCCTTTCAACTCTCCAGCATGCACCATCCGTCTCAGGAATCTCCTACTTTACCCGAATCCTCCGCCACGGATTCTGGCTACTACAGTCCCGCTGGAGGAG
TCGCTCACGGCTACTGCTCGCCCAGCTCCACCTCTTACGGGAAGCCTCTGAATGCCTATCAGTACCAGTACCCGGGTGTAAACGGATCTGCTGGAAATTACTCGACAAAATCCT
ACACAGATTACAGCTCCTACACGACCCCCTCCTACCACCAATATGCTGGGACTTACAGCAGAGTGCAAGCCCAGCCGAGTCCACAAGAAAAAGAGATAAGCGAGCCTGAGGT
GAGGATGGTGAATGGCAAGCCGAAGAAAGTGAGAAAACCCCGGACAATCTACTCCAGCTTCCAGCTGGCCGCCTTGCAGAGGCGGTTTCAGAACACACAGTACCTGGC----------
----

Dlx6a input file:

>dlx6a_Batgra
CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCGCAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCAGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCACGCTCGCCGAGCATCCCCCCAATCTGGGATGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA
>dlx6a_Boumic
CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA
>dlx6a_Cypfro

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCTCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Cyplep

AGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAGATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAAC
CACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTAGGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAA
GTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAACCCATCCCGGGCTCCATGTCGCTGTCCCCGCGCTCTCCGAGCATCCCCCCAATCTGGGACGTGTCGG
CCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATTCCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Cyafur

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAATAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Astbur

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCTCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Ctehor

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCAT
GTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCTGGCTACTCTCACTGGTATTC
CTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Neofur

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Neopul

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Limsta

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Permic

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCTCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Loblab

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCTGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Tromoo

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACCACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCAT
GTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCTGGCTACTCTCACTGGTATTC
CTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Tylpol

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACTCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACCACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCTGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Altfas

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG

ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Lepelo

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTTCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Varmoo

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTTCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Gnaper

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCCATCCCGGGCTCCA
TGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATT
CCTCCCCACATCAA----------

>dlx6a_Oretan

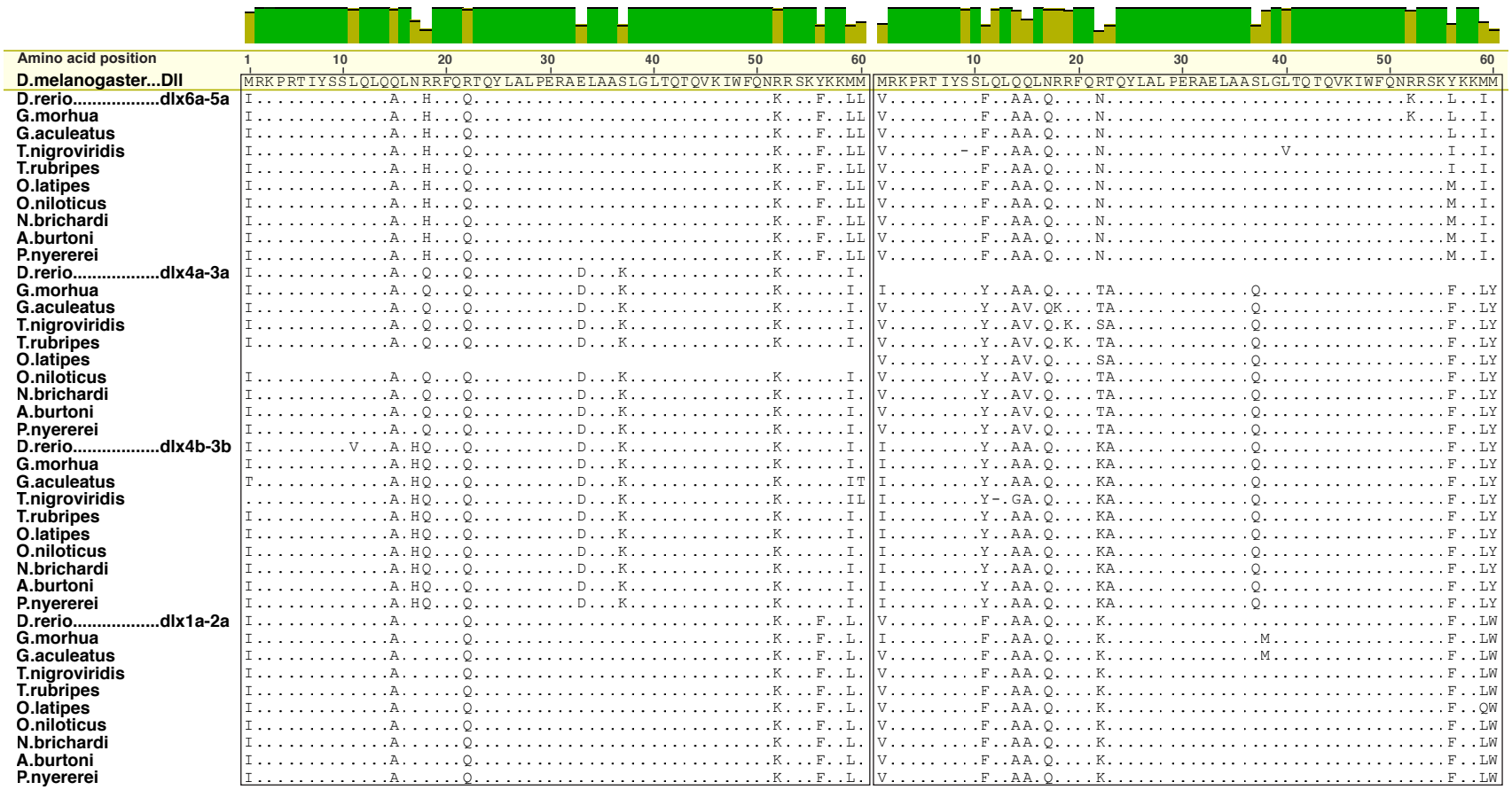
CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACCCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAGAGCAGCAGAAGACGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAG
ATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAATCACCGTTTCCAGCAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTA
GGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCTGAAGCAAGGCAGTAACCCGCATGAGAGCGAGCCCATCCCGGGCTCCAT
GTCCCTGTCCCCACGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAGTAAACATGCCGGCCAACAGCTACATGCCCGGCTACTCTCACTGGTATTC
CTCCCCACATCAAGACTCAATGCAGAGA

>dlx6a_Petfam

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACTCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAG----------
CAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAGATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAGCAA
ACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTAGGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCTGCT
GAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCTATCCCGGGCTCCATGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAGGAG
TAAACATGCCGGCCAACAGCTACATGCCTGGCTACTCTCACTGGTATTCC----------

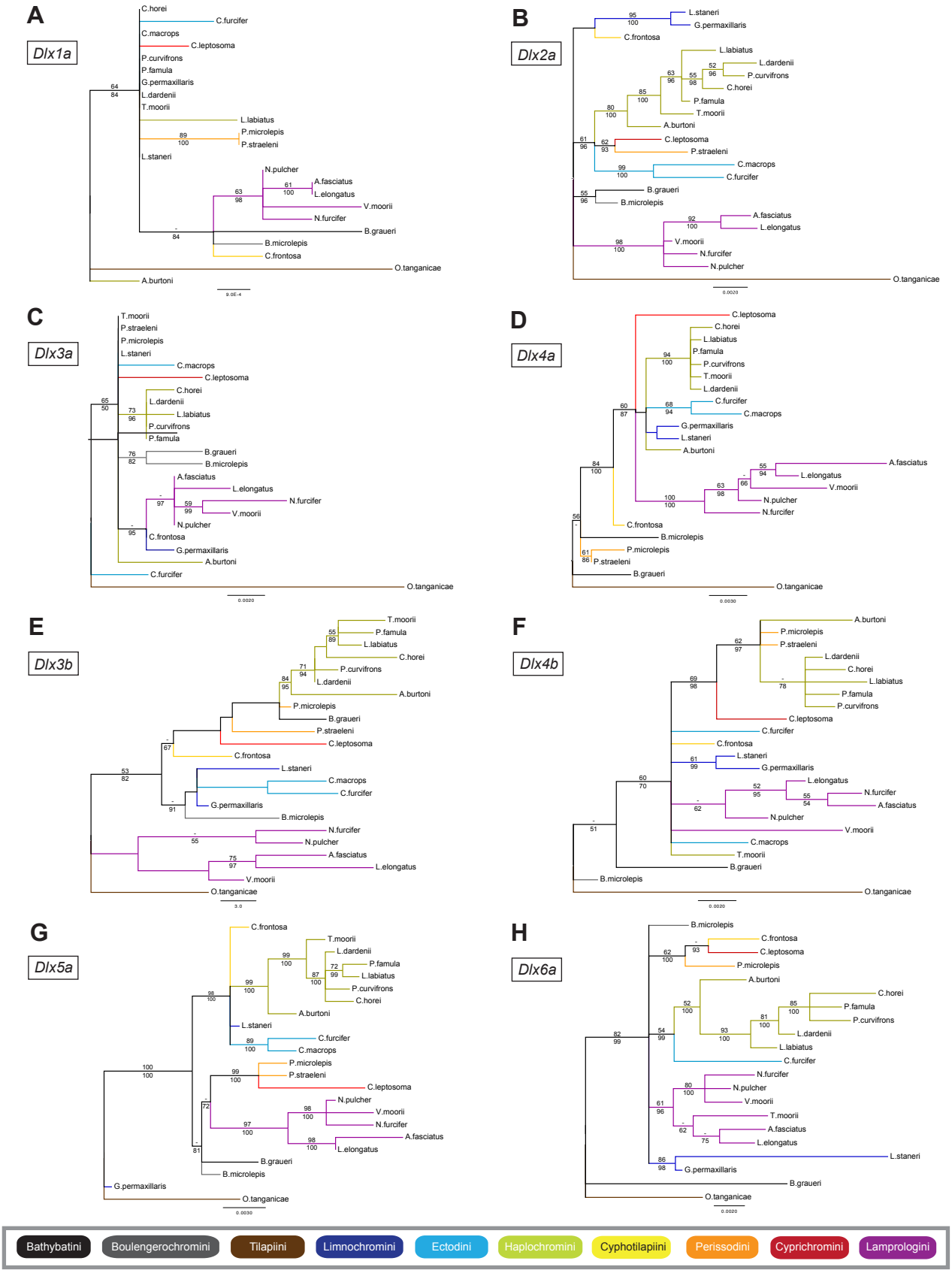
>dlx6a_Psecur

CTGGAGGCTCAGGACTCGTCCAAGTCTGCTTTCATGGAGTTTGGACAGCAGTCGCACTCACAGCAGAGCTCCCCATCGATGGGCAGCGGCCACTACCCGCTGCACTGTCTCCAC
TCCGGCTCACACTCTCACCACCAGCACGACAACACTCCGTACCCCGGGAGCAACACCTACAACAGGTCGTTACCTTACCCTTACGTGAGCCATCCACACCACAGCCCGTACCTG
CCGTCCTATCACAACAACATGGGAGGACAGACAAGGTTAGACGGCACAG----------
CGACAGTGATCGAAAACGGGGAGATTCGTTTTAACGGCAAAGGCAAGAAGATTCGCAAACCTCGGACAATTTATTCCAGTTTGCAGCTTCAAGCACTGAACCACCGTTTCCAG
CAAACACAGTACCTCGCTTTACCGGAGCGCGCCGAGCTGGCCGCCTCTCTAGGACTGACGCAAACACAGgTAAAAATTTGGTTTCAGAATAAAAGGTCAAAGTTCAAGAAGCT
GCTGAAGCAAGGCGGTAACCCGCACGAGAGCGAGCCTATCCCAGGCTCCATGTCCCTGTCCCCGCGCTCGCCGAGCATCCCCCCAATCTGGGACGTGTCGGCCTCATCCAAAG
GAGTAAACATGCCGGCCAACAGCTACATGCCTGGCTACTCTCACTGGTATTCCTCCCCACATCAAGACTCAATGCAGAGA



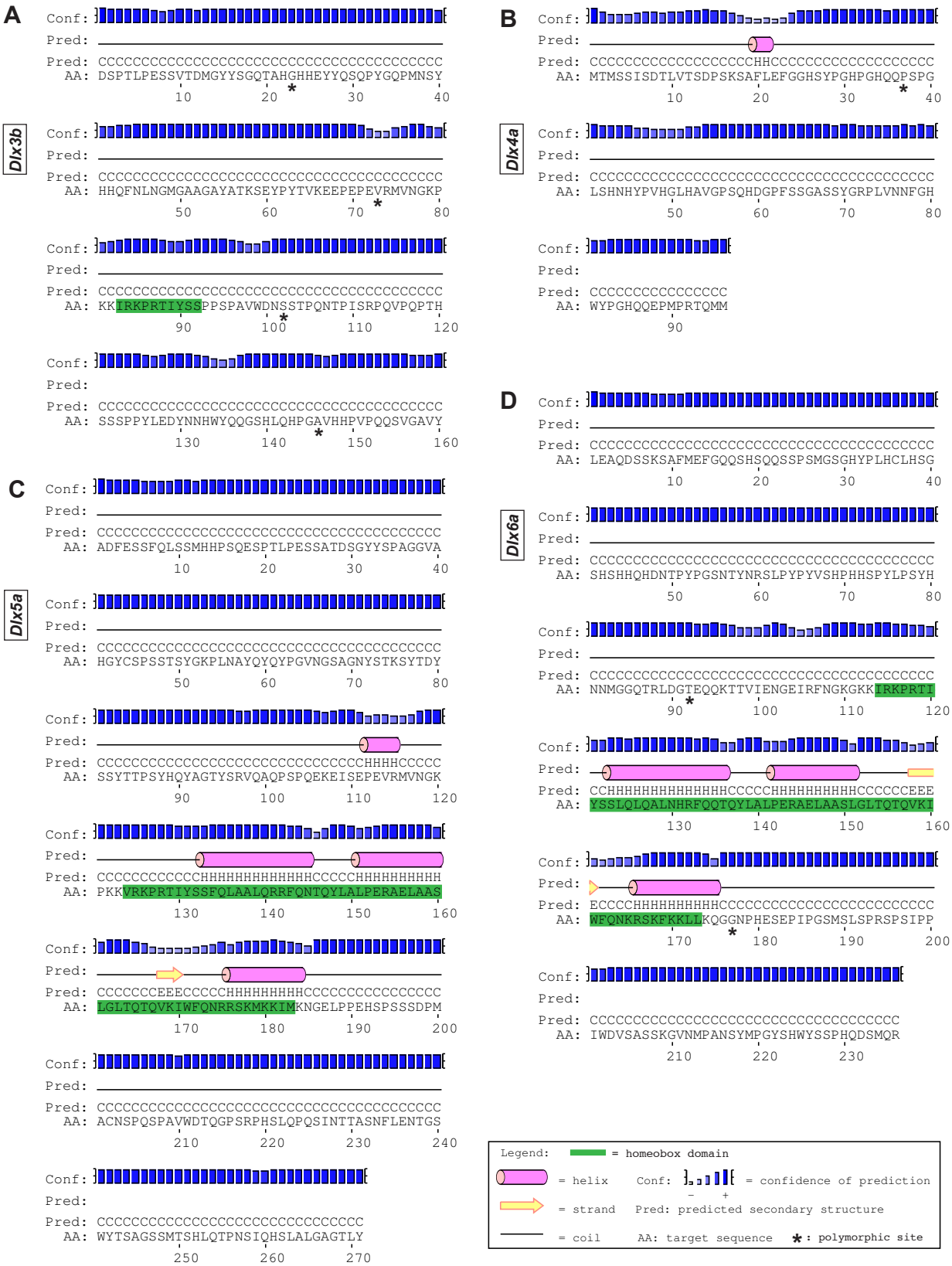
Additional file 5:

Protein comparison of the teleost dlx homeobox domains. Depicted are the amino acid sequences of the homeobox domains for each of the four teleost clusters: dlx1a-dlx2a, dlx4a-dlx3a, dlx4b-dlx3b and dlx6a-dlx5a in comparison with the single Dll homeobox sequence (here depicted in duplo) of *Drosophila melanogaster*. Sequences can be divided in two groups; dlx1a, dlx4a, dlx4b and dlx6a versus dlx2a, dlx3a, dlx3b and dlx5a. The two sixty amino acid long homeobox domains of each cluster are depicted in separate boxes. The top graph displays the mean pairwise identity of all sequences (i.e., green = 100% identity and brown ≥ 30% identity). Numbers represent the amino acid position within the homeobox.



Additional file 6:
Maximum likelihood gene trees based on 23 cichlid species for the eight *dlx* loci. Bootstrap values (PAUP*) and Bayesian posterior probabilities (MrBayes) above 50% are shown respectively above and below the branches. A color key for the ten studied cichlid lineages is given in the box below the figure. (**a**) *Dlx1a* (737 base pairs (bp); TPM3uf model). Two major polytomies were recovered. The lamprologines cluster together with the Boulengerochromini, Bathybatini and the Cyphotilapiini. *A. burtoni* is found at the base with *O. tanganicae*. (**b**) *Dlx2a* (1371 bp; HKY + I

model). Polytomous tree with all members of the lineages Lamprologines, Ectodines, Haplochromines and Limnochromines recovered as monophyletic clades. (**c**) *Dlx3a* (666 bp; HKY model). Polytomous tree, with only the Lamprologines recovered as monophyletic clade. (**d**) *Dlx4a* (1166 bp; TPM3uf + I + G). Polytomous relationships were observed between multiple lineages, although most lineages are monophyletic except the Haplochromines (**e**) *Dlx3b* (1972 bp; GTR + I + G). Moderately resolved tree. (**f**) *Dlx4b* (722 bp; TPM3uf). Mostly polytomous relationships between species, except the Limnochromini and most members of the Lamprologines. (**g**) *Dlx5a* (1538 bp; TIM2 + G). Basal polytomy divides ingroup species except *G. permaxillaris*, in two big clades. (**h**) *Dlx6a* (1710 bp; TIM3 + G). Limnochromines, Lamprologines and Haplochromines recovered as monophyletic clades, although the relationships between lineages are largely polytomous.



Additional file 7:

Four partially sequenced cichlid Dlx proteins. Depicted are the amino acid sequences of *Astatotilapia burtoni* (**a**, **c**, **d**) and *Ctenochromis horei* (**b**). Secondary structure predictions were obtained from the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/). (**a**) Dlx3b. (**b**) Dlx4a. (**c**) Dlx5a. (**d**) Dlx6a.

CHAPTER 4

IMMUNE-RELATED FUNCTIONS OF THE *HIVEP* GENE FAMILY IN EAST AFRICAN CICHLID FISHES

ARTICLE

Immune-Related Functions of the *Hivep* Gene Family in East African Cichlid Fishes

Eveline T. Diepeveen,*,[1,2] Olivia Roth,[†] and Walter Salzburger*
*Zoological Institute, University of Basel, 4051 Basel, Switzerland, and [†]Evolutionary Ecology of Marine Fishes, Helmholtz Centre of Ocean Research Kiel (GEOMAR), D-24105 Kiel, Germany

ABSTRACT Immune-related genes are often characterized by adaptive protein evolution. Selection on immune genes can be particularly strong when hosts encounter novel parasites, for instance, after the colonization of a new habitat or upon the exploitation of vacant ecological niches in an adaptive radiation. We examined a set of new candidate immune genes in East African cichlid fishes. More specifically, we studied the signatures of selection in five paralogs of the *human immunodeficiency virus type I enhancer-binding protein* (*Hivep*) gene family, tested their involvement in the immune defense, and related our results to explosive speciation and adaptive radiation events in cichlids. We found signatures of long-term positive selection in four *Hivep* paralogs and lineage-specific positive selection in *Hivep3b* in two radiating cichlid lineages. Exposure of the cichlid *Astatotilapia burtoni* to a vaccination with *Vibrio anguillarum* bacteria resulted in a positive correlation between immune response parameters and expression levels of three *Hivep* loci. This work provides the first evidence for a role of *Hivep* paralogs in teleost immune defense and links the signatures of positive selection to host–pathogen interactions within an adaptive radiation.

KEYWORDS
positive selection
immune response
adaptive radiation
molecular evolution

The interplay between hosts and their parasites (*i.e.*, macroparasites, bacteria, and viruses) represents one of the strongest biological interactions (Haldane 1949). Pathogens impose strong selection pressures on their hosts and have the potential to rapidly change the genotypic composition of host populations, which may ultimately alter the structure of entire ecosystems (Thompson 1988; Ebert and Hamilton 1996; Keesing *et al.* 2010). To counteract the permanently evolving pathogen virulence, hosts evolve resistance through diverse immune response mechanisms (Hamilton 1980). These include the discrimination between self and nonself, facilitating the recognition of pathogen-derived epitopes (Altizer *et al.* 2003; Boots *et al.* 2008). Invertebrates and vertebrates share the architecture of the innate immune system, *i.e.*, a conserved immediate defense mechanism including Toll-like receptors, lysozymes, and cellular defenses (Janeway *et al.* 2001). Immune memory, however, is the hallmark of the adaptive immune system of vertebrates (Cooper and Alder 2006; Flajnik and Kasahara 2010). The unique somatic diversification of receptors of the immunoglobulin family during ontogeny [*i.e.*, V(D)J recombination] mediates a dramatic increase in the number of foreign pathogen epitopes that the adaptive immune system can remember and attack (Janeway *et al.* 1996; Zhu *et al.* 2012).

Because of the constant interplay between host–parasite adaptation and counteradaption, immune genes are commonly characterized by signatures of positive selection through elevated rates of adaptive protein evolution (Hughes and Nei 1988; Hughes and Nei 1989; Jansa *et al.* 2003; Schlenke and Begun 2003; Sawyer *et al.* 2004; Nielsen *et al.* 2005; Jiggins and Kim 2007; Sackton *et al.* 2007; Tschirren *et al.* 2011). Selection on immune system diversification can be particularly strong when hosts encounter novel pathogens that induce primary immune challenges. This is the case after the colonization of a new habitat or upon the exploitation of vacant ecological niches (Scharsack *et al.* 2007; Matthews *et al.* 2010; Jones *et al.* 2012). For instance, it has been shown that migratory birds that encounter two or more different parasite faunas have larger immune defense organs (*e.g.*, bursa and spleen) than closely related resident birds (Møller and Erritzøe 1998). Freshwater sticklebacks differ in their immune competence potential depending on their ecotype (*e.g.*, lower parasite diversity in rivers than

doi: 10.1534/g3.113.008839
Manuscript received July 30, 2013; accepted for publication October 7, 2013; published Early Online October 18, 2013.

Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.113.008839/-/DC1
All generated cichlid *Hivep* sequences have been deposited into GenBank under accession nos. KF049218–KF049416.
[1]Present address: Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720.
[2]Corresponding author: Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, CA 94720-3144. E-mail: eveline.diepeveen@berkeley.edu

in lakes) (Scharsack *et al.* 2007), which is furthermore supported by a correlation between major histocompatibility complex (MHC) genotype and foraging habitat in benthic and limnetic ecomorphs (Matthews *et al.* 2010). Given the recognized evolutionary importance of the immune system (Janeway *et al.* 2001; Rodríguez *et al.* 2012) and the range of available functional and theoretical knowledge, the next step would be to assess to which degree immune genes contribute to, or even trigger, macroevolutionary events such as divergence, rapid speciation, and adaptive radiation.

East African cichlid fishes are a classic example of adaptive radiation (Schluter 2000). Because of their great numbers of closely related endemic species and their high levels of phenotypic and ecological diversity, cichlids are an important model system to study the genetic basis of diversification, adaption, and speciation (Kornfield and Smith 2000; Kocher 2004; Seehausen 2006; Salzburger 2009; Santos and Salzburger 2012). Previous studies of cichlid adaptive radiations have mainly focused on the understanding of ecologically important traits (and their genetic basis), such as the feeding apparatus (Terai *et al.* 2002b; Albertson *et al.* 2005; Fraser *et al.* 2008; Fraser *et al.* 2009; Muschick *et al.* 2011; Muschick *et al.* 2012), as well as on sexually selected traits such as coloration and pigmentation (Terai *et al.* 2002a; Terai *et al.* 2003; Salzburger *et al.* 2007; Roberts *et al.* 2009). Fewer studies have addressed the evolution of the immune system or, more generally, physiology in relation to diversification and rapid speciation in cichlids (Blais *et al.* 2007; Gerrard and Meyer 2007; Dijkstra *et al.* 2011). Dijkstra *et al.* (2011), for example, showed that divergence in coloration is accompanied by differentiation in immune function in Lake Victoria cichlids, and divergence in alleles of the MHC has previously been proposed as trigger of speciation in Lake Malawian cichlids through MHC-mediated mate choice (Blais *et al.* 2007). Several genes related to the immune system, including MHC loci, have been found to show signs of positive selection in East African cichlids (Gerrard and Meyer 2007), suggesting a role for immune genes during cichlid adaptive radiations.

In this study, we focused on the function and molecular characterization of a novel family of immune genes in (cichlid) fishes, which have been implicated to have immunological and developmental functions in mammals and insects (Seeler *et al.* 1994; Wu *et al.* 1996; Torres-Vazquez *et al.* 2001). In a previous study that focused on a candidate gene family for neural crest-derived structures in cichlids (*i.e.*, the endothelin family of ligands and receptors), we detected strong signatures of positive selection in a gene adjacent to one of the focal loci, the zinc finger protein *human immunodeficiency virus type I enhancer-binding protein 1* (*Hivep1*) (Diepeveen and Salzburger 2011). *Hivep1* is a transcription factor with functions in a variety of biological and developmental processes, *e.g.*, *HIV-1* gene expression (Maekawa *et al.* 1989; Muchardt *et al.* 1992; Seeler *et al.* 1994), in the *Decapentaplegic* signaling pathway important for cell fate specification during embryogenesis (Grieder *et al.* 1995; Dai *et al.* 2000; Marty *et al.* 2000; Torres-Vazquez *et al.* 2001), in V(D)J recombination of immunoglobulins (Wu *et al.* 1993; Wu *et al.* 1996), and in MHC enhancer binding (Baldwin *et al.* 1990; William *et al.* 1995). Although a single copy of this gene is found in *Drosophila*, mammals are typically characterized by three copies (Hicar *et al.* 2001; Dürr *et al.* 2004). Teleost fish, however, possess up to five duplicates (see Braasch *et al.* 2009), which is in accordance with the 3R hypothesis of a teleost-specific genome duplication event after the 2R duplications in the vertebrate linage (Sidow 1996; Taylor *et al.* 2003; Meyer and Van De Peer 2005; Volff 2005).

The goal of the current study was threefold. First, we characterized the signatures of selection (*i.e.*, d_N/d_S ratios) in the five *Hivep* paralogs in 40 East African cichlid fish species to determine whether adaptive protein evolution is commonly observed in the *Hivep* gene family. To this end, we performed phylogenetic analyses of the *Hivep* loci and estimated d_N/d_S ratios on both codon sites and in individual cichlid lineages. Second, we examined the role of the *Hivep* paralogs in the immune defense in the cichlid *Astatotilapia burtoni*. We evaluated the functional connection between *Hivep* expression levels and several cellular immune parameters after an experimental vaccination with *Vibrio anguillarum* bacteria. This fish pathogen was chosen to simulate a novel immune challenge, as the host was expected to be immunologically naïve against these *Vibrio* bacteria. Finally, we examined putative pleiotropic developmental functions through analyses of *cis*-regulatory regions to obtain insights into other functions of the *Hivep* paralogs in teleosts that could be linked to the observed signatures of adaptive protein evolution and related our findings to the explosive speciation events in East African cichlid fishes.

MATERIALS AND METHODS

Sampling, DNA and RNA extraction, and housing conditions

Samples for the DNA analyses were collected during two expeditions to Lake Tanganyika in 2007 and 2008 using a standard operating procedure described by Muschick *et al.* (2012). In total, 40 different cichlid species from 14 different lineages, including all major cichlid lineages in East Africa (so-called tribes) (Muschick *et al.* 2012) were examined (Supporting Information, Table S1). RNA for the gene expression assays was extracted from gill, brain, and liver tissue of adult *A. burtoni* (laboratory strain, both sexes; see *Experimental Vaccination* section). DNA and RNA extractions were performed as described elsewhere (Diepeveen and Salzburger 2011), with one exception: the tissue homogenization during the RNA extraction was performed on a BeadBeater (FastPrep-24; MP). Animals being part of the experimental vaccination study were kept under standard conditions (12 hr light, 12 hr dark, 25°) in the animal facilities at the Zoological Institute in Basel before transportation to the Helmholtz Centre for Ocean Research Kiel, where they were kept under the following conditions: 14 hr light; 10 hr dark; and 25° for ≥38 hr before the start of the experimental vaccination.

Loci, PCR amplification, and sequencing

Previously, five nuclear *Hivep* paralogs (*i.e.*, *Hivep1*, *Hivep2a*, *Hivep2b*, *Hivep3a*, and *Hivep3b*) were identified in teleost fishes (Braasch *et al.* 2009). Ensemble (versions 61 and 67) sequences from the following species were downloaded: zebrafish (*Danio rerio*); cod (*Gadus morhua*); medaka (*Oryzias latipes*); spotted green pufferfish (*Tetraodon nigroviridis*); fugu (*Takifugu rubripes*); tilapia (*Oreochromis niloticus*); and stickleback (*Gasterosteus aculeatus*) (Table S2). For two loci, we performed in-house tblastx searches on the server of the Zoological Institute (University of Basel) to determine Hivep protein sequences in the preliminary cichlid genomes of *A. burtoni*, *Neolamprologus brichardi*, and *Pundamilia nyererei* (BROAD Institute, unpublished data). These teleost and cichlid sequences were aligned with Codon Code Aligner 3.7.1 (CodonCode Corporation) to determine exon–intron structure and to design cichlid-specific primers (Table S3).

Subsequent PCR and sequencing reactions were performed as described elsewhere (Diepeveen and Salzburger 2011). PCR products were visualized with GelRed (Biotium) on a 1.5% agarose gel. Sequences were aligned and visually inspected using Codon Code Aligner 3.7.1 (CodonCode Corporation) and exon/intron boundaries were determined based on homology with the obtained other teleost sequences. Total sequenced regions (TSR), protein-coding regions, and concatenated

(TSRs of all five loci) data sets were constructed. All generated cichlid *Hivep* sequences have been deposited into GenBank (GenBank KF049218–KF049416) (Table S1).

Phylogenetic analyses and tests for selection pressure

Phylogenetic analyses and tests for selection pressure were performed as described elsewhere (Diepeveen and Salzburger 2011; Diepeveen *et al.* 2013). In short, models of nucleotide substitution were chosen based on likelihood ratio tests (LRTs) conducted in jModeltest 0.1.1 (Guindon and Gascuel 2003; Posada 2008) and used in maximum likelihood searches in PAUP* (Swofford 2002) and Bayesian Inference in MrBayes 3.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) for each individual paralog and for the concatenated dataset. Bootstrap analyses with 100 replicates were performed in PAUP* and MrBayes was run for 10,000,000 generations. *Tylochromis polylepis* and/or *Oreochromis tanganicae* were used as the outgroup in these analyses (Salzburger *et al.* 2002). The consensus tree based on the concatenated dataset was used as a common input tree in the subsequent analyses.

Both site and branch-site models, as implemented in Codeml, Phylogenetic Analysis by Maximum Likelihood (PAML) 4.2 (Yang 1997; Yang 2007), were used to test for selection pressure. The nonsynonymous/synonymous substitution rate ratio, ω or d_N/d_S, the proportion of sites assigned to an ω category, the $p_{0,1,2}$, and the p and q parameters of the β distribution were estimated for all five datasets under different models. LRTs of the following model comparisons were performed to detect sites under positive selection: M1a (nearly neutral) with M2a (positive selection); M7 (β) with M8 (β and $\omega_s \geq 1$); and M8a (β and $\omega_s = 1$) with M8. The comparison between M0 (one ratio) and M3 (discrete) was used as a test of variable ω among sites. Posterior probabilities for site classes were calculated with the Bayes empirical Bayes (BEB) (Yang *et al.* 2005). Next, LRTs between the null model ($\omega_s = 1$) and the alternative model ($\omega_s \geq 1$) were performed to determine if focal, or foreground, lineages evolved under non-neutral evolution. These foreground branches were chosen based on the results from the phylogenetic and PAML analyses.

Subsequent sliding window analyses (window size = 20) were performed with TreeSAAP (selection on amino acid properties) 3.2 (Woolley *et al.* 2003) for the four loci for which positively selected sites were observed with the PAML analyses (*Hivep1*, *Hivep2b*, *Hivep3a*, and *Hivep3b*). Amino acids were categorized based on 31 physicochemical properties to identify regions of positive selection. Selection on amino acids was subsequently screened for positive destabilizing selection by means of categorizing the substitutions into eight categories (categories 1–8) based on the magnitude of radicality (*i.e.*, 1 is the most conservative amino acid substitutions and 8 is the most radical). The three highest categories (6–8; $P \leq 0.001$) were used as indicative of radical amino acid substitutions. Next, these substitutions were analyzed with the program SIFT (sorting intolerant from tolerant) (Ng and Henikoff 2003) to screen for possible effect on protein function.

Analyzing *cis*-regulatory regions

The five *Hivep* sequences from *A. burtoni* were compared with the obtained teleost sequences of *O. niloticus*, *O. latipes*, *T rubripes*, *T. nigroviridis*, and *D. rerio* with mVISTA (Mayor *et al.* 2000; Frazer *et al.* 2004). Sequences were globally aligned with Shuffle-LAGAN (Brudno *et al.* 2003) and the minimum sequence similarity was set to 50%. Intragenic conserved noncoding elements were predicted and analyzed with rVISTA (Loots *et al.* 2002) to identify potential transcription factor binding sites.

Experimental vaccination and immune response measurements

To examine the expression patterns of the *Hivep* paralogs after an experimental vaccination, we exposed adult cichlid fish of the species *A. burtoni* to *Vibrio* bacteria following Roth *et al.* (2011). *V. anguillarum* was physically isolated from the stomach of freshly caught broad-nosed pipefish (*Syngnathus typhle*) (Roth *et al.* 2012). Strain confirmation was performed via sequencing of the *16S rRNA*, *recA*, and *pyrH* loci (GenBank reference numbers provided in Roth *et al.* 2012). On day 1 of the experiment, fish of both sexes were randomly assigned to either the experimental treatment (12 individuals) or the control treatment (11 individuals) and injected intraperitoneally with either 50 μl heat-killed (60 min at 65°) *V. anguillarum* (phylotype S6M4; 10^6 cells/ml dissolved in phosphate-buffered saline (PBS), *i.e.*, experimental treatment) or 50 μl PBS (*i.e.*, control treatment), respectively, according to the methods of Roth *et al.* (2012). Fish were tagged subcutaneously with visible implant elastomer tags (Northwest Marine Technology) according to treatment and kept in a single aquarium system. After ~21 hr of exposure, fish were killed with MS222 and weight and standard length were noted as in Birrer *et al.* (2012). Blood was collected from the caudal vessel in heparinized capillaries (Na-heparinized; Brand GMBH + Co. KG), followed by extraction of the head kidneys and spleen, which were forced through a 40-μm nylon sieve to prepare cell suspensions for subsequent cellular immune measurements. All steps were performed on ice. Cells were washed twice with RPMI medium (10 min, 600 rpm, 4°) and resuspended in a final volume of 450 μl.

The number of lymphocytes and monocytes (as proxies for immune response in the form of inflammation and/or stress to the treatment) were measured in blood, head kidneys, and spleen tissues by means of flow cytometry (FACSCalibur; Becton Dickinson) with pre-assessed cichlid-specific settings for each tissue type. The proportions of monocytes, lymphocytes, and the lymphocyte/monocyte ratio were calculated. Furthermore, the activity of the relative number of lymphocytes in the $G_{2\text{-}M}$ and synthesis (S) phases of the proliferation cycle compared to the relative number of lymphocytes in the $G_{0\text{-}1}$ phase was measured by killing cells in ethanol and subsequent labeling of the DNA with propidium iodide (Sigma Aldrich) as described by Roth *et al.* (2011). Lymphocytes were identified by their characteristic FSC/SSC pattern (*i.e.*, cell volume and inner complexity). Proliferating cells in the $G_{2\text{-}M}$ phase were distinguished from $G_{0\text{-}1}$ and S phase cells by a more intense red fluorescence because of their higher DNA content. To test whether the obtained data were normally distributed, D'Agnostino and Pearson omnibus normality tests as implemented in GraphPad Prism version 5.0a for Mac OS X (http://www.graphpad.com) were conducted. Outliers with values outside 2 SDs from the mean were removed (*i.e.*, up to three individuals per treatment group and tissue type).

The experiment was performed according to current national legislation of the Ministerium für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein (project entitled "Effects of global change on the immunological interaction of pipefish and cichlids with their natural bacteria communities"). One fish from the control treatment died during the experiment.

Gene expression assays and analyses

Gill, brain, and liver tissues of the 22 experimental animals were extracted and directly stored in RNA later (Invitrogen). RNA extraction and reverse-transcriptase were conducted as described elsewhere (Diepeveen and Salzburger 2011). Subsequent gene expression analyses were performed by means of quantitative PCR on a BioMark HD system at the Genetic Diversity Centre of the ETH Zurich, following the manufacturer's protocol. Levels of gene expression were measured in

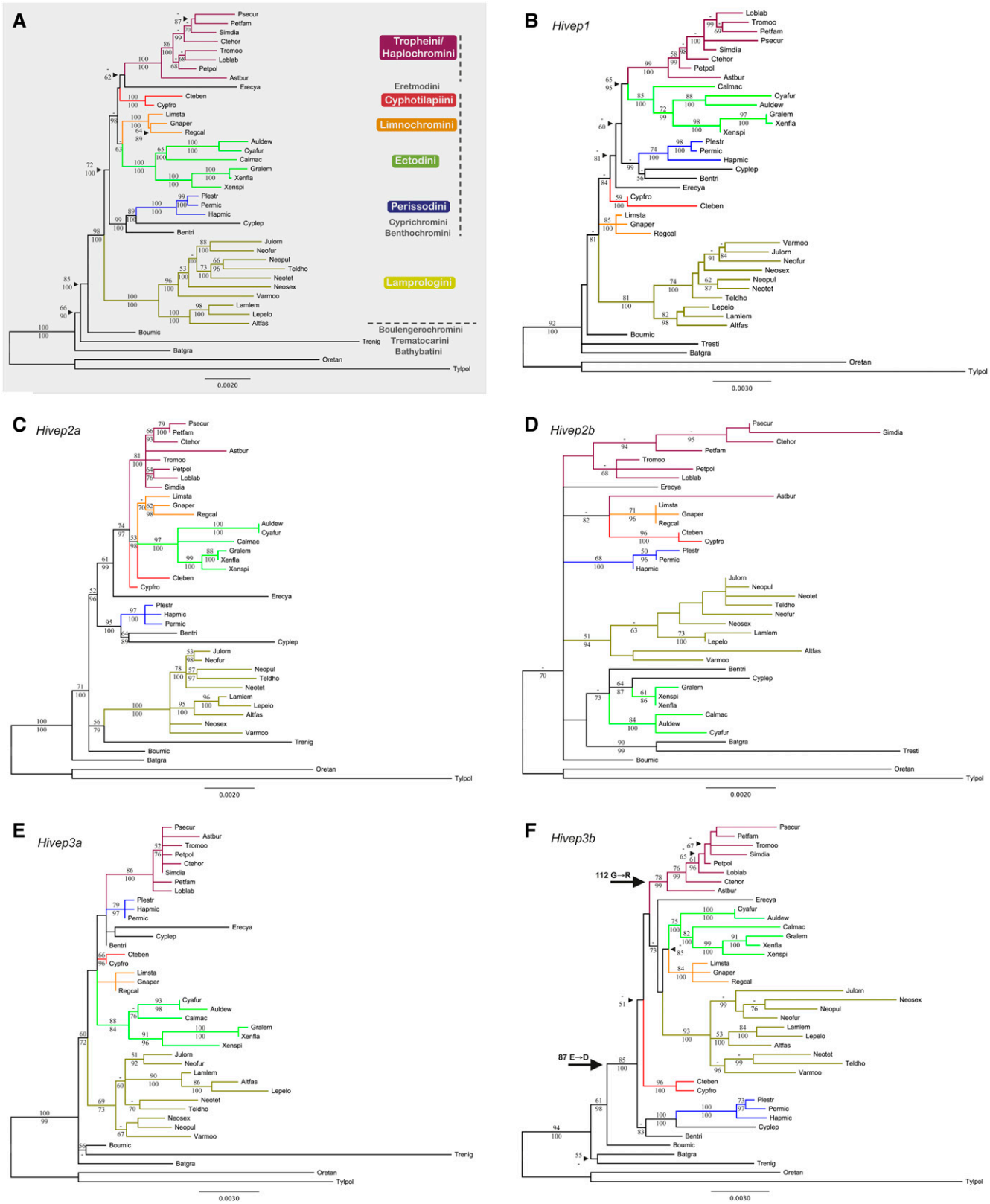


Figure 1 Phylogenetic hypotheses based on maximum likelihood for the concatenated dataset and the individual *Hivep* loci consisting of 40 taxa. (A) Concatenated dataset (13,543 base pairs (bp); best-fitting model of nucleotide substitution: HKY+I+G). Lineages are recovered with maximum support values, whereas relationships within and between lineages are supported with relative high values. The horizontal dotted line separates the five most basal species from the derived lineages: the lamprologines, the eretmodines, and the species belonging to the C-lineage, with the latter marked by the vertical dotted line. (B) *Hivep1* (3440 bp; TPM1uf+I+G) well-resolved with all major lineages recovered with high support

48.48 dynamic array integrated fluidic circuits with EvaGreen DNA binding dye. Primers were designed and tested for the five focal *Hivep* loci, two housekeeping loci [*i.e.*, elongation factor 1 (*EF1*) and ribosomal protein SA 3 (*RpSA3*)] (Colombo *et al.* 2013) and four control loci with demonstrated immune-related functions [*i.e.*, allograft inflammatory factor 1 (*AIF1*), anti-inflammatory response (Watano *et al.* 2001); coagulation factor II receptor-like 1 (*F2RL1*), inflammation and immunity (Rothmeier and Ruf 2012); interleukin 10 (*IL10*), immunosuppression (Jankovic and Trinchieri 2007); Toll-like receptor 5 (*TLR5*), pathogen recognition (Janeway and Medzhitov 2002; Akira *et al.* 2006)] (Table S3). Data were visualized, amplification plots were checked, and outliers were removed with the Fluidigm Real-Time PCR analysis software version 3.1 (Fluidigm). Further comparative analyses were performed with the qBase^{PLUS2} software package (Biogazelle). *EF1* and *RpSA3* were used as reference targets for the multiple reference gene normalization approach as implemented in qBase^{PLUS2} software (Biogazelle). Three different positive controls were included in this study; RNA was extracted from whole nonexperimental *A. burtoni* juveniles and two mixes composed of nine samples (gill, liver, and brain tissues of three randomly chosen individuals) each for the control group and the experimental group separately. Variation between PCR replicates and deviation of normalization factors were checked and outliers with values outside 2 SDs from the mean were removed. Data were controlled for inter-run variation.

Unpaired *t* tests were performed for the control immune genes expression levels between the control and experimental treatment groups with the qBase^{PLUS2} software (Biogazelle). To test for a correlation between the expression of *Hivep* paralogs [*i.e.*, quality-controlled and normalized relative quantities (CNRQ, here just RQ)] and the immune response measurements, Pearson correlations were calculated in GraphPad Prism in the control and experimental treatment groups.

RESULTS

Phylogenetic analyses of cichlid *Hivep* sequences

To examine the molecular evolutionary history of the *Hivep* paralogs, we performed maximum likelihood and Bayesian Inference phylogenetic analyses based on the total sequenced region per locus, and the concatenated dataset including all loci. The phylogenetic topologies of the obtained partial cichlid *Hivep* gene sequences and the concatenated dataset of 13.5 kb are displayed in Figure 1, A–F. Generally, the observed topologies of the gene trees, and the concatenated tree in particular, correspond with the available species trees (Salzburger *et al.* 2002; Takahashi 2003; Salzburger and Meyer 2004; Clabaut *et al.* 2005; Salzburger *et al.* 2005; Muschick *et al.* 2012), with *T. polylepis*, *O. tanganicae*, *B. graueri*, *B. microlepis*, and *T. nigrifrons* as most basal species, followed by the Lamprologini, the Eretmodini, and the species belonging to the "C-lineage" (Salzburger *et al.* 2002; Clabaut *et al.* 2005; Day *et al.* 2008) (Figure 1). As previously observed (Diepeveen and Salzburger 2011), *E. cyanostictus* was found at a different position within the C-lineage, whereas this species has been commonly resolved outside the C-lineage in previous studies (Salzburger *et al.* 2002; Clabaut *et al.* 2005; Day *et al.* 2008). Also, the relationships between the individual lineages of the C-lineage altered between the individual gene trees. Long branches were observed for *T. polylepis*, the lamprologines, individual lamprologine species, different branches within the ectodines in several gene trees, and for the haplochromines in *Hivep2b* and *T. nigrifrons* in the *Hivep3a* gene tree.

Selection pressure on sites and branches

To investigate signatures of selection pressure in the *Hivep* paralogs, we used site and branch-site models (Yang 1997, 2007) to obtain, *e.g.*, nonsynonymous/synonymous substitution rate ratios (d_N/d_S). The maximum likelihood parameter estimations for ω, $p_{0,1,2}$ and p, and q under different evolutionary models can be found in Table 1 for all five *Hivep* loci. Estimations of ω under the one ratio model (M0) suggest that the *Hivep* genes evolved under purifying selection, with ω ranging from 0.093 (*Hivep2a*) to 0.303 (*Hivep1*). A small proportion of sites, 1.4% (*Hivep2a*) to 11.8% (*Hivep3b*), was estimated to have evolved neutrally ($\omega = 1$) under the neutral model (M1a). By using models that allow ω to vary among sites (M2a, M3, and M8), up to 4.3% of sites were detected to have evolved with $\omega > 1$ in *Hivep1*, *Hivep2b*, *Hivep3a*, and *Hivep3b*, with more than 89.4% of sites evolving under purifying selection.

Likelihood ratio tests of several model comparisons (Table 2) were performed to detect positively selected amino acids. This approach resulted in the rejection of the null models in the M1a *vs.* M2a, M7 *vs.* M8, and M8a *vs.* M8 comparisons for all loci except *Hivep2a*. Positively selected sites were detected with the BEB for *Hivep1* (16 sites), *Hivep2b* (2 sites), *Hivep3a* (18 sites), and *Hivep3b* (13 sites).

LRTs of the branch-site analyses were performed to test whether focal lineages evolved under non-neutral evolution. Significant LRTs were only observed for *Hivep3b* (Table 3), indicating that although the ω ratios do vary among sites for three of the other four *Hivep* loci, they do not seem to vary significantly among lineages. For *Hivep3b*, the following two branches were observed with $\omega > 1$: the derived lineages (excluding the five basal species; $P < 0.001$) and the haplochromines ($P = 0.031$) (Table 3 and Figure 1).

Sliding windows and amino acid substitution characteristics

To visualize regions with elevated d_N/d_S values and to connect such regions with the physiochemical properties of the respective amino acid substitutions, we performed sliding window analyses. The sliding window plots of *Hivep1*, *Hivep2b*, *Hivep3a*, and *Hivep3b* are depicted in Figure 2, A–D. Regions of positive selection (z-score ≥ 3.09 corresponding with $P \leq 0.001$) were observed for all four loci, with highest

values. (C) *Hivep2a* (3143 bp; TIM2+G). The lamprologines plus the five most basal species are found basal of the C-lineage plus Eretmodini. All major lineages are monophyletic, except the Cyphotilapiini. (D) *Hivep2b* (1517 bp; TrN+I+G). Mostly unresolved tree with a basal polytomy, excluding the two outgroup species from all other species. Polytomous relationships were further found for the haplochromine and ectodine lineages. (E) *Hivep3a* (2142 bp; TPM1uf+G). The lamprologines plus the five most basal species are found basal of the C-lineage plus Eretmodini. (F) *Hivep3b* (3301 bp; HKY+I+G). The lamprologines are positioned within the C-lineage. Black arrows represent the two branches for which $\omega > 1$ was found in the branch-site analyses and their lineage-specific amino acid substitutions. Bootstrap values (PAUP*) and Bayesian posterior probabilities (MrBayes) >50% are shown, respectively, above and below the branches. Cichlid lineage names and a color key for the six cichlid lineages with more than one species included in this study are provided in the gray box in (A). Abbreviations of species names consist of the first three characters of the genus name followed by the first three characters of the species name (Table S1 shows full species names). Branch lengths of *T. polylepis* were shortened by 50% in all phylogenies and for *T. nigrifrons* in (E).

Table 1 Site model parameter estimates for the five *Hivep* paralogs

Locus	Parameter Estimates Using Different Models						
	M0 (One Ratio)	M1a (Neutral)[a]	M2a (Selection)[a]	M3 (Discrete)[a]	M7 (β)[a]	M8 (β and ω)[a]	M8a (β and $\omega_s = 1$)[a]
Hivep1	$\omega = 0.303$	$p_0 = 0.896$, $\omega_0 = 0.000$ $p_1 = 0.104$, $\omega_1 = 1.000$	$p_0 = 0.899$, $\omega_0 = 0.012$ $p_1 = 0.099$, $\omega_1 = 1.000$ **$p_2 = 0.002$, $\omega_2 = 21.464$**	$p_0 = 0.997$, $\omega_0 = 0.106$ **$p_1 = 0.003$, $\omega_1 = 18.416$** **$p_2 = 0.000$, $\omega_2 = 62.335$**	$p = 0.005$, $q = 0.046$	$p = 0.013$, $q = 0.091$ $p_0 = 0.998$ **$p_1 = 0.002$, $\omega = 21.496$**	$p = 0.005$, $q = 4.470$ $p_0 = 0.896$ $p_1 = 0.104$, $\omega = 1.000$
Hivep2a	$\omega = 0.093$	$p_0 = 0.986$, $\omega_0 = 0.080$ $p_1 = 0.014$, $\omega_1 = 1.000$	$p_0 = 0.986$, $\omega_0 = 0.080$ $p_1 = 0.001$, $\omega_1 = 1.000$ $p_2 = 0.012$, $\omega_2 = 1.000$	$p_0 = 0.543$, $\omega_0 = 0.000$ $p_1 = 0.270$, $\omega_1 = 0.203$ $p_2 = 0.186$, $\omega_2 = 0.203$	$p = 0.554$, $q = 5.217$	$p = 0.554$, $q = 5.221$ $p_0 = 1.000$ $p_1 = 0.000$, $\omega = 1.00$	$p = 0.672$, $q = 6.412$ $p_0 = 0.999$ $p_1 = 0.001$, $\omega = 1.000$
Hivep2b	$\omega = 0.120$	$p_0 = 0.942$, $\omega_0 = 0.000$ $p_1 = 0.058$, $\omega_1 = 1.000$	$p_0 = 0.941$, $\omega_0 = 0.000$ $p_1 = 0.059$, $\omega_1 = 1.000$ **$p_2 = 0.000$, $\omega_2 = 38.425$**	$p_0 = 0.957$, $\omega_0 = 0.000$ **$p_1 = 0.043$, $\omega_1 = 1.439$** **$p_2 = 0.000$, $\omega_2 = 38.861$**	$p = 0.005$, $q = 0.049$	$p = 0.010$, $q = 0.160$ $p_0 = 1.000$ **$p_1 = 0.000$, $\omega = 38.446$**	$p = 0.005$, $q = 2.777$ $p_0 = 0.942$ $p_1 = 0.058$, $\omega = 1.000$
Hivep3a	$\omega = 0.251$	$p_0 = 0.919$, $\omega_0 = 0.000$ $p_1 = 0.081$, $\omega_1 = 1.000$	$p_0 = 0.909$, $\omega_0 = 0.000$ $p_1 = 0.090$, $\omega_1 = 1.000$ **$p_2 = 0.001$, $\omega_2 = 31.219$**	$p_0 = 0.984$, $\omega_0 = 0.040$ **$p_1 = 0.016$, $\omega_1 = 4.264$** **$p_2 = 0.000$, $\omega_2 = 47.201$**	$p = 0.005$, $q = 0.046$	$p = 0.005$, $q = 0.049$ $p_0 = 0.999$ **$p_1 = 0.001$, $\omega = 31.868$**	$p = 0.005$, $q = 0.047$ $p_0 = 1.000$ $p_1 = 0.000$, $\omega = 1.000$
Hivep3b	$\omega = 0.277$	$p_0 = 0.882$, $\omega_0 = 0.000$ $p_1 = 0.118$, $\omega_1 = 1.000$	$p_0 = 0.894$, $\omega_0 = 0.014$ $p_1 = 0.102$, $\omega_1 = 1.000$ **$p_2 = 0.004$, $\omega_2 = 13.855$**	$p_0 = 0.000$, $\omega_0 = 0.000$ $p_1 = 0.992$, $\omega_1 = 0.090$ **$p_2 = 0.008$, $\omega_2 = 10.193$**	$p = 0.005$, $q = 0.046$	$p = 0.023$, $q = 0.170$ $p_0 = 0.996$ **$p_1 = 0.004$, $\omega = 13.814$**	$p = 0.005$, $q = 2.855$ $p_0 = 0.882$ $p_1 = 0.118$, $\omega = 1.000$

The maximum likelihood parameter estimations for ω, $p_{0,1,2}$ and p and q under different evolutionary models for all five *Hivep* loci. Estimations of ω under the one ratio model (M0) suggest that the *Hivep* genes evolved under purifying selection. A small proportion of sites was estimated to have evolved neutrally ($\omega = 1$) under the M1a model in all loci. By using the M2a, M3, and M8 models, a small proportion of sites were detected to have evolved under $\omega > 1$ in *Hivep1*, *Hivep2b*, *Hivep3a*, and *Hivep3b*. *Hivep*, *human immunodeficiency virus type I enhancer-binding protein*.

[a] $p_{0\text{-}2}$ are the proportions of sites assigned to the ω category or to a beta distribution with p and q as parameters; ω ratios >1 and their corresponding proportions are depicted in bold.

z-scores for *Hivep2b*, and the most numerous regions with a z-score ≥ 3.09 observed for *Hivep3a*. Interestingly, not all of these retrieved regions of positive selection correspond with the obtained positively selected sites as identified by the PAML analyses and *vice versa*. Relative few regions of positive selection are observed in the ZAS domains that contain the zinc fingers that bind specific DNA motifs. Notable exceptions are the ZAS-N domain of *Hivep2b* and the ZAS-C domain of *Hivep3b* (Figure 2); the latter is furthermore characterized by a positively selected site identified by the PAML analyses. Most commonly observed positively selected amino acid properties among the four paralogs affect the alpha-helical tendencies, the compressibility, the equilibrium constant (ionization of COOH), and the surrounding hydrophobicity. The SIFT analyses of the observed substitutions to screen for possible effect on protein function showed that all substitutions are tolerant and thus have no predicted damaging effect on function (data not shown).

Analyzing *cis*-regulatory regions

We investigated noncoding regions within the *Hivep* paralogs for potential *cis*-regulatory elements to determine possible binding sites for transcription factors, indicative of putative functional involvement in signaling pathways. Vistaplots of *Hivep1*, *Hivep2a*, *Hivep2b*, and *Hivep3b* are depicted in Figure S1, A–D. Because of a limited number of retrieved teleost sequences for *Hivep3a*, the Vistaplot was not informative for this locus and was therefore excluded from further analyses. For all four analyzed loci, conserved noncoding elements (CNEs) were observed in *A. burtoni*. Interestingly, a similar pattern of two CNEs surrounding a single exon was observed in all loci (arrows in Figure S1). Although this pattern seems common among teleost fish for *Hivep1* and *Hivep2a*, for *Hivep2b* and *Hivep3b* this pattern seems to be restricted to cichlid fishes (*O. niloticus* is the reference sequence in these analyses). A third cichlid-specific CNE was observed in a subsequent intron in both *Hivep1* and *Hivep2a*, whereas for *Hivep3b* two more cichlid-specific CNEs were identified.

Because the particular pattern of two CNEs surrounding an exon was observed in all four analyzed loci, the subsequent search for potential transcription factor binding sites was mainly focused on these regions to determine any overlap in possible function of these regions. The analyses resulted in similar hits among *Hivep* paralogs and suggested a possible association between the *Hivep* paralogs and different types of signaling pathways involved in, *e.g.*, sex determination [*androgen receptor* (*AR*); pre-B-cell leukemia transcription factor 1 (*PBX1*); sex-determining region Y protein (*SRY*)], immune system [*B-cell lymphoma 6 protein* (*BCL6*); H2.0-like homeobox protein (*HB24*); signal transducer and activator of transcription1,3,5a (*STAT1,3,5a*)], developmental patterning [homeobox protein Hox-A3 (*Hoxa3*); homeobox protein MSX-1 (*Msx1*)], and several members of the paired box protein Pax (*PAX*) and bone morphogenetic protein (*BMP*) pathways.

Experimental vaccination and immune response measurements

We performed experimental vaccinations to test whether the *Hivep* paralogs are involved in an induced immune response. The experimental vaccination was realized by exposure to heat-killed *V. anguillarum* for ~21 hr, following the methods of Birrer *et al.* (2012). Several immune response measurements were performed to determine induction of immune defense dynamics. The lymphocyte/monocyte ratio and the relative number of lymphocytes in the $G_{2\text{-}M}$ and S phases of the proliferation cycle were measured in blood, spleen, and head kidney (Figure 3). Data were normally distributed. The experimental treatment resulted in an elevated lymphocyte/monocyte ratio in blood

Table 2 LRT statistics of three site model comparisons and positively selected sites

Locus	Test	LRT ($2\Delta l$)	*P*	Selected Sites (BEB)*
Hivep1	M1a *vs.* M2a	210.047	<**0.001**	**37S**, **49Q**, **61V**, **81N**, *106T*, *114R*, *130T*, **248A**, **292Q**, **472Q**, *530V*, *546L*, **558T**, **582P**, *587H*, **656N**
	M7 *vs.* M8	210.084	<**0.001**	**37S**, **49Q**, **61V**, **81N**, *106T*, *114R*, **130T**, **248A**, **292Q**, **472Q**, **530V**, **546L**, **558T**, **582P**, *587H*, **656N**
	M8a *vs.* M8	210.024	<**0.001**	See M7 *vs.* M8 comparison
Hivep2a	M1a *vs.* M2a	0.000	1.000	0
	M7 *vs.* M8	0.000	1.000	0
	M8a *vs.* M8	0.007	0.932	0
Hivep2b	M1a *vs.* M2a	35.537	<**0.001**	**143P**
	M7 *vs.* M8	37.555	<**0.001**	**143P**, *330Q*
	M8a *vs.* M8	34.250	<**0.001**	See M7 *vs.* M8 comparison
Hivep3a	M1a *vs.* M2a	304.843	<**0.001**	**62S**, **63A**, **64A**, *82S*, **141A**, **200S**, **265I**, *329Q*, **343P**, *371D*, **436V**, **445A**, **484T**, **565E**, *573T*, **703P**
	M7 *vs.* M8	307.001	<**0.001**	**62S**, **63A**, **64A**, *82S*, **141A**, *168N*, **200S**, **265I**, *329Q*, **343P**, *371D*, **436V**, *437K*, **445A**, **484T**, **565E**, *573T*, **703P**
	M8a *vs.* M8	307.003	<**0.001**	See M7 *vs.* M8 comparison
Hivep3b	M1a *vs.* M2a	111.745	<**0.001**	**87E**, **112G**, *218H*, **286T**, **326S**, *335I*, **352G**, *392G*, **399P**, *401P*, **403R**, **447I**, **511K**
	M7 *vs.* M8	113.003	<**0.001**	**87E**, **112G**, *218H*, **286T**, **326S**, **335I**, **352G**, *392G*, **399P**, **401P**, **403R**, **447I**, **511K**
	M8a *vs.* M8	111.749	<**0.001**	See M7 *vs.* M8 comparison

LRTs resulted in the rejection of the null models in the M1a *vs.* M2a, M7 *vs.* M8, and M8a *vs.* M8 comparisons for all loci except *Hivep2a*. Positively selected sites were detected for *Hivep1*, *Hivep2b*, *Hivep3a*, and *Hivep3b*. LRT, likelihood ratio test; BEB, Bayes empirical Bayes; *Hivep*, *human immunodeficiency virus type I enhancer-binding protein*.

[a] *P* = 0.01 (bold) and *P* = 0.05 (italic).

(P = 0.008; unpaired t test) and spleen (P = 0.018), indicative of a higher proportion of cells from the adaptive immune system (*i.e.*, immune response). A higher proportion of lymphocytes in the S and $G_{2\text{-}M}$ phases was found in the head kidney of the experimental group (P = 0.005), indicative of induced lymphocyte proliferation.

Gene expression assays

We measured the expression levels of four control immune loci, *AIF1*, *F2RL1*, *IL10*, and *TLR5* in liver and gill tissues. These relative expression levels are depicted in Figure 4. For *AIF1* and *TLR5*, we found significantly higher levels of relative expression in liver (P = 0.014 and $P < 0.001$; unpaired t test) and gills ($P < 0.001$ and P = 0.006) in the experimental treatment group.

To analyze the effect of *Vibrio* exposure on the expression levels of the *Hivep* paralogs in detail, we assessed their expression levels in relation to an immune response parameter (*i.e.*, lymphocyte/monocyte ratio) per treatment group (*i.e.*, control and experimental). Four correlations were significant between the lymphocyte/monocyte ratio of the spleen and the relative expression of *Hivep1* (liver; Pearson r = 0.798, P = 0.018), *Hivep1* (gills; Pearson r = 0.794, P = 0.011), *Hivep2a* (Pearson r = 0.745, P = 0.021), and *Hivep3b* (Pearson r = 0.852, P = 0.007) (Figure 5). In these cases, the expression level of the *Hivep* paralogs thus correlates positively with the level of the immune response parameter.

DISCUSSION

In this study, we examined the molecular evolutionary history of the *Hivep* gene family members in relation to their presumed immune-related function in a renowned model system for evolutionary biology, the East African cichlid fishes. We performed comparative phylogenetic

Table 3 Parameter estimations and LRTs for the null and alternative hypotheses of the branch-site model for two different cichlid lineages for *Hivep3b*

Clade	Model	Site Class	0	1	2a	2b	LRT (*P*)
DL	Model A (Null)	Proportion	0.766	0.123	0.096	0.015	
		Background ω	0.000	1.000	0.000	1.000	
		Foreground ω	0.000	1.000	1.000	1.000	
	Model A (Alternative)	Proportion	0.897	0.087	0.014	0.001	18.509 (<0.001)
		Background ω	0.070	1.000	0.070	1.000	
		Foreground ω	0.070	1.000	7.974	7.974	
HC	Model A (Null)	Proportion	0.725	0.167	0.088	0.020	
		Background ω	0.000	1.000	0.000	1.000	
		Foreground ω	0.000	1.000	1.000	1.000	
	Model A (Alternative)	Proportion	0.806	0.185	0.008	0.002	4.628 (0.031)
		Background ω	0.000	1.000	0.000	1.000	
		Foreground ω	0.000	1.000	24.33	24.33	

LRTs of the branch-site analyses indicate that *Hivep3b* evolved under non-neutral evolution ($\omega > 1$) in the following two focal lineages: the derived lineages (excluding the five basal species) and the haplochromines. LRT, likelihood ratio test; *Hivep*, *human immunodeficiency virus type I enhancer-binding protein*; DL, derived lineage; HC, Haplochromini.

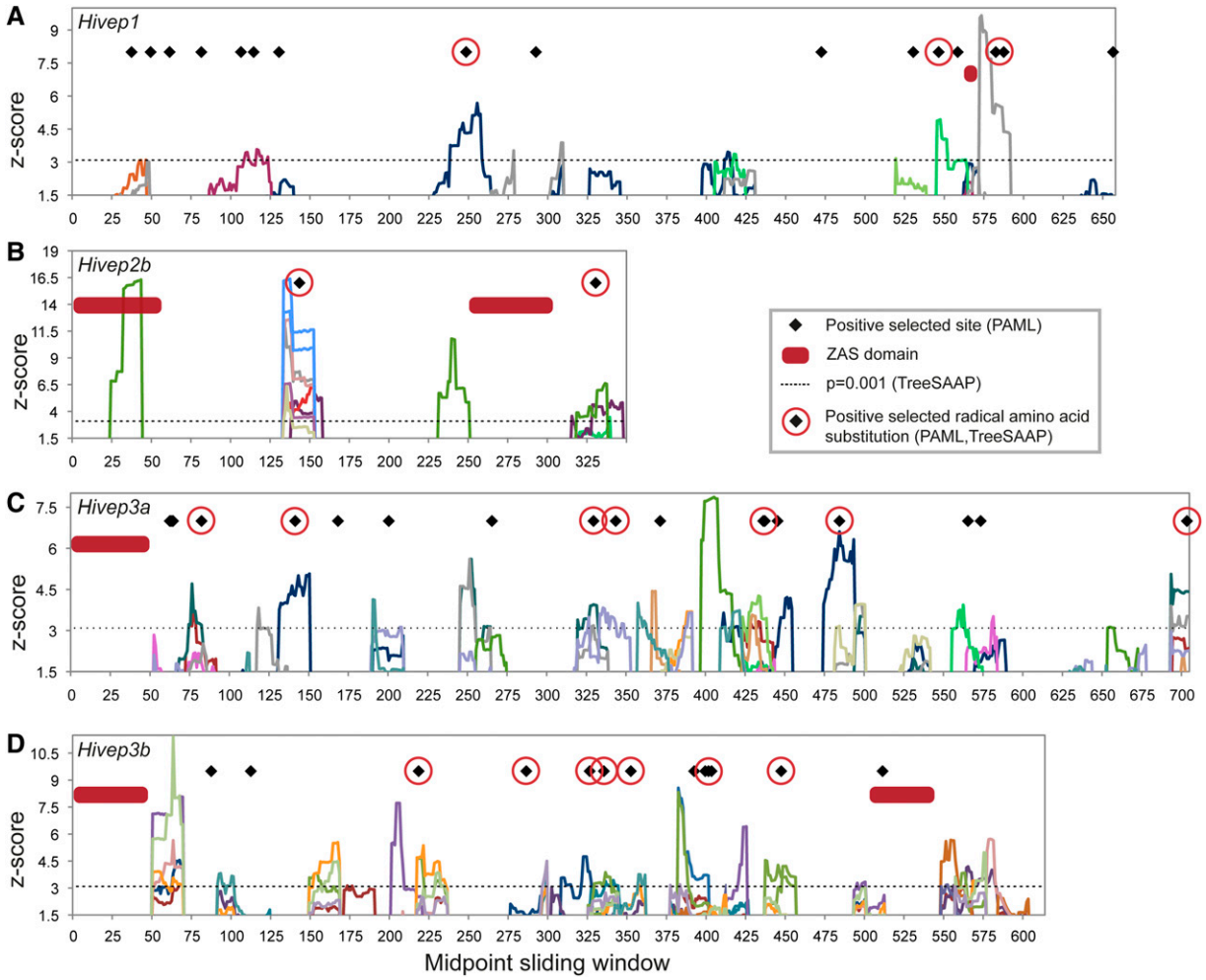


Figure 2 Sliding window plots and radical amino acid properties for four *Hivep* paralogs. (A) For *Hivep1*, multiple sliding windows were observed with z-score ≥ 3.09. For several positively selected sites identified with PAML, no radical substitutions were found and *vice versa*. (B) In *Hivep2b*, four regions of positive selection are observed of which two correspond with positively selected sites identified by the PAML analyses. (C) The observed regions of positive selection for *Hivep3a* are characterized by many different physicochemical properties. (D) Many radical amino acid substitutions were found for *Hivep3b* that correspond with most of the positively selected sites identified by PAML analyses. Each physicochemical amino acid property is individually color coded (see Figure S2 for details). Black diamonds (♦) represent positively selected amino acid sites obtained by the PAML analyses and red circles around them represent positively selected radical nonsynonymous substitutions (category 6-8). The dotted line at z-score = 3.09 represents $P = 0.001$, whereas z-score = 1.64 represents $P = 0.05$. Red rectangles represent the following ZAS domains: Hivep1 ZAS-C; Hivep2b ZAS-N, ZAS-C; Hivep3a ZAS-N; and Hivep3b ZAS-N, ZAS-C.

analyses and detailed screens of d_N/d_S ratios, analyzed putative *cis*-regulatory regions within the loci, and, in particular, investigated the expression levels of the *Hivep* paralogs after an experimental vaccination with *V. anguillarum*. We show, for the first time to our knowledge, that the *Hivep* paralogs play a putative role in the response to vaccination in fish, and that they are characterized by signatures of long-term positive selection. Our findings regarding the *Hivep* paralogs indicate that they are important candidate genes for immune-related functions in teleost fish and suggest broader implications in relation to speciation events, such as the adaptive radiations in East African cichlid fishes.

Exposure to *V. anguillarum* causes an immune response in *A. burtoni*

To test whether the exposure to a vaccination with heat-killed *V. anguillarum* resulted in an upregulation of the cellular fish immune response, we measured lymphocyte/monocyte ratios, the proportions of proliferating lymphocytes, and the expression levels of four control immune genes with demonstrated functions in the inflammatory response and immunity (Watano *et al.* 2001; Janeway and Medzhitov 2002; Akira *et al.* 2006; Jankovic and Trinchieri 2007; Rothmeier and Ruf 2012) (see *Materials and Methods* section). Consistent with an elevated immune response upon *Vibrio* vaccination, we found an increased lymphocyte production in the head kidney, the organ where clonal lymphocyte production takes place (Rombout *et al.* 2005; Abdel-Aziz *et al.* 2010). We also found a higher proportion of lymphocytes both in blood and spleen, indicating lymphocyte migration toward peripheral organs. Although lymphocytes are transported via blood, the spleen is the major lymphoid tissue associated with the clearance of blood-borne antigens (Press and Evensen 1999; Whyte 2007). Finally, the significant upregulation of both *AIF1* and *TLR5* in the *Vibrio*-exposed group indicates activation of the immune system (Watano *et al.* 2001; Janeway and Medzhitov 2002; Akira *et al.* 2006). However, we did not find a significant upregulation for two other immune genes with demonstrated functions in the immune response,

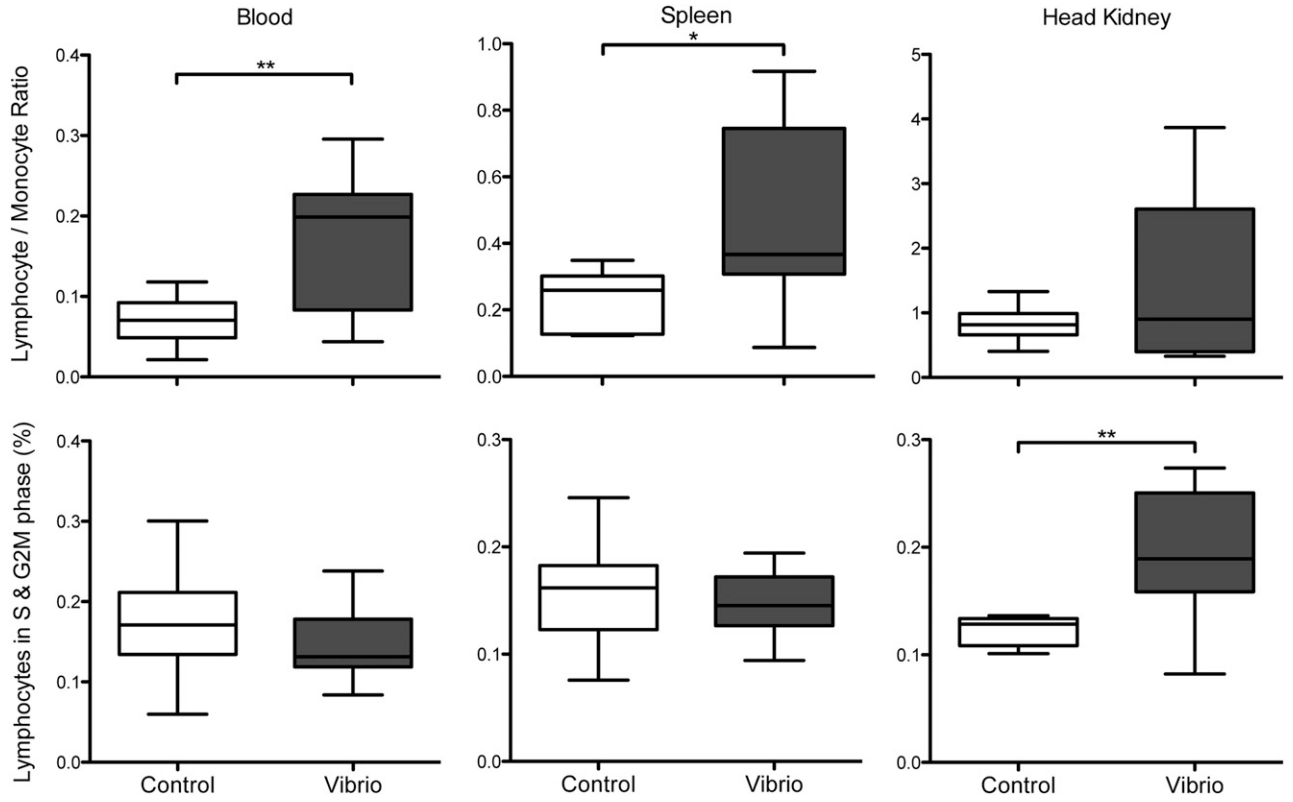


Figure 3 Immune response measurements after the experimental vaccination in *A. burtoni* adults. Lymphocyte/monocyte ratios (top) and proportions of cells in the S and $G_{2\text{-}M}$ phases (bottom) measured in blood (left), spleen (center), and head kidney (right) for the control treatment (white boxplot) and experimental treatment (*Vibrio*; gray boxplot). $^{*}P < 0.05$; $^{**}P < 0.01$. Depicted are the median, lower and upper quartiles (box), and the minimum and maximum observed values (error bars).

F2RL1 and *IL10* (Déry *et al.* 1998; Savan *et al.* 2003; Rothmeier and Ruf 2012). These were possibly missed by the choice of our measuring time point (21 hr after vaccination), as suggested by Savan *et al.* (2003), who found elevated *IL10* expression in carp liver tissue after LPS stimulation only within the first 6 hr of incubation.

Hivep expression levels correlate with cellular immune response parameters in an East African cichlid fish

Although several functions of the *Hivep* paralogs have been demonstrated in the fruitfly, *Xenopus* frog, human, and mouse (Wu 2002), the *Hivep* paralogs have, so far, not been examined in teleost fishes. We tested, for the first time, whether there is a correlation between immune response parameters and the expression level of the *Hivep* paralogs in fish as an indicator of putative function(s) in the immune response.

Although our study does not determine the exact function of the *Hivep* paralogs within the immune response, the positive correlations between the lymphocyte/monocyte ratio and the expression levels of three *Hivep* paralogs indicate that the expression of —at least— *Hivep1*, *Hivep2a*, and *Hivep3b* is upregulated upon the experimental vaccination. This implies that *Hivep* paralogs play a role during the immune response of fish. These results provide, to our knowledge, the first indication of an immunological function of the *Hivep* paralogs in teleost fish, which is congruent with preliminary findings in pipefish (O. Roth, personal communication). The *Hivep* gene family thus offers a potential novel family of immune genes for teleost fish that, when their functions are characterized in more detail, can be used in future immunological screens.

Other functional implications

The experimental vaccination did not lead to upregulated expression levels of all five *Hivep* paralogs. We found no correlation between the expression levels of *Hivep2b* and *Hivep3a* and the immune response measurements. These paralogs either have immunological functions beyond the scope of our experimental design or are not involved in the immune response in teleost fishes. Previously, it had been shown that *Hivep* genes are involved in functions other than the immune system in insects and vertebrates, *e.g.*, in murine osteoclastogenesis (Liu *et al.* 2011), in *BMP*/*Dpp* signaling (Grieder *et al.* 1995; Dai *et al.* 2000; Marty *et al.* 2000; Torres-Vazquez *et al.* 2001; Yao 2006; Saita *et al.* 2007; Yin *et al.* 2010), and in the development of the nervous system (Campbell and Levitt 2003; Takagi *et al.* 2006). Interestingly, several of the potential transcription factor binding sites identified within the observed CNEs correspond with these known functions of *Hivep* paralogs. For instance, we found multiple hits for components of the *BMP* signaling pathway, as well as several other developmental patterning loci, suggesting a putative role of the *Hivep* paralogs in developmental patterning and bone formation in cichlid fishes. *Hivep* paralogs have been found to play a role in the specification of *Drosophila* wing and halter discs (Torres-Vazquez *et al.* 2001), multiple *dpp*-dependent patterning events of both *Drosophila* ectoderm and mesoderm (Arora *et al.* 1995), and male tail patterning in *Caenorhabditis elegans* (Liang *et al.* 2007). The roles of *Hivep* in the *BMP* pathway, possibly through alternative splicing (Hicar *et al.* 2001; Hong and Wu 2005; Yin *et al.* 2010), together with several indications of functions in appendage specification and patterning make them candidate genes for fin patterning and anal fin egg-spot

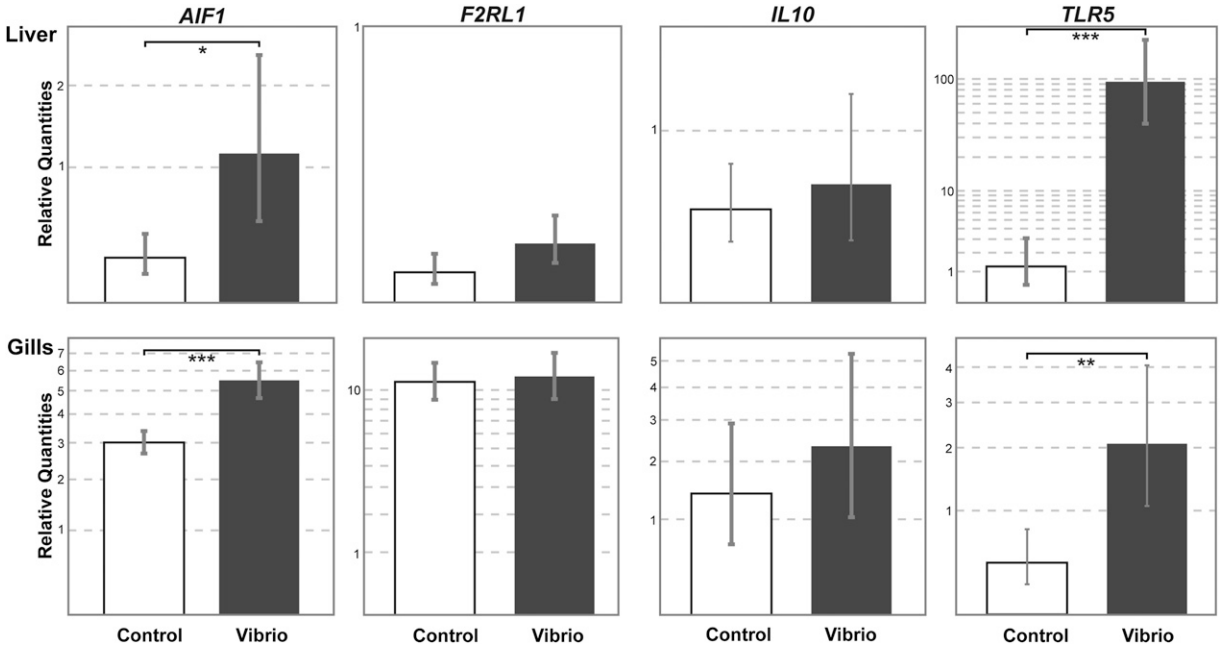


Figure 4 Gene expression assays for the control immune genes in *A. burtoni* adults. The relative gene expression levels (relative quantities) of *AIF1*, *F2RL1*, *IL10*, and *TLR5* measured in liver (top) and gills (bottom) for the control treatment (white bars) and experimental treatment (*Vibrio*, blue bars). $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. Depicted are the mean and the 95% CI (error bars).

formation, a sexually selected trait involved in courtship and spawning behavior and intrasexual communication of haplochromine cichlid species (Wickler 1962; Hert 1989; Salzburger *et al.* 2005; Salzburger 2009; Theis *et al.* 2012). Future detailed expression and functional analyses should elucidate whether the *Hivep* paralogs are involved the development of egg spots in haplochromine cichlid fishes.

Implications of positive selection on *Hivep* paralogs, immune genes, and speciation events

Positive selection, or adaptive sequence evolution, is the hallmark of evolutionary change and molecular adaptation. By comparing the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) of protein coding genes, the selection regime (*i.e.*, neutral, purifying or positive) per amino acid can be inferred (Yang and Bielawski 2000). This method is widely used and has led to the identification of many cases of positive selection (Yang and Bielawski 2000). Genes involved in (evading) defensive systems or immunity are typically found with signatures of positive selection (Endo *et al.* 1996; Yang and Bielawski 2000; Schlenke and Begun 2003; Nielsen 2005; Nielsen *et al.* 2005; Biswas and Akey 2006; Yang 2006; Jiggins and Kim 2007; Montoya-Burgos 2011). As discussed, several functions within the immune response have been described for the *Hivep* paralogs in other species, and our detailed inferences of the d_N/d_S ratios provide evidence for positive selection acting on four out of five *Hivep* paralogs. Interestingly, signs of positive selection have been found before in vertebrate *Hivep* paralogs. *Hivep2* has been found with a signature of positive selection in *Tetraodon* (Montoya-Burgos 2011) and the cow lineage (Toll-Riera *et al.* 2011), whereas *Hivep3* showed signs of positive selection in the human lineage (Vamathevan *et al.* 2008). At least for the human *Hivep* paralog, it has been suggested that the immune function is the cause for the signature of positive selection. Together, these results indicate that it is likely that the immune-related functions of the *Hivep* paralogs are the cause for the elevated d_N/d_S ratios observed across vertebrate lineages, including the 14 cichlid lineages examined here.

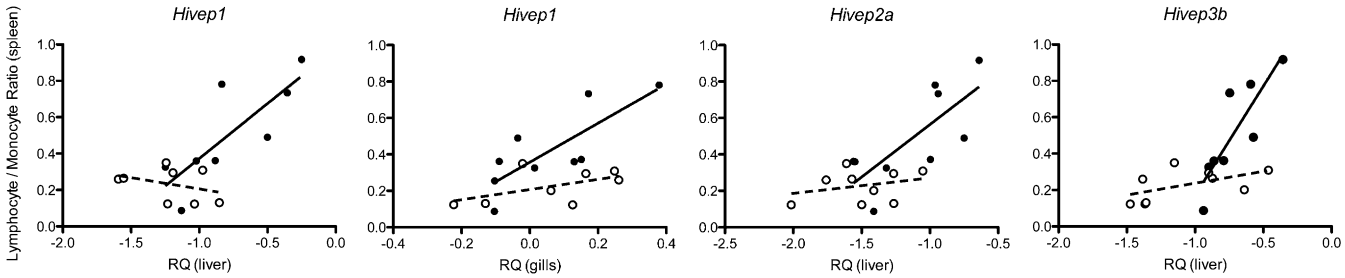


Figure 5 Correlations between immune response measurements and *Hivep* gene expression levels in *A. burtoni* adults. The relative gene expression levels [relative quantities (RQ)] of *Hivep1*, *Hivep2a*, and *Hivep3b* measured in gills and/or liver correlated with the observed lymphocyte/monocyte ratios measured in spleen for the control treatment (open circles and dashed fitted trend) and experimental treatment (closed circles and black fitted trend lines). Significant correlations were only observed for the experimental group: *Hivep1* (liver; Pearson $r = 0.798$, $P = 0.018$); *Hivep1* (gills; Pearson $r = 0.794$, $P = 0.011$); *Hivep2a* (Pearson $r = 0.745$, $P = 0.021$); and *Hivep3b* (Pearson $r = 0.852$, $P = 0.007$).

Positively selected genes are typically only loosely connected to reproductive isolation in *Drosophila* (Wu and Ting 2004). This is in contrast to the vertebrate MHC loci, known for their signatures of positive selection (Hughes and Nei 1988; Hughes and Nei 1989; Yang and Bielawski 2000; Montoya-Burgos 2011), which have been proposed as pleiotropic speciation genes (Eizaguirre *et al.* 2009). Because these genes are involved in adaptation to novel habitats in response to different pathogenic communities and assortative mating via mate choice, their pleiotropic effects are hypothesized to induce and accelerate speciation (Eizaguirre *et al.* 2009). Our work shows that several of the *Hivep* paralogs also have putative pleiotropic roles in immune defense and an important sexually selected trait—the anal fin egg-spot in East African haplochromine cichlids—subject to mate choice (Hert 1989; Couldridge 2002; but see Theis *et al.* 2012). Mate choice for the most attractive male anal fin could thus select a certain *Hivep* genotype and thereby facilitate adaptation to pathogenic environments at the same time. Similar to the MHC loci, the *Hivep* paralogs might have played important roles during the explosive speciation events of cichlid fishes and therefore are exciting new putative speciation genes.

Hivep3b: Selective patterns in haplochromines and other derived cichlid lineages

That we found evidence of lineage-specific positive selection acting on *Hivep3b* indicates that this locus underwent adaptive protein evolution in both the derived cichlid lineages, including the lamprologines, eretmodines, and the C-lineage, and the most species-rich cichlid lineage, the haplochromines. Adaptive protein evolution underlies the adaptive evolution of traits and is thus ultimately responsible for species divergence and evolutionary innovation (Yang 2006). Interestingly, the elevated d_N/d_S ratios were observed in lineages that are characterized by explosive speciation and diversification events (Salzburger *et al.* 2005; Day *et al.* 2008), which can be seen as further support for the hypothesis that the pleiotropic functions of the *Hivep* paralogs —*Hivep3b* specifically— can be linked to speciation events. During such events genes could have been recruited to perform altered functions to generate novel or modified traits, which ultimately may have played a role in the divergence between species. A lineage-specific amino acid substitution in *Hivep3b* was observed for all the haplochromines (position 112 G → R) and the derived lineages (position 87 E → D), as well as several substitutions in a subset of the species belonging to these lineages. Functional analyses are now needed to test whether these substitutions have a fitness advantage for these species and, above all, their function in these cichlid lineages.

ACKNOWLEDGMENTS

We thank past and current members of the Salzburger group for their contribution to sampling during fieldwork; Brigitte Aeschbach and Nicolas Boileau for assistance during laboratory work; and Verena Klein, Anne Beemelmanns, and Susanne Landis for help with the fish dissections and the experimental vaccinations. We also thank Jörn Scharsack for support in establishing the immune parameter measurements. The Inflammation Center at the University of Kiel kindly provided access to the flow cytometer facilities to perform immune measurements. Gene expression data analyzed in this article were generated in the Genetic Diversity Centre of ETH Zurich. We express our gratitude to Louis Du Pasquier, Christian Michel, Christophe Eizaguirre, Joost Raeymaekers, two anonymous reviewers, and the Editor for their constructive comments and suggestions. We also thank the BROAD Institute for sharing unpublished cichlid genome sequence data with the community. This work was supported by the European Research Council (starting grant "INTERGENADAPT" to W.S.), the Swiss National Science Foundation (grant 3100A0_122458 to W.S.), the Freiwillige Akademische Gesellschaft Basel (dissertation support grant to E.T.D.), the Volkswagenstiftung (grant to O.R.), and the German Research Foundation (grant to O.R. and W.S.).

LITERATURE CITED

Abdel-Aziz, E.-S. H., S. B. S. Abdu, T. E.-S. Ali, and H. F. Fouad, 2010 Haemopoiesis in the head kidney of tilapia, *Oreochromis niloticus* (Teleostei: Cichlidae): A morphological (optical and ultrastructural) study. Fish Physiol. Biochem. 36: 323–336.

Akira, S., S. Uematsu, and O. Takeuchi, 2006 Pathogen recognition and innate immunity. Cell 124: 783–801.

Albertson, R. C., J. T. Streelman, T. D. Kocher, and P. C. Yelick, 2005 Integration and evolution of the cichlid mandible: The molecular basis of alternate feeding strategies. Proc. Natl. Acad. Sci. USA 102: 16287–16292.

Altizer, S., C. D. Harvell, and E. Friedle, 2003 Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol. Evol. 18: 589–596.

Arora, K., H. Dai, S. G. Kazuko, J. Jamal, M. B. O'Connor *et al.*, 1995 The *drosophila schnurri* gene acts in the *Dpp/TGFβ* signaling pathway and encodes a transcription factor homologous to the human *MBP* family. Cell 81: 781–790.

Baldwin, A. S., Jr., K. P. LeClair, H. Singh, and P. A. Sharp, 1990 A large protein containing zinc finger domains binds to related sequence elements in the enhancers of the *class I Major Histocompatibility Complex* and *Kappa Immunoglobulin* genes. Mol. Cell. Biol. 10: 1406–1414.

Birrer, S. C., T. B. H. Reusch, and O. Roth, 2012 Salinity change impairs pipefish immune defence. Fish Shellfish Immunol. 33: 1238–1248.

Biswas, S., and J. M. Akey, 2006 Genomic insights into positive selection. Trends Genet. 22: 437–446.

Blais, J., C. Rico, C. van Oosterhout, J. Cable, G. F. Turner *et al.*, 2007 MHC adaptive divergence between closely related and sympatric African cichlids. PLoS One 2: e734.

Boots, M., A. Best, M. R. Miller, and A. White, 2008 The role of ecological feedbacks in the evolution of host defence: What does theory tell us? Philos. T. R. Soc. B 364: 27–36.

Braasch, I., J. N. Volff, and M. Schartl, 2009 The endothelin system: Evolution of vertebrate-specific ligand-receptor interactions by three rounds of genome duplication. Mol. Biol. Evol. 26: 783–799.

Brudno, M., S. Malde, A. Poliakov, C. B. Do, O. Couronne *et al.*, 2003 Glocal alignment: finding rearrangements during alignment. Bioinformatics 19: i54–i62.

Campbell, D. B., and P. Levitt, 2003 Regionally restricted expression of the transcription factor *c-myc intron 1* binding protein during brain development. J. Comp. Neurol. 467: 581–592.

Clabaut, C., W. Salzburger, and A. Meyer, 2005 Comparative phylogenetic analyses of the adaptive radiation of Lake Tanganyika cichlid fish: Nuclear sequences are less homoplasious but also less informative than mitochondrial DNA. J. Mol. Evol. 61: 666–681.

Colombo, M., E. T. Diepeveen, M. Muschick, E. Santos, A. Indermaur *et al.*, 2013 The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol. Ecol. 22: 670–684.

Cooper, M. D., and M. N. Alder, 2006 The evolution of adaptive immune systems. Cell 124: 815–822.

Couldridge, V., 2002 Experimental manipulation of male eggspots demonstrates female preference for one large spot in *Pseudotropheus lombardoi*. J. Fish Biol. 60: 726–730.

Dai, H., C. Hogan, B. Gopalakrishnan, J. Torres-Vazquez, M. Nguyen *et al.*, 2000 The zinc finger protein *Schnurri* acts as a *Smad* partner in mediating the transcriptional response to *Decapentaplegic*. Dev. Biol. 227: 373–387.

Day, J. J., J. A. Cotton, and T. G. Barraclough, 2008 Tempo and mode of diversification of lake Tanganyika cichlid fishes. PLoS One 3: e1730.

Déry, O., C. U. Corvera, M. Steinhoff, and N. W. Bunnett, 1998 Proteinase-activated receptors: novel mechanisms of signaling by serine proteases. Am. J. Physiol. 274: C1429–C1452.

Diepeveen, E. T., and W. Salzburger, 2011 Molecular characterization of two endothelin pathways in East African cichlid fishes. J. Mol. Evol. 73: 355–368.

Diepeveen, E. T., F. D. Kim, and W. Salzburger, 2013 Sequence analyses of the *distal-less homeobox* gene family in East African cichlid fishes reveal signatures of positive selection. BMC Evol. Biol. 13: 153.

Dijkstra, P. D., G. F. Wiegertjes, M. Forlenza, I. van der Sluijs, and H. A. Hofmann, 2011 The role of physiology in the divergence of two incipient cichlid species. J. Evol. Biol. 24: 2639–2652.

Dürr, U., K. A. Henningfeld, T. Hollemann, W. Knöchel, and T. Pieler, 2004 Isolation and characterization of the *Xenopus HIVEP* gene family. Eur. J. Biochem. 271: 1135–1144.

Ebert, D., and W. D. Hamilton, 1996 Sex against virulence: The coevolution of parasitic diseases. Trends Ecol. Evol. 11: 79–82.

Eizaguirre, C., T. L. Lenz, A. Traulsen, and M. Milinski, 2009 Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. Ecol. Lett. 12: 5–12.

Endo, T., K. Ikeo, and T. Gojobori, 1996 Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. 13: 685–690.

Flajnik, M. F., and M. Kasahara, 2010 Origin and evolution of the adaptive immune system: Genetic events and selective pressures. Nat. Rev. Genet. 11: 47–59.

Fraser, G. J., R. F. Bloomquist, and J. T. Streelman, 2008 A periodic pattern generator for dental diversity. BMC Biol. 6: 32.

Fraser, G. J., C. D. Hulsey, R. F. Bloomquist, K. Uyesugi, N. R. Manley *et al.*, 2009 An ancient gene network is co-opted for teeth on old and new jaws. PLoS Biol. 7: e31.

Frazer, K. A., L. S. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak, 2004 VISTA: Computational tools for comparative genomics. Nucleic Acids Res. 32: W273–W279.

Gerrard, D. T., and A. Meyer, 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.

Grieder, N. C., D. Nellen, R. Burke, K. Basler, and M. Affolter, 1995 *Schnurri* is required for *Drosophila Dpp* signaling and encodes a zinc finger protein similar to the mammalian transcription factor *PRDII-BF1*. Cell 81: 791–800.

Guindon, S., and O. Gascuel, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52: 696–704.

Haldane, J. B. S., 1949 Disease and evolution. Ric. Sci. 19 (Suppl. A): 68–76.

Hamilton, W. D., 1980 Sex *vs.* non-sex *vs.* parasite. Oikos 35: 282–290.

Hert, E., 1989 The function of egg-spots in an African mouth-brooding cichlid fish. Anim. Behav. 37: 726–732.

Hicar, M. D., Y. Liu, C. E. Allen, and L.-C. Wu, 2001 Structure of the human zinc finger protein *HIVEP3*: Molecular cloning, expression, exon–intron structure, and comparison with paralogous genes *HIVEP1* and *HIVEP2*. Genomics 71: 89–100.

Hong, J., and L.-C. Wu, 2005 Structural characterization of the gene encoding the large zinc finger protein *ZAS3*: Implication to the origin of multiple promoters in eukaryotic genes. Biochim. Biophys. Acta. 1681: 74–87.

Huelsenbeck, J. P., and F. Ronquist, 2001 MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at *Major Histocompatibility Complex class I* loci reveals overdominant selection. Nature 335: 167–170.

Hughes, A. L., and M. Nei, 1989 Nucleotide substitution at *Major Histocompatibility Complex class II* loci: Evidence for overdominant selection. Proc. Natl. Acad. Sci. USA 86: 958–962.

Janeway, C. A., C. C. Goodnow, and R. Medzhitov, 1996 Danger–Pathogen on the premises! Immunological tolerance. Curr. Biol. 6: 519–522.

Janeway, C. A., and R. Medzhitov, 2002 Innate immune recognition. Annu. Rev. Immunol. 20: 197–216.

Janeway, C. A., M. J. Walport, and P. Travers, 2001 *Immunobiology: The immune system in health and disease*, Garland Publisher, New York, USA.

Jankovic, D., and G. Trinchieri, 2007 *IL-10* or not *IL-10*: that is the question. Nat. Immunol. 8: 1281–1283.

Jansa, S. A., B. L. Lundrigan, and P. K. Tucker, 2003 Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *mus*. J. Mol. Evol. 56: 294–307.

Jiggins, F. M., and K. W. Kim, 2007 A screen for immunity genes evolving under positive selection in *Drosophila*. J. Evol. Biol. 20: 965–970.

Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady *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.

Keesing, F., L. K. Belden, P. Daszak, A. Dobson, C. Drew Harvell *et al.*, 2010 Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature 468: 647–652.

Kocher, T. D., 2004 Adaptive evolution and explosive speciation: The cichlid fish model. Nat. Rev. Genet. 5: 288–298.

Kornfield, I., and P. F. Smith, 2000 African cichlid fishes: model systems for evolutionary biology. Annu. Rev. Ecol. Syst. 31: 163–196.

Liang, J., L. Yu, J. Yin, and C. Savage-Dunn, 2007 Transcriptional repressor and activator activities of *SMA-9* contribute differentially to *BMP*-related signaling outputs. Dev. Biol. 305: 714–725.

Liu, S., F. Madiai, K. V. Hackshaw, C. E. Allen, J. Carl *et al.*, 2011 The large zinc finger protein *ZAS3* is a critical modulator of osteoclastogenesis. PLoS One 6: e17161.

Loots, G. G., I. Ovcharenko, L. Pachter, I. Dubchak, and E. M. Rubin, 2002 rVista for comparative sequence-based discovery of functional transcription factor binding sites. Genome Res. 12: 832–839.

Maekawa, T., H. Sakura, T. Sudo, and S. Ishii, 1989 Putative metal finger structure of the *human immunodeficiency virus type 1 enhancer binding protein HIV-EP1*. J. Biol. Chem. 264: 14591.

Marty, T., B. Müller, K. Basler, and M. Affolter, 2000 *Schnurri* mediates *Dpp*-dependent repression of *brinker* transcription. Nat. Cell Biol. 2: 745–749.

Matthews, B., L. J. Harmon, L. M'Gonigle, K. B. Marchinko, and H. Schaschl, 2010 Sympatric and allopatric divergence of MHC genes in threespine stickleback. PLoS One 5: e10948.

Mayor, C., M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin *et al.*, 2000 VISTA: Visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16: 1046–1047.

Meyer, A., and Y. Van de Peer, 2005 From 2R to 3R: Evidence for a fish-specific genome duplication (FSGD). Bioessays 27: 937–945.

Montoya-Burgos, J. I., 2011 Patterns of positive selection and neutral evolution in the protein-coding genes of *Tetraodon* and *Takifugu*. PLoS One 6: e24800.

Muchardt, C., J. S. Seeler, A. Nirula, D. L. Shurland, and R. B. Gaynor, 1992 Regulation of *human immunodeficiency virus enhancer* function by *PRDII-BF1* and *c-rel* gene products. J. Virol. 66: 244–250.

Muschick, M., M. Barluenga, W. Salzburger, and A. Meyer, 2011 Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol. Biol. 11: 116.

Muschick, M., A. Indermaur, and W. Salzburger, 2012 Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22: 2362–2368.

Møller, A. P., and J. Erritzøe, 1998 Host immune defence and migration in birds. Evol. Ecol. 12: 945–953.

Ng, P., and S. T. Henikoff, 2003 SIFT: Predicting amino acid changes that affect protein functions. Nucleic Acids Res. 31: 3812–3814.

Nielsen, R., 2005 Molecular signatures of natural selection. Annu. Rev. Genet. 39: 197–218.

Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton *et al.*, 2005 A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3: e170.

Posada, D., 2008 jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25: 1253–1256.

Press, C. M., and O. Evensen, 1999 The morphology of the immune system in teleost fishes. Fish Shellfish Immun. 9: 309–318.

Roberts, R. B., J. R. Ser, and T. D. Kocher, 2009 Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326: 998–1001.

Rodríguez, R. M., A. López-Vázquez, and C. López-Larrea, 2012 Immune systems evolution, pp. 237–251 in *Sensing in Nature: Advances in experimental medicine and biology*, edited by C. López-Larrea. Landes Bioscience and Springer Science and Business Media. New York, USA.

Rombout, J. H. W. M., H. B. T. Huttenhuis, S. Picchietti, and G. Scapigliati, 2005 Phylogeny and ontogeny of fish leucocytes. Fish Shellfish Immunol. 19: 441–455.

Ronquist, F., and J. P. Huelsenbeck, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Roth, O., J. P. Scharsack, I. Keller, and T. B. H. Reusch, 2011 Bateman's principle and immunity in a sex-role reversed pipefish. J. Evol. Biol. 24: 1410–1420.
Roth, O., I. Keller, S. H. Landis, W. Salzburger, and T. B. H. Reusch, 2012 Hosts are ahead in a marine host-parasite coevolutionary arms race: Innate immune system adaptation in pipefish *syngnathus typhle* against *Vibrio* phylotypes. Evolution 66: 2528–2539.
Rothmeier, A. S., and W. Ruf, 2012 *Protease-activated receptor 2* signaling in inflammation. Semin. Immunopathol. 34: 133–149.
Sackton, T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark *et al.*, 2007 Dynamic evolution of the innate immune system in *Drosophila*. Nat. Genet. 39: 1461–1468.
Saita, Y., T. Takagi, K. Kitahara, M. Usui, K. Miyazono *et al.*, 2007 Lack of *Schnurri-2* expression associates with reduced bone remodeling and osteopenia. J. Biol. Chem. 282: 12907–12915.
Salzburger, W., 2009 The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 18: 169–185.
Salzburger, W., A. Meyer, S. Baric, E. Verheyen, and C. Sturmbauer, 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., and A. Meyer, 2004 The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften 91: 277–290.
Salzburger, W., T. Mack, E. Verheyen, and A. Meyer, 2005 Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 5: 17.
Salzburger, W., I. Braasch, and A. Meyer, 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.
Santos, M. E., and W. Salzburger, 2012 How cichlids diversify. Science 338: 619–621.
Savan, R., D. Igawa, and M. Sakai, 2003 Cloning, characterization and expression analysis of *interleukin-10* from the common carp, *Cyprinus carpio L.* Eur. J. Biochem. 270: 4647–4654.
Sawyer, S. L., M. Emerman, and H. S. Malik, 2004 Ancient adaptive evolution of the primate antiviral DNA-editing enzyme *APOBEC3G*. PLoS Biol. 2: e275.
Scharsack, J. P., M. Kalbe, C. Harrod, and G. Rauch, 2007 Habitat-specific adaptation of immune responses of stickleback (*Gasterosteus aculeatus*) lake and river ecotypes. P. Roy. Soc. B: Biol. Sci. 274: 1523–1532.
Schlenke, T. A., and D. J. Begun, 2003 Natural selection drives *drosophila* immune system evolution. Genetics 164: 1471–1480.
Schluter, D., 2000 *The Ecology of Adaptive Radiation*, edited by P. H. Harvey, R. M. May. Oxford: Oxford University Press.
Seehausen, O., 2006 African cichlid fish: a model system in adaptive radiation research. Proc. Biol. Sci. 273: 1987–1998.
Seeler, J. S., C. Muchardt, A. Suessle, and R. B. Gaynor, 1994 Transcription factor *PRDII-BF1* activates *human immunodeficiency virus type 1* gene expression. J. Virol. 68: 1002–1009.
Sidow, A., 1996 Gen(om)e duplications in the evolution of early vertebrates. Curr. Opin. Genet. Dev. 6: 715–722.
Swofford D. L., 2002 Phylogenetic analysis using parsimony (∗and other methods). Sinauer Associates, Sunderland, Massachusetts.
Takagi, T., W. Jin, K. Taya, G. Watanabe, K. Mori *et al.*, 2006 *Schnurri-2* mutant mice are hypersensitive to stress and hyperactive. Brain Res. 1108: 88–97.
Takahashi, T., 2003 Systematics of Tanganyikan cichlid fishes (Teleostei: Perciformes). Ichthyol. Res. 50: 367–382.
Taylor, J. S., I. Braasch, T. Frickey, A. Meyer, and Y. Van de Peer, 2003 Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res. 13: 382–390.
Terai, Y., N. Morikawa, K. Kawakami, and N. Okada, 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., N. Morikawa, and N. Okada, 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.
Terai, Y., N. Morikawa, K. Kawakami, and N. Okada, 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.
Theis, A., W. Salzburger, and B. Egger, 2012 The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni*. PLoS One 7: e29878.
Thompson, J. N., 1988 Variation in interspecific interactions. Annu. Rev. Ecol. Syst. 19: 65–87.
Toll-Riera, M., S. Laurie, and M. M. Alba, 2011 Lineage-specific variation in intensity of natural selection in mammals. Mol. Biol. Evol. 28: 383–398.
Torres-Vazquez, J., S. Park, R. Warrior, and K. Arora, 2001 The transcription factor *Schnurri* plays a dual role in mediating *Dpp* signaling during embryogenesis. Development 128: 1657–1670.
Tschirren, B., L. Raberg, and H. Westerdahl, 2011 Signatures of selection acting on the innate immunity gene *Toll-like receptor 2* (*TLR2*) during the evolutionary history of rodents. J. Evol. Biol. 24: 1232–1240.
Vamathevan, J. J., S. Hasan, R. D. Emes, H. Amrine-Madsen, D. Rajagopalan *et al.*, 2008 The role of positive selection in determining the molecular cause of species differences in disease. BMC Evol. Biol. 8: 273.
Volff, J. N., 2005 Genome evolution and biodiversity in teleost fish. Heredity 94: 280–294.
Watano, K., K. Iwabuchi, S. Fujil, N. Ishimori, S. Mitsuhashi *et al.*, 2001 *Allograft inflammatory factor-1* augments production of *interleukin-6*, *-10* and *-12* by a mouse macrophage line. Immunogenetics 104: 307–316.
Whyte, S. K., 2007 The innate immune response of finfish–A review of current knowledge. Fish Shellfish Immunol. 23: 1127–1151.
Wickler, W., 1962 "Egg-dummies" as natural releasers in mouth-breeding cichlids. Nature 194: 1092–1093.
William, R., E. Leisha, I. Alexander, B. Melissa, I. Ichiro *et al.*, 1995 The *Myc* intron-binding polypeptide associates with *RFX1* in vivo and binds to the *Major Histocompatibility Complex Class II* promoter region, to the *Hepatitis B virus enhancer*, and to regulatory regions of several distinct viral genes. Mol. Cell. Biol. 15: 3041–3048.
Woolley, S., J. Johnson, M. J. Smith, K. A. Crandall, and D. A. McClellan, 2003 TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671–672.
Wu, C.-I., and C.-T. Ting, 2004 Genes and speciation. Nat. Rev. Genet. 5: 114–122.
Wu, L.-C., 2002 ZAS: C2H2 zinc finger proteins involved in growth and development. Gene Expr. 10: 137–152.
Wu, L.-C., C. H. Mak, N. Dear, T. Boehm, L. Foroni *et al.*, 1993 Molecular cloning of a zinc finger protein which binds to the heptamer of the signal sequence for V (D) J recombination. Nucleic Acids Res. 21: 5067–5073.
Wu, L.-C., L. Yiling, J. Strandtmann, M. Chi-Ho, B. Lee *et al.*, 1996 The mouse DNA binding protein *Rc* for the *Kappa B* motif of transcription and for the V(D)J recombination signal sequences contains composite DNA–protein interaction domains and belongs to a new family of large transcriptional proteins. Genomics 35: 415–424.
Yang, Z., 1997 PAML, A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13: 555–556.
Yang, Z., and J. P. Bielawski, 2000 Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15: 496–503.
Yang, Z., W. Wong, and R. Nielsen, 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22: 1107–1118.
Yang, Z., 2006 *Computational molecular evolution*, Oxford University Press, Oxford.
Yang, Z., 2007 PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591.
Yao, L. C., 2006 *Schnurri* transcription factors from *Drosophila* and vertebrates can mediate *Bmp* signaling through a phylogenetically conserved mechanism. Development 133: 4025–4034.
Yin, K., L. Yu, and C. Savage-Dunn, 2010 Alternative trans-splicing of *Caenorhabditis elegans sma-9*/*schnurri* generates a short transcript that provides tissue-specific function in *BMP* signaling. BMC Mol. Biol. 11: 46.
Zhu, C., V. Lee, A. Finn, K. Senger, A. A. Zarrin *et al.*, 2012 Origin of *immunoglobulin* isotype switching. Curr. Biol. 22: 872–880.

Communicating editor: A. Rokas

CHAPTER 4

IMMUNE-RELATED FUNCTIONS OF THE *HIVEP* GENE FAMILY IN EAST AFRICAN CICHLID FISHES

SUPPORTING MATERIAL

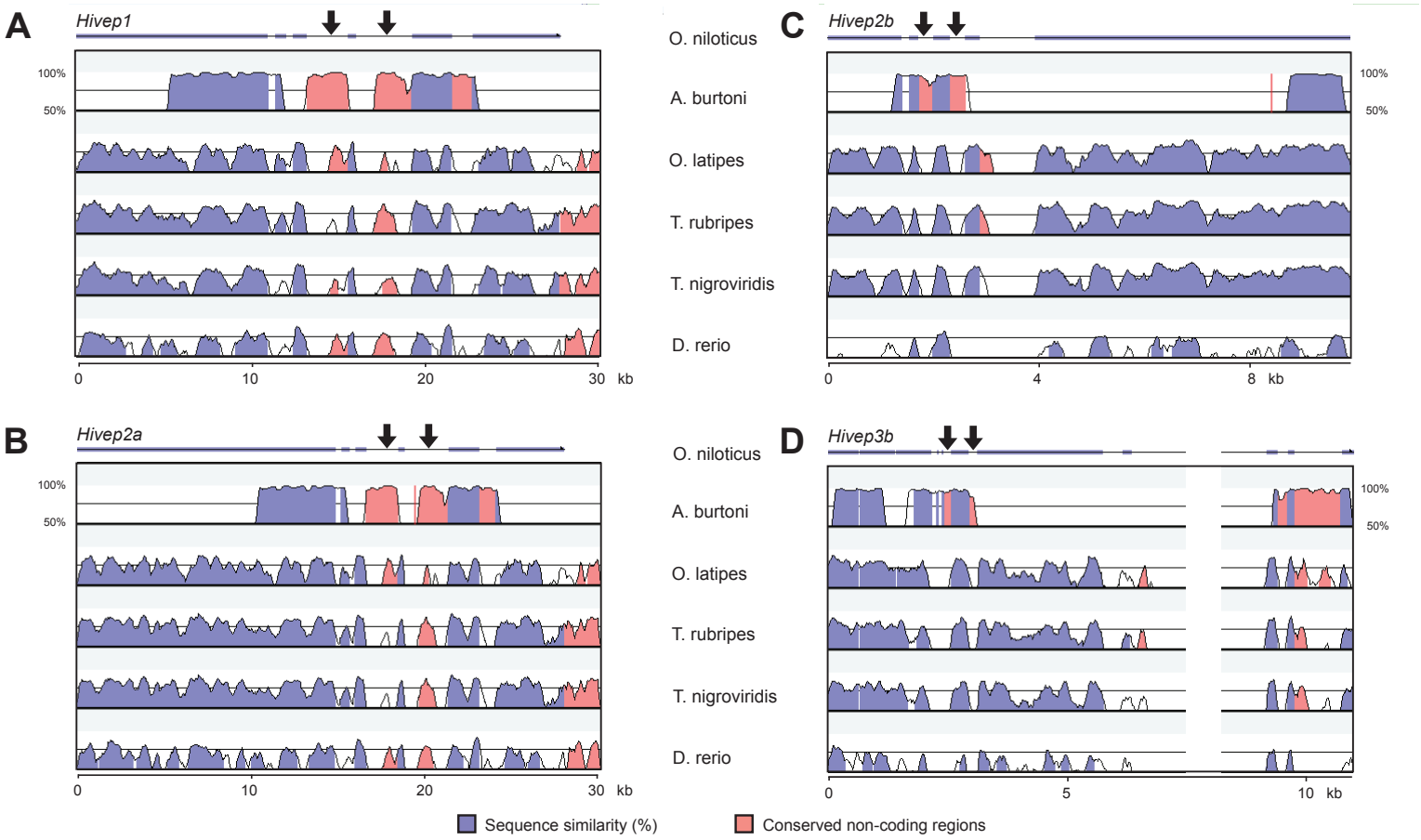
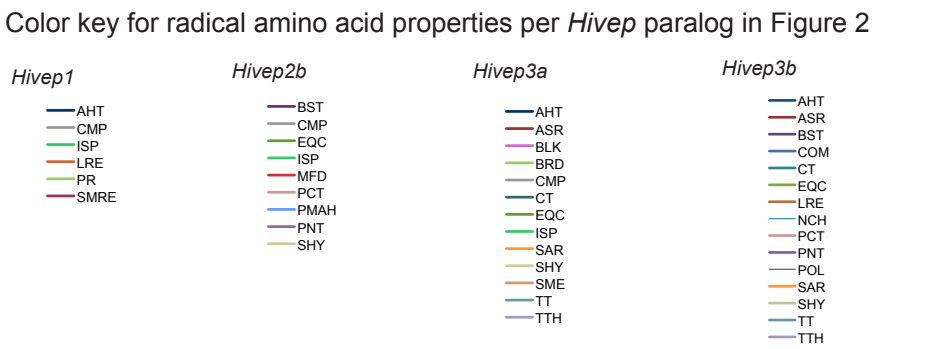


Figure S1 Vistaplots for four *Hivep* paralogs. Sequence similarity with a minimum of 50% identity between *A. burtoni* (top panel), *O. latipes* (second panel), *T. rubripes* (mid panel), *T. nigroviridis* (fourth panel), *D. rerio* (bottom panel) and *O. niloticus* (exon-intron structure on top of figure) are displayed in purple. Pink regions indicate conserved non-coding regions (CNEs). Arrows indicate the two identified CNEs in all four paralogs. Three CNEs were observed in *Hivep1* (A) and *Hivep2a* (B) in *A.burtoni*, of which two were commonly observed among teleosts. (C) Two cichlid-specific CNEs were observed in *Hivep2b.* (D) In *Hivep3b* four CNEs were observed.



AHT=Alpha-helical tendencies
ASR=Average number of surrounding residues
BLK=Bulkiness
BRD=Buriedness
BST=Beta-structure tendencies
CMP=Compressibility
COM=Composition
CT=Coil tendencies
EQC=Equilibrium constant (ionization of COOH)
ISP=Isoelectric point
LRE=Long-range non-bonded energy
MFD=Mean r.m.s. fluctuation displacement
NCH=Normalized consensus hydrophobicity
PCT=Power to be at the C-terminal
PMAH=Power to be at the middle of alpha-helix
PNT=Power to be at the N-terminal
POL=Polarity
PR=Polar requirement
SAR=Solvent accessible reduction ratio
SHY=Surrounding hydrophobicity
SMRE=Short and medium range non-bonded energy
TT=Turn tendencies
TTH=Thermodynamic transfer hydrophobicity

Figure S2 Overview of the radical amino acid properties identified by the TreeSAAP analyses for four *Hivep* paralogs. Colors correspond to the sliding window plots in Fig. 2 and amino acid properties abbreviations are displayed for each paralog individually.

Table S1 Species names, Tribes and GenBank accession numbers of the five sequenced *Hivep* paralogs

Species	Tribe	*Hivep1*	*Hivep2a*	*Hivep2b*	*Hivep3a*	*Hivep3b*
Bathybates graueri	Bathybatini	KF049218	KF049276	KF049316	KF049357	KF049398
Benthochromis tricoti	Benthochromis	KF049245	KF049264	KF049303	KF049373	KF049385
Boulengerochromis microlepis	Boulengerochromini	KF049229	KF049258	KF049297	KF049337	KF049377
Ctenochromis benthicola	Cyphotilapiini	KF049241	KF049267	KF049306	KF049347	KF049388
Cyphotilapia frontosa	Cyphotilapiini	KF049237	KF049293	KF049333	KF049371	KF049415
Cyprichromis leptosoma	Cyprichromini	KF049227	KF049265	KF049304	KF049345	KF049386
Aulocranus dewindtii	Ectodini	KF049244	KF049263	KF049335	KF049344	KF049384
Callochromis macrops	Ectodini	KF049228	KF049268	KF049307	KF049348	KF049389
Cyathopharynx furcifer	Ectodini	KF049238	KF049296	KF049336	KF049341	KF049381
Grammatotria lemairii	Ectodini	KF049249	KF049272	KF049311	KF049352	KF049393
Xenotilapia flavipinnis	Ectodini	KF049256	KF049288	KF049328	KF049366	KF049410
Xenotilapia spiloptera	Ectodini	KF049254	KF049286	KF049326	KF049364	KF049408
Eretmodus cyanostictus	Eretmodini	KF049239	KF049295	KF049298	KF049338	KF049378
Astatotilapia burtoni	Haplochromini	KF049240	KF049260	KF049300	KF049340	KF049380
Ctenochromis horei	Haplochromini	KF049246	KF049290	KF049330	KF049368	KF049412
Altolamprologus fasciatus	Lamprologini	KF049243	KF049289	KF049329	KF049367	KF049411
Julidochromis ornatus	Lamprologini	KF049248	KF049271	KF049310	KF049351	KF049392
Lamprologus lemairii	Lamprologini	KF049255	KF049273	KF049312	KF049353	KF049394
Lepidiolamprologus elongatus	Lamprologini	KF049242	KF049294	KF049334	KF049372	KF049416
Neolamprologus furcifer	Lamprologini	KF049230	KF049275	KF049315	KF049356	KF049397
Neolamprologus pulcher	Lamprologini	KF049231	KF049274	KF049314	KF049355	KF049396
Neolamprologus sexfasciatus	Lamprologini	KF049250	-*	KF049313	KF049354	KF049395
Neolamprologus tetracanthus	Lamprologini	KF049251	KF049278	KF049318	KF049374	KF049400
Telmatochromis dhonti	Lamprologini	KF049257	KF049284	KF049324	KF049375	KF049406
Variabilichromis moorii	Lamprologini	KF049219	KF049292	KF049332	KF049370	KF049414
Gnathochromis permaxillaris	Limnochromini	KF049232	KF049269	KF049308	KF049349	KF049390
Limnochromis staneri	Limnochromini	KF049220	KF049262	KF049302	KF049343	KF049383
Reganochromis calliurus	Limnochromini	KF049252	KF049282	KF049322	KF049376	KF049404
Haplotaxodon microlepis	Perissodini	KF049247	KF049270	KF049309	KF049350	KF049391
Perissodus microlepis	Perissodini	KF049233	KF049279	KF049319	KF049359	KF049401
Plecodus straeleni	Perissodini	KF049221	KF049261	KF049301	KF049342	KF049382
Oreochromis tanganicae	Tilapiini	KF049234	KF049277	KF049317	KF049358	KF049399
Trematocara nigrifrons	Trematocarini	KF049235	KF049280	KF049320	KF049360	KF049402
Lobochilotes labiatus	Tropheini	KF049222	KF049287	KF049327	KF049365	KF049409
Petrochromis famula	Tropheini	KF049236	KF049281	KF049321	KF049361	KF049403
Petrochromis polyodon	Tropheini	KF049253	KF049283	KF049323	KF049362	KF049405
Pseudosimochromis curvifrons	Tropheini	KF049223	KF049259	KF049299	KF049339	KF049379
Simochromis diagramma	Tropheini	KF049224	KF049291	KF049331	KF049369	KF049413
Tropheus moori	Tropheini	KF049225	KF049266	KF049305	KF049346	KF049387
Tylochromis polylepis	Tylochromini	KF049226	KF049285	KF049325	KF049363	KF049407

* The *Hivep2a* sequence for *N. sexfasciatus* could not be submitted to GenBank, due to a lack of completeness

Table S2 Ensemble accession numbers or genomic location of teleost *Hivep* sequences used for primer design, phylogenetic and/or Vista analyses

Species	Ensemble/GenBank accession numbers and preliminary genome information				
	Hivep1	***Hivep2a***	***Hivep2b***	***Hivep3a***	***Hivep3b***
Danio rerio	ENSDARG00000079528	ENSDARG00000039987	ENSDARG00000018773	ENSDARG00000075928	ENSDARG00000037154
Gadus morhua				ENSGMOG00000005293	
Gasterosteus aculeatus		ENSGACG00000011974	ENSGACG00000011743		ENSGACG00000007350
Tetraodon nigrovirdis	ENSTNIG00000004491	ENSTNIG00000010718	ENSTNIG00000019322		ENSTNIG00000009667
Takifugu rubripes	scaffold_69:866682-876734:1	ENSTRUG00000015889	ENSTRUG00000007039	ENSTRUG00000010896	ENSTRUG00000013983
Oryzias latipes	ENSORLG00000009270	ENSORLG00000018133	ENSORLG00000008843		ENSORLG00000019868
Oreochromis niloticus	ENSONIG00000013692	ENSONIG00000018691	ENSONIG00000000567		ENSONIG00000006743
Astatotilapia burtoni[1]			scaffold_477	scaffold_91	
Neolamprologus brichardi[1]			scaffold_57	scaffold_36	
Pundamilia nyererei[1]			scaffold_69	scaffold_147	

[1] v1 assembly of the preliminary whole genome sequences by the Cichlid Genome Consortium (cichlid.umd.edu/CGCindex.html)

Table S3 Primer names and Sequences used for Sequencing (top) and Gene-expression assay (bottom)

Locus	Forward primer	5' - 3' Sequence	Reverse Primer	5' - 3' Sequence
Hivep1	Hivep1_2F93	CGAGAGAACATCCAAAACAAAC	Hivep1_C4R74	ATCCCACATTCCTCACAGA
	Hivep1_4F105	AACACATCCGCACCCACTC	Hivep1_5R31	GCCTTGGATTTCATGTGCTT
	Hivep1_4F144	GCGTCCACTGCAACTTCTC	Hivep1_5R59	CCCCATCTCCAAGCATTTC
	Hivep1_5F7	TGACCAAGCACATGAAATCC	Hivep1_6R471	GGCACATCAAAGTCAGGGTAG
	Hivep1_6F159	CAGCACATCTCTTCCAAA	Hivep1_7R131	GGTCTCTGTTCCTGGGGTTT
Hivep2a	Hivep2a_1F73	CCCAGCTCNTTTGGACAGTA	Hivep2a_2R484	AGGTTAGGAGGGGTGGAAGT
	Hivep2a_2F219	GCCTGACTACCCTGAAGCAA	Hivep2a_3R133	TTTCAGGCAACTCCATAGGC
	Hivep2a_8F129	CCGACCCTATGTCTGCAAGT	Hivep2a_9R95	GTCTGCCTCAGTGTCCTCAA
Hivep2b	Hivep2b_1F6	AGCAGAAACCACAAAAACCT	Hivep2b_1R500	CTTGATAGGACTTTGCGTTTG
	Hivep2b_2F42	CGGGAAGAAGGTGAAAGACA	Hivep2b_5R129	CGTCCTCATCGTCATCTTCA
Hivep3a	Hivep3a_1F1	CAGAAACGTGAGCGTAAG	Hivep3a_1R936	ATTTGARGGTGGGTGATGAA
	Hivep3a_1F856	GTTGGCGATGTGGGCTAT	Hivep3a_1R841	TGATGTTTTGCTGTGGTGGT
	Hivep3a_1F1297	TGCCTAKCACATCAACCCATC	Hivep3a_1R2243	CTTTCYCTCCACCCAAYTGC
Hivep3b	Hivep3b_1F8	GAGAAGAAACCACAAAAAC	Hivep3b_1R123	GGCTTCTCCTTTGGGATCA
	Hivep3b_1F132	AGCCTGCCATAGAGGTTCC	Hivep3b_4R50	AATGTCTGACTTCGCTGAGG
	Hivep3b_F40	CGGCAGGGGAAAGTACATA	Hivep3b_R223	CCTTGCTTAGAGTGCCTGCT

Locus	Forward sequence	Reverse sequence
Hivep1	ATGCCCGCAGCAGGAA	TCCATGTAGATGTCCAAGCTGTTT
Hivep2a	TGCCATGTCCCAGTCACTTAA	GAGCGTGAAAAGTAGCCAGAGTCT
Hivep2b	CGAGGCTAGTCGTGCAATTCA	AGAGTCACTGCTTCTTTTTTCAGACA
Hivep3a	GAGCAGCCAGCTAGCCAAAC	ATAAACTGGCTTCAGGTGGCATT
Hivep3b	CCAGACAGCCACAGCAACAA	CCCCACATGTTCCACATTCA
EF1	GCCCCTGCAGGACGTCTA*	CGGCCGACGGGTACAGT*
RpSA3	AGACCAATGACCTGAAGGAAGTG*	TCTCGATGTCCTTGCCAACA*
AIF1	GGTTTGAAGAGGATGTTGGAGAA	CAACCACCTCTGCCATCATTT
IL10	TGGCGCTCGCGTCTTT	CGGCAGCAGCGGTTGT
F2RL1	TGGGAATAAAGCACACAAGGAA	CATCAAAGAACCGACACAAAGTTG
TLR5	GCATGTTGCTGGTCTCAATTTG	GAAGTGAGGCCCTTGAAAACAC

* Primers from Colombo *et al*. 2013

PART II

NEXT GENERATION SEQUENCING APPROACHES

CHAPTER 5

TWO DECADES OF MOLECULAR ECOLOGY: WHERE ARE WE AND WHERE ARE WE HEADING?

MEETING REVIEW

MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 5656–5659

NEWS AND VIEWS

MEETING REVIEW

Two Decades of Molecular Ecology: where are we and where are we heading?

EVELINE T. DIEPEVEEN* and WALTER SALZBURGER*

Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

The twentieth anniversary of the journal *Molecular Ecology* was celebrated with a symposium on the current state and the future directions of the field. The event, organized by Tim Vines and Loren Rieseberg, took place on the opening day of the First Joint Congress on Evolutionary Biology organized by the American Society of Naturalists (ASN), the Canadian Society for Ecology and Evolution (CSEE), the European Society for Evolutionary Biology (ESEB), the Society for the Study of Evolution (SSE) and the Society of Systematic Biologists (SSB) in Ottawa (Canada) from 6–10 July 2012. The get together of these five societies created a truly international and exciting "Evolution conference" and the ideal framework for the *Molecular Ecology* symposium. Its thirteen talks were grouped into the five different subject areas of the journal: Speciation and Hybridization; Landscape Genetics, Phylogeography and Conservation; Ecological Genomics and Molecular Adaptation; Kinship, Parentage and Behaviour; Ecological Interactions. Each session was followed by a panel discussion on the future direction of the subfield. That more than 300 colleagues registered for this special symposium illustrates the broad interest in, and appreciation of, molecular ecology – both the field and the journal.

Keywords: ecological genomics, ecological interactions, landscape genetics, molecular adaptation, phylogeography, speciation

Received 24 August 2012; revised 14 September 2012; accepted 19 September 2012

Next-generation (Sequencing in the field of) Molecular Ecology

The overlapping and dominating theme of almost all sessions (except maybe Kinship, Parentage and Behaviour) was next-generation sequencing (NGS). Taking the symposium as a snapshot of what is going on in the community, then it seems as if NGS has 'conquered' the field of molecular ecology (see also Fig. 1). At the same time, the different talks exemplified the broad applicability of NGS and the various questions that can be addressed using NGS technologies. In the following, we summarize the symposium by focussing on how NGS is used in each of the subfields.

Alex Buerkle (University of Wyoming) presented some recent work on reproductive isolation in *Lycaeides* butterflies and *Manacus* birds based on genome-wide single-nucleotide polymorphism (SNP) screens of both parental species and hybrids (see Gompert *et al.* 2012). They found that many loci show signs of extreme introgression along hybrid zones and that there are many fine-scaled genomic footprints of differentiation and reproductive isolation, calling into question the genomic island view of speciation. Alex Widmer (ETH Zurich) talked about his work on a hybrid zone in *Silene*. Previous amplification fragment length polymorphism (AFLP) genomic scans revealed that outlier markers were typically found on the sex chromosomes. In the absence of a reference genome, they built a reference transcriptome to determine gene expression differences between sexes. This way they were able to show that dosage compensation does occur in *Silene* by overexpression of the X-chromosome in males when the Y-chromosome is down-regulated (Muyle *et al.* 2012). Tatiana Giraud (University of South Paris) used NGS to sequence expressed sequence tag (EST) libraries of four *Microbotryum* species – fungal pathogens specialized on different host plants (Caryophyllaceae). These libraries were subjected to a genome-wide dN/dS analysis, which revealed 42 loci under positive selection that could be implicated with functions in the host–parasite interactions (Aguileta *et al.* 2010).

In the second session, Rose Andrew (University of British Columbia) talked about her research on dune sunflowers (*Helianthus*) in which she applied RAD sequencing to twenty subpopulations to identify putative adaptive loci. Several strong peaks of F_{ST} outliers were detected, leading to the conclusion that seed mass and vegetation cover were both associated with the same genomic region. Next, Victoria Sork (University of California, Los Angeles) stated that NGS is the essential tool for the development of a new subfield: landscape genomics. With NGS, the whole genome can be screened for adaptive genetic variation in relation to geographic patterns, making candidate gene approaches somewhat redundant. Furthermore, massive NGS genome scans provide valuable information on neutral loci that can be used for, for example, inferring demographic processes.

The field of ecological genomics and molecular adaptation has gained tremendously from the NGS revolution. Jon Slate (University of Sheffield), for instance, showed not

Correspondence: Walter Salzburger, Fax: +41 61 267 0301;
E-mail: walter.salzburger@unibas.ch

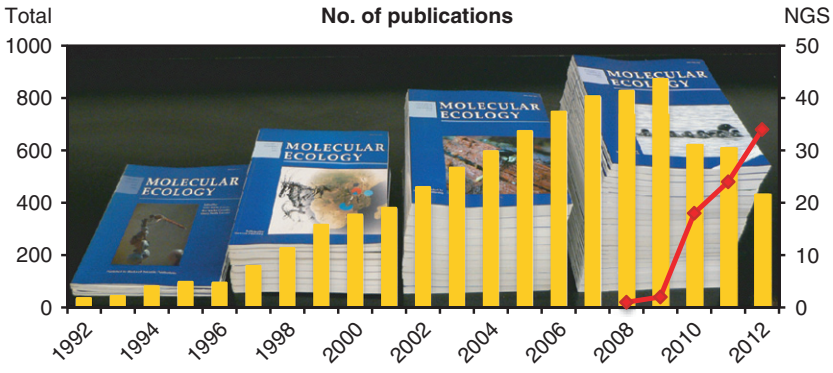


Fig. 1 The total number of publications (yellow bars; left *y*-axis) and the number of studies containing the keyword next-generation sequencing (NGS; red diamonds; right *y*-axis) in *Molecular Ecology* and *Molecular Ecology Resources* (previously *Molecular Ecology Notes*) per year. Numbers for 2012 are until August. The photograph was kindly provided by Louis Bernatchez.

only that NGS allows associating loci to phenotypes, but that we now have the tools to distinguish between patterns of genetic drift and natural selection. With 'gene-dropping' simulations, he and his co-workers recently showed that the frequency and excess of heterozygotes observed in Soay sheep cannot be explained by drift alone and that selection plays a role in shaping coat pattern (Gratten *et al.* 2012). That NGS is a rapid and effective way to identify candidate genes was demonstrated by Aurelie Bonin (LECA, Université Joseph Fourier). They combined NGS approaches (i.e. genome scans, transcriptomics) with admixture mapping and QTL analyses to narrow down the list of candidate genes for insecticide resistance in the yellow fever mosquito *Aedes aegypti*. These genes will now be studied in more detail.

Brent Emerson (IPNA-CSIC, Tenerife) gave a nice overview on the field of ecological interactions, where NGS is used to determine resource diversity. For example, diet analyses can now be performed by sequencing gut contents or faecal matter and, in some cases, even complete focal individuals. Several specialized techniques can be applied, such as sequencing distinct diagnostic fragments (with the aid of e.g. chloroplast- or invertebrate-specific primers, if the consumer is a vertebrate, or the usage of species-specific blocking primers). Furthermore, Graham Stone (University of Edinburgh) showed that besides prey species, also symbionts can reliably be recovered and identified using NGS techniques.

Limitations of NGS

Next-generation sequencing success stories, like the ones presented during the *Molecular Ecology* symposium, have the potential of giving a false impression about the presumed ease of applying the technique and/or analysing NGS data. Fortunately, most speakers also commented on some of the difficulties they encountered when applying NGS. Limitations of NGS techniques are widely accepted and discussed in the literature (see e.g. Ekblom & Galindo 2011; Harrison 2012). Besides the well-known issues such as short read lengths, occasional poor read quality, the sheer amount of data to be managed and analysed and/or the lacking user-friendliness of analytic tools, several other problems were emphasized by the speakers.

One major problem encountered in many NGS studies is the high percentage of nonannotated and/or un-mapped loci, which is especially the case with nonmodel organisms lacking reference genomes and other such resources. Often, a researcher can only speculate about functions of the discovered genes of effect or use indirect evidence from other organisms. Clearly, large-scale functional validation experiments and better comparative tools are needed to recover better annotations for nonmodel organisms.

Another widely discussed topic was brought forward by Alex Buerkle. His simulations on sequencing depth vs. allele frequencies estimates showed that the latter are easily biased, despite adequate coverage. He argued that for many population genetic studies, a coverage of 1× is sufficient, as the individual genotype does not play that much of a role in most analyses. The great advantage of a 1× approach is that many more individuals can be included in a study, reducing overall costs tremendously. It is needless to say that there are many other situations, where one would aim for high coverage. As Bryan Carstens (Louisiana State University) outlined in his talk, higher coverage does give a higher confidence when the data is, for instance, used for *de novo* assembly and SNP calling.

Bryan Carstens further pointed out that NGS is a great tool for phylogeographic studies, but that analyses should be performed more rigorously. He proposed a probabilistic method for model selection and suggested that an averaged approach of parameter estimates in relative proportions to the probabilities per model should be used.

Although the examples so far focussed on functional validation, study design and analytic methods, several speakers called for refined theories. Louis Bernatchez (Université Laval), for example, emphasized the need for an extended theory on the evolutionary causes and consequences of the molecular complexity that links the genotype not only to the phenotype, but also to phenotypic plasticity and nongenetic inheritance. In general, a more holistic approach would be necessary, said Louis.

Future directions of the field

Next-generation sequencing obviously revolutionized the field (although, in our view, the number of publications seems to somewhat lack behind the large number of people

who have submitted samples for NGS or already obtained such data), and there is no reason to believe that method development will decelerate anytime soon. This leads to the question: where will we go from here?

The revolution in genotyping technologies, from isozymes and AFLPs to deep sequencing, is probably the best example for the advances that have been made in our field over the last twenty years, said Loren Rieseberg (University of British Columbia). Analytical methods developed and improved alongside the rise of new experimental techniques. And so did the journal *Molecular Ecology*.

Only four issues (258 pages in total) were published in its first year of existence, 1992, compared to over 5000 pages in the twenty-four issues since 2007 (see also Fig. 1). But we are not there yet, concluded Loren, as the gap between molecular biology and ecology is still substantial. We have learned more about evolution (and lately also ecology) by studying molecules. Now, it is time to increase our efforts of studying ecology as a way to increase our knowledge on the function of molecules. These goals can only be reached if we improve the integration of disciplines and methods, another important conclusion of the symposium put forward by several speakers.

The revolution of sequencing technology will continue, and we are on the doorstep of single-molecule sequencing or 'third-generation sequencing'. This method of sequencing single strands of DNA without prior amplification has great potential for the field in general and for population genetics/genomics in particular. Single-strand sequencing directly produces phased haplotypes allowing more accurate estimations of population genetic parameters and the determination of recombination rates and recombination breakpoints. The technique still needs to improve and to become more efficient, especially with respect to error rates that currently counterbalance the ultra-long reads produced, which would, in theory, simplify (*de novo*) assembly, shorten sequencing times and further reduce costs (see e.g. Schadt *et al*. 2010).

Technological and analytical advances are beginning to change the way in which ecology is studied. With available sequencing methods, it is already possible to genetically characterize ecologically divergent populations in detail (see e.g. Roesti *et al*. 2012). Today's multilocus data sets will soon be replaced by whole genome population samples (see e.g. Jones *et al*. 2012), making it possible to link, at a large scale, alleles throughout the genome to particular phenotypes, geographic patterns and ecological parameters. Importantly, new sequencing techniques will become more and more applicable to nonmodel organisms, which are the main targets of interest in ecology. Thus, in the field of molecular ecology, we can shortly start focussing on the effect of ecology on a single individual ('s genome) instead of studying alleles at a single locus or few loci only; and we can do so in nonmodel organisms and, hence, across a large range of taxa.

Furthermore, technical advances will open the opportunity to study topics that have stayed somewhat untouched or isolated until now. In the next twenty years, we should, for example, focus on epigenomics and plasticity in relation to phenotype and/or genotype. Another challenge is the development of toolkits to study a variety of organisms. As discussed above, elucidating the link between genotype and phenotype is very difficult, if not impossible without proper genome annotation. The integration of ecological metadata with genomic data sets is another challenge, just like the development of better analytical tools for comparative population genomics, which will further support the development of new subfields like landscape genomics.

In conclusion, the field has challenging but very excited times ahead, and together with the ongoing revolution of technological, analytical and methodological tools, it will stay an exciting field for the next twenty years. We fully agree with Loren Rieseberg that the future of the field is bright indeed!

The videos and slides from the symposium as well as the Online Forum can be found on: www.molecularecologist.com/2012/10/molecular-ecology-online-forum-2012/

References

Aguileta G, Lengelle J, Marthey S *et al*. (2010) Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens. *Molecular Ecology*, **19**, 292–306.

Ekblom R, Galindo J (2011) Application of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Gompert Z, Lucas LK, Nice CC, Fordyce JA, Forister ML, Buerkle CA (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Gratten J, Pilkington JG, Brown EA, Clutton-brock TH, Pemberton JM, Slate J (2012) Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Molecular Ecology*, **21**, 2977–2990.

Harrison RJ (2012) Understanding genetic variation and function-the application of next generation sequencing. *Seminars in Cell & Developmental Biology*, **23**, 230–236.

Jones FC, Grabherr M, Chan YF *et al*. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB (2012) Rapid De Novo Evolution of X Chromosome Dosage Compensation in *Silene latifolia*, a Plant with Young Sex Chromosomes. *PLoS Biology*, **10**, 4.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. *Human Molecular Genetics*, **19**, R227–R240.

Eveline Diepeveen is a PhD student in the group of Walter Salzburger and uses molecular genetic and genomic approaches to study the genetic basis of naturally and sexually

selected traits in cichlid fishes with special interest in the selective forces acting on genes involved. Walter Salzburger is Professor of Zoology and Evolutionary Biology at the University of Basel. The research of his team focuses on the genetic basis of adaptation, evolutionary innovation and animal diversification. The main model systems in the laboratory are three spine stickleback fish, Antarctic notothenioids and the exceptionally diverse assemblages of cichlid fishes.

doi: 10.1111/mec.12093

CHAPTER 6

THE ECOLOGICAL AND GENETIC BASIS OF CONVERGENT THICK-LIPPED PHENOTYPES IN CICHLID FISHES

ARTICLE

MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 670–684 doi: 10.1111/mec.12029

The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes

MARCO COLOMBO,*[1] EVELINE T. DIEPEVEEN,*[1] MORITZ MUSCHICK,*‡ M. EMILIA SANTOS,* ADRIAN INDERMAUR,* NICOLAS BOILEAU,* MARTA BARLUENGA† and WALTER SALZBURGER*

Zoological Institute, University of Basel, Vesalgasse 1, 4051, Basel, Switzerland, †Museo Nacional de Ciencias Naturales, CSIC, José Gutierrez Abascal 2, 28006, Madrid, Spain

Abstract

The evolution of convergent phenotypes is one of the most interesting outcomes of replicate adaptive radiations. Remarkable cases of convergence involve the thick-lipped phenotype found across cichlid species flocks in the East African Great Lakes. Unlike most other convergent forms in cichlids, which are restricted to East Africa, the thick-lipped phenotype also occurs elsewhere, for example in the Central American Midas Cichlid assemblage. Here, we use an ecological genomic approach to study the function, the evolution and the genetic basis of this phenotype in two independent cichlid adaptive radiations on two continents. We applied phylogenetic, demographic, geometric morphometric and stomach content analyses to an African (*Lobochilotes labiatus*) and a Central American (*Amphilophus labiatus*) thick-lipped species. We found that similar morphological adaptations occur in both thick-lipped species and that the 'fleshy' lips are associated with hard-shelled prey in the form of molluscs and invertebrates. We then used comparative Illumina RNA sequencing of thick vs. normal lip tissue in East African cichlids and identified a set of 141 candidate genes that appear to be involved in the morphogenesis of this trait. A more detailed analysis of six of these genes led to three strong candidates: *Actb*, *Cldn7* and *Copb*. The function of these genes can be linked to the loose connective tissue constituting the fleshy lips. Similar trends in gene expression between African and Central American thick-lipped species appear to indicate that an overlapping set of genes was independently recruited to build this particular phenotype in both lineages.

Keywords: adaptive radiation, cichlid species flocks, convergent evolution, East Africa, ecological genomics, RNAseq

Received 9 March 2012; revision received 4 July 2012; accepted 15 July 2012

Introduction

Adaptive radiation is the rapid evolution of an array of species from a common ancestor as a consequence of the emerging species' adaptations to distinct ecological niches (Simpson 1953; Schluter 2000; Gavrilets & Losos 2009). It is typically triggered by ecological opportunity in form of underutilized resources—just as being provided after the colonization of a new habitat, the extinction of antagonists and/or the evolution of a novel trait, which is then termed an evolutionary 'key innovation' (Gavrilets & Vose 2005; Gavrilets & Losos 2009; Losos & Ricklefs 2009; Losos 2010; Yoder *et al.* 2010; Matschiner *et al.* 2011). Whatever the circumstances were that initiated an adaptive radiation, there is always a strong link between adaptively relevant traits and the habitat and/or foraging niche (a 'phenotype–environment correlation'; Schluter 2000). In the most illustrative examples of adaptive radiation, the Darwin's finches on the Galapagos archipelago, the *Anolis* lizards on the

Correspondence: Walter Salzburger, Fax: +41 61 267 0301; E-mail: walter.salzburger@unibas.ch

‡ Present address: Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, S10 2TN, UK.

[1]These authors contributed equally to this work.

Caribbean islands and the cichlid fishes of the East African Great Lakes, this correlation exists between beak-shape and food source (finches), limb morphology and twig diameter (anoles), and the architecture of the mouth and jaw apparatus and foraging mode (cichlids) (Schluter 2000; Butler *et al.* 2007; Grant & Grant 2008; Losos 2009; Salzburger 2009).

An interesting aspect of many adaptive radiations is the frequent occurrence of convergent (or parallel) evolution (Schluter & Nagel 1995; Harmon *et al.* 2005; Arendt & Reznick 2008; Losos 2011; Wake *et al.* 2011). For example, similar ecotype morphs of anoles lizards have evolved independently on different Caribbean islands (Losos *et al.* 1998; Harmon *et al.* 2005; Losos & Ricklefs 2009), benthic–limnetic and lake–stream species pairs of threespine sticklebacks emerged repeatedly in and around postglacial lakes (Rundle *et al.* 2000; Berner *et al.* 2010; Roesti *et al.* 2012), and a whole array of convergent forms of cichlid fish emerged between the lakes of East Africa (Kocher *et al.* 1993; Salzburger 2009). Such instances of convergent evolution are generally interpreted as the result of the action of similar selection regimes in isolated settings (Schluter & Nagel 1995; Rundle *et al.* 2000; Nosil *et al.* 2002; Harmon *et al.* 2005; Losos 2011). It has further been suggested that if radiations are truly replicated (i.e. driven by adaptive processes), convergence in morphology should tightly be associated with convergence in ecology and behaviour (Johnson *et al.* 2009).

The species flocks of cichlid fishes in the East African Great Lakes Victoria, Malawi and Tanganyika represent the most species-rich extant adaptive radiations in vertebrates (Kocher 2004; Seehausen 2006; Salzburger 2009). Several hundreds of endemic cichlid species have emerged in each lake within a period of several millions of years (as is the case for Lake Tanganyika; Salzburger *et al.* 2002; Genner *et al.* 2007) to <150 000 years (as in Lake Victoria; Verheyen *et al.* 2003). The various endemic cichlid species differ greatly in the morphology of the trophic apparatus (mouth form and shape, jaw structure and dentition) as well as in coloration and pigmentation, suggesting that both natural and sexual selection are jointly responsible for adaptive radiation and explosive speciation in cichlids (Salzburger 2009). Interestingly, convergent forms that emerged in independent cichlid adaptive radiations often show very similar coloration patterns in addition to matching body shapes and mouth morphologies (Kocher *et al.* 1993; Stiassny & Meyer 1999; Salzburger 2009). This has led to speculations whether selection alone is sufficient to explain convergence, or whether genetic or developmental constraints have contributed to the morphogenesis of these matching phenotypes (Brakefield 2006).

The present study focuses on the morphology, ecology and the genetic basis of a peculiar mouth trait in cichlid fishes, which has evolved multiple times: hypertrophied ('fleshy') lips (see Box 1 in Salzburger 2009). The exact function of the thick lips in cichlids is unknown, although this feature is generally implicated in a specific foraging mode (Fryer 1959; Fryer & Iles 1972; Arnegard *et al.* 2001). Fleshy lips are often interpreted as an adaptation for feeding on invertebrates and crustaceans hidden in crannies, with the lips being used to seal cracks and grooves to facilitate the sucking of prey (Barlow & Munsey 1976; Ribbink *et al.* 1983; Seehausen 1996; Konings 1998). Alternatively, it has been suggested that hypertrophied lips protect from mechanical shocks (Greenwood 1974; Yamaoka 1997), and that they function as taste receptors (Arnegard *et al.* 2001) or as mechanoreceptors (Fryer 1959; Fryer & Iles 1972). [Note, however, that there is no increase in sensory cells in lip tissue (Greenwood 1974).]

It is remarkable that thick-lipped species appear to be a common outcome of cichlid adaptive radiations. For example, the large cichlid assemblages in East Africa all contain at least one such taxon (Lake Victoria: *Haplochromis chilotes*; Lake Malawi: *Chilotilapia euchilus, Abactochromis labrosus*, *Otopharynx pachycheilus, Placidochromis milomo*, *Protomelas ornatus*; Lake Tanganyika: *Lobochilotes labiatus*). In addition, cichlids featuring hypertrophied lips are known from, for example, the Midas Cichlid (*Amphilophus* spp.) assemblage in the large lakes of Nicaragua, where a thick-lipped species (*A. labiatus*) is common in rocky habitats (Fig. 1). Occasionally, hypertrophied lips are also observed in other related cichlids in Nicaragua, such as in the riverine species *Tomacichla tuba* (Villa 1982) or in *Astatheros rostratus* (pers. obs.). Additional riverine representatives with hypertrophied lips are also found in South America (*Crenicichla tendybaguassu*) and Western Africa (*Thoracochromis albolabris*). Hypertrophied lips are not unique to cichlids, though. For example, the adaptive radiation of the sailfin silverside fish (Telmatherinidae) in the Malili lakes of Sulawesi (Herder *et al.* 2006) and the barbs of Lake Tana in Ethiopia (Sibbing *et al.* 1998; de Graaf *et al.* 2008) also produced thick-lipped species.

Members of the family Cichlidae are distributed in the Southern hemisphere, with a few ancestral lineages in India, Sri Lanka and Madagascar and two exceptionally species-rich clades, one in Central and South America and one in Africa (Salzburger & Meyer 2004). This biogeographical pattern is consistent with a Gondwanan origin of the Cichlidae, dating the split between American and African representatives to ~100 Ma (Salzburger & Meyer 2004; Sereno *et al.* 2004; Genner *et al.* 2007). This set-up opens the possibility to study the ecological and genetic basis of a convergent trait across one of the

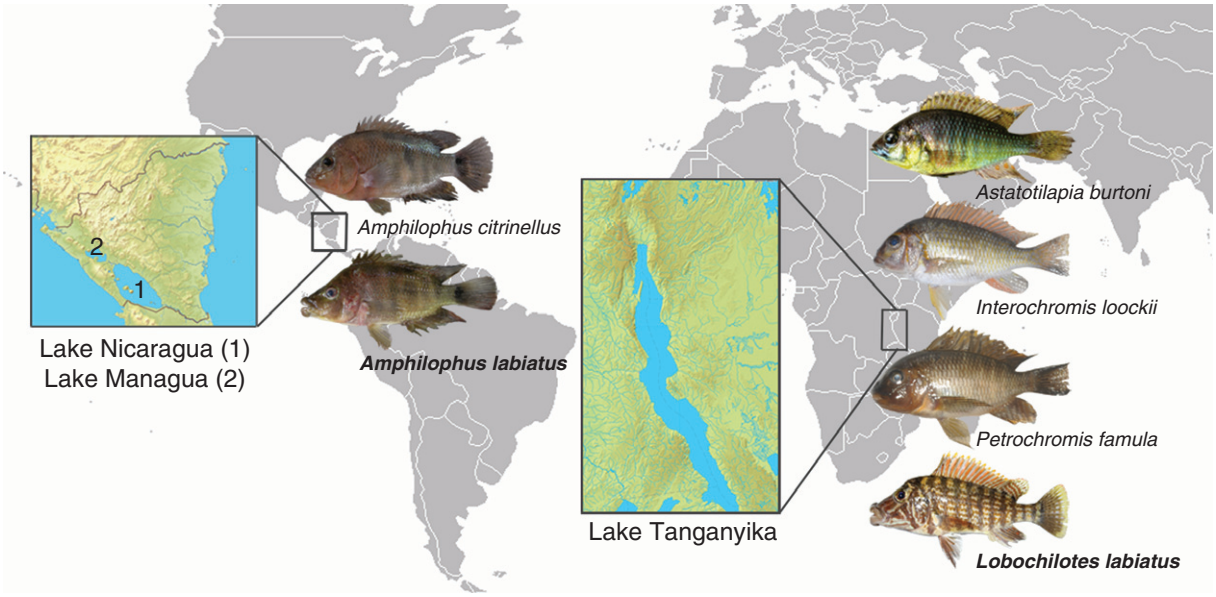


Fig. 1 Map of the Southern hemisphere showing the two study systems, the Midas Cichlid (*Amphilophus* sp.) species complex in Nicaragua, Central America, and the Tropheini in Lake Tanganyika, East Africa.

largest possible phylogenetic and geographical distances in cichlids and, hence, in the complete absence of gene flow and outside the influence of ancestral polymorphism and/or standing genetic variation.

Here, we applied an integrative approach in two cichlid fish radiations, the one of the Tropheini in East African Lake Tanganyika and the Midas Cichlid assemblage in Nicaragua, to uncover the ecological and genetic basis of the thick-lipped phenotype. More specifically, we compared the two '*labiatus*' species to one another and to their sister species by means of geometric morphometric and stomach content analyses; we placed them in their respective radiations by phylogenetic and demographic analyses; and we provide field observations on foraging strategies for one of them (*L. labiatus*). To study the genetic basis of hypertrophied lips, we first applied comparative transcriptome analyses (RNA-seq) on the basis of Illumina next-generation sequencing of juvenile and adult individuals of the African species *L. labiatus* (in comparison with a closely related species for which a genome sequence is available). In a second step, we tested candidate genes identified by RNAseq in representatives of both radiations in a quantitative real-time PCR environment.

Materials and methods

Study species

This study focuses on two thick-lipped species, *Lobochilotes labiatus* from East African Lake Tanganyika and *Amphilophus labiatus* from Nicaragua. *Lobochilotes labiatus* is a member of the rock-dwelling Tanganyikan cichlid tribe Tropheini and therefore part of the most species-rich group of cichlids, the haplochromines, which include the Tanganyikan Tropheini, many riverine species and the species flocks of Lakes Victoria and Malawi (Salzburger *et al.* 2002, 2005). The Tropheini themselves underwent a subradiation within Lake Tanganyika (see e.g. Sturmbauer *et al.* 2003). *Amphilophus labiatus* is part of the Midas Cichlid assemblage in Nicaragua and occurs in the large Central American lakes Managua and Nicaragua, where it co-occurs with the most common species in the area, *A. citrinellus* (Barlow 1976; Barluenga & Meyer 2010). For this study, we sampled a total of 84 and 74 specimens of the Central American species *Amphilophus citrinellus* and *A. labiatus*, respectively, and 143 specimens of *L. labiatus* plus 14 additional Haplochromini/Tropheini specimens from Lake Tanganyika. Exact sampling locations and dates for specimens used for the genetic analysis and GenBank accession numbers are provided in Appendix S1.

Sampling, DNA and RNA extraction

Sampling of *L. labiatus* and other Tropheini species was performed between 2007 and 2011 in the Southern part of Lake Tanganyika, East Africa; *A. labiatus* and its congeners were collected in September 2009 in the two large Nicaraguan lakes Managua and Nicaragua (see Appendix S1 for details). Fishes were processed in the field following our standard operating procedure: fishes were individually labelled, measured (total and standard length) and weighted and a photograph was taken from the left side

of each specimen using a Nikon P5000 or a Nikon D5000 digital camera (fins were spread out using clips); then, a piece of muscle tissue and a fin-clip were taken as DNA sample and preserved in ethanol; fishes were then dissected and RNA samples from lip and other tissues were preserved in RNAlater (Ambion); the whole intestinal tract was removed and stored in ethanol.

For DNA extraction, we either applied a high-salt extraction method (Bruford *et al.* 1998) or used a MagnaPure extraction robot (Roche, Switzerland) following the manufacturer's protocol. RNA was extracted according to the Trizol method with either Trizol (Invitrogen) or TRI reagent (Sigma). Lip tissue was homogenized with a PRO200 Homogenizer (PRO Scientific Inc.) or with a BeadBeater (FastPrep-24; MP Biomedicals). DNase treatment following the DNA Free protocol (Ambion) was performed to remove any genomic DNA from the samples. Subsequent reverse transcription was achieved by using the High Capacity RNA-to-cDNA kit (Applied Biosystems). For the *A. burtoni* samples, up to two individuals (adults) or up to eight individuals (juveniles) were used per sample, due to a diminutive amount of lip tissue extracted from these fishes. All other samples were taken from a single specimen.

Phylogenetic and demographic analyses

We first wanted to phylogenetically place the thick-lipped species into the respective clade of East African and Nicaraguan cichlids. We thus performed a phylogenetic analysis of the Tanganyikan cichlid tribe Tropheini (see also Sturmbauer *et al.* 2003) and used haplotype genealogies to reconstruct the evolutionary history in the much younger *Amphilophus* species assemblage in Nicaragua, where phylogenetic analyses are not expedient due to the lack of phylogenetic signal (see also Barluenga *et al.* 2006; Barluenga & Meyer 2010). We also performed mismatch analyses within *A. citrinellus*, *A. labiatus* and *L. labiatus* to compare their demographic histories.

We amplified three gene segments for each of the three focal species and additional Tropheini/Haplochromini species: the first segment of the noncoding mtDNA control region and two nuclear loci containing coding and noncoding DNA (a segment each of the *endothelin receptor 1*, *ednrb1* and the *phosphatidin phosphatase 1*, *phpt1*). We used previously published primers L-Pro-F (Meyer *et al.* 1994) and TDK-D (Lee *et al.* 1995) for the control region and ednrb1F and ednrb1R (Lang *et al.* 2006) for *ednrb1*, and so far unpublished primers 38a_F (5′-AGC AGG GTT GAC CTT CTC AA-3′) and 38a_R (5′-TGG CTA AAA TCC CCG ATG TA-3′) for *phpt1*. Polymerase chain reaction (PCR) amplification, purification and cycle sequencing were performed as described elsewhere (Diepeveen & Salzburger 2011); an ABI 3130*xl* capillary genetic analyzer (Applied Biosystems) was used for DNA sequencing.

The resulting sequences were complemented with already available sequences. In the case of the Tropheini, we also included available sequences of the mitochondrial NADH dehydrogenase subunit 2 gene (ND2) (see Appendix S1 for GenBank accession numbers). Sequences were aligned with MAFFT (Katoh & Toh 2008) resulting in a total length of 2345 bp for the Tropheini (control region: 371 bp; ND2: 1047 bp; ednrb1: 538 bp; *phpt1:* 389 bp) and 1620 bp for *Amphilophus* (control region: 371 bp; ednrb1: 743 bp; *phpt1:* 469 bp). Maximum-likelihood and Bayesian inference phylogenetic analyses of the Tropheini were performed for each gene segment separately (not shown) and for a concatenated alignment with PAUP* (Swofford 2003) and MRBAYES (Ronquist & Huelsenbeck 2003), respectively. The appropriate model of sequence evolution was detected with JMODELTEST (Posada 2008) applying the Akaike Information Criterion (AIC). A maximum-likelihood bootstrap analysis with 100 pseudoreplicates was performed in PAUP*, and MR. BAYES was run for eight million generations with a sample frequency of 100 and a burn-in of 10%. We then used MESQUITE (www.mesquiteproject.org) to map feeding specializations on the resulting maximum-likelihood topology and to reconstruct ancestral character states with parsimony. Data on feeding mode from the Haplochromini/Tropheini species other than *L. labiatus* are based on Brichard (1989), Nori (1997), Yamaoka (1997) and Konings (1998).

Haplotype genealogies for the *Amphilophus* data set were constructed following the method described in the study by Salzburger *et al.* (2011) on the basis of a maximum-likelihood tree and sequences of the mitochondrial control region and the nuclear *ednrb1* gene (*phpt1* was not used here due to the limited number of haplotypes found). Mismatch analyses were performed on the basis of mtDNA sequences with ARLEQUIN 3.0 (Excoffier *et al.* 2005).

Geometric morphometric analyses

In order to test for similarities in overall body shape between the thick-lipped forms from Central America and East Africa, we performed geometric morphometric analyses on the basis of digital images. Body shape was quantified in a set of 58 *A. citrinellus*, 27 *A. labiatus* and 27 *L. labiatus* using 17 homologous landmarks (see Appendix S2; note that lip shape was not assessed to prevent a bias). Data acquisition was carried out using TPSDIG (Rohlf 2006), and data were analysed with MORPHOJ (Klingenberg 2011). For all shape comparisons, we used the residuals of a within-species regression of shape on centroid size to reduce allometric effects within species, in

order to retain shape differences between differently sized species. For the same reason, we only included *L. labiatus* individuals with a body size larger than 12 cm total length. We then performed a discriminant function analysis between all pairs of species and a principal component analysis (PCA). To identify morphological changes associated with the enlarged lip phenotype, we compared *A. labiatus* to its closest relative, *A. citrinellus*. In the case of *L. labiatus*, we made use of our new phylogeny of the Tropheini (Fig. 2a) and body shape data of *L. labiatus* and its nine closest relatives [*Petrochromis macrognathus*, *P. polyodon*, *P. ephippium*, *Lobochilotes labiatus*, *Simochromis diagramma*, *S. babaulti*, *Gnathochromis pfefferi*, *Pseudosimochromis curvifrons*, *Limnotilapia dardenni* and *Ctenochromis horei* (M. Muschick, A. Indermaur & W. Salzburger, unpublished data)] to reconstruct the landmark configuration of the direct ancestor to *L. labiatus*. This was carried out in MorphoJ using branch length-weighted squared-change parsimony. The changes in landmark configurations along a discriminant function (Nicaraguan species) or along the shape-change vector from the estimated ancestral shape to *L. labiatus* were increased threefold to produce Fig. 3. The shape differences between species shown in Fig. 3 accurately reflect the shape-change vectors for landmark positions. Outlines were interpolated and added to Fig. 3 to help the reader envision these shape differences in the context of fish body shape.

Stomach and gut content analyses

To assess trophic specialization of the thick-lipped cichlid species, we performed comparative stomach and gut content analyses. To this end, stomachs and guts were opened step-by-step. First, the stomach was opened and emptied under a binocular followed by the remaining parts of the intestine. All items were grouped into seven food categories: hard-shelled (crustaceans, snails, mussels), small arthropods (insects and zooplankton), fish scales, fish remains, plant seeds and plant material other than seeds. For each specimen, the wet weight of each food category was measured on a Kern ALS 120-4 scale (Kern, Germany) and was then used to calculate Schoener's index of proportional diet overlap (Schoener 1970). We analysed stomach and gut contents in a total of 159 specimens: *A. citrinellus* ($N = 58$; of which 25 had contents), *A. labiatus* ($N = 62$; 34) and *L. labiatus* ($N = 39$; 29). We note that such an analysis has the drawback that it only covers food uptake in the last few hours or days before sampling.

Field observations in Lobochilotes labiatus

The feeding behaviour of *L. labiatus* was observed at our field site near Mpulungu, Zambia, in concrete ponds ($1.5 \times 1.5 \times 1$ m). The purpose of these observations under semi-natural conditions and with wild specimens was to document if and how the lips are used in processing the main prey item identified in the stomach content analyses. The ponds were equipped with stones of ~20–30 cm diameters that covered the ground and formed caves as they occur naturally in the habitat of *L. labiatus*. Each pond was stocked with five to six freshly caught and unharmed adult individuals of *L. labiatus*. After an acclimatization period of at least 4 days, fish were offered snails of different sizes and their feeding behaviour was recorded with two underwater cameras (Canon Ixus 65 with WP-DC3 underwater case; Olympus μ tough-6000) for a period of 1 h each.

Comparative gene expression assays using RNAseq

For the identification of differentially expressed genes in thick-lipped species, we performed RNA sequencing (RNAseq) comparing lip tissue from a thick-lipped species to lip tissue from a reference species. We decided to perform these experiments in the African species *L. labiatus* and to use the closely related species *Astatotilapia burtoni* as reference taxon for several reasons such as the availability of laboratory strains and of sufficient RNA samples from adult and juvenile individuals. Most importantly, we chose this set-up because of the availability of various genomic resources for *A. burtoni*, such as a whole-genome sequence and a set of ~50 000 partly annotated expressed sequence tags (ESTs) (Salzburger *et al.* 2008; Baldo *et al.* 2011), which is crucial for the analysis and interpretation for RNAseq data. Such resources are currently not publicly available for *Amphilophus*.

In a first step, RNA was extracted from adult and juvenile individuals of *L. labiatus* and *A. burtoni* (see above for the RNA extraction protocol). RNA quality and quantity were determined on a NanoDrop 1000 spectrophotometer (Thermo Scientific) and by gel electrophoresis. RNA samples were pooled to create four samples subjected to RNA sequencing (RNAseq): (i) *A. burtoni* adult ($N = 3$); (ii) *A. burtoni* juvenile ($N = 1$); (iii) *L. labiatus* adult ($N = 2$); and (iv) *L. labiatus* juvenile ($N = 3$). Five micrograms of RNA per RNAseq sample was sent for Illumina sequencing at the Department of Biosystems Science and Engineering (D-BSSE), University of Basel and ETH Zurich. For library construction and sequencing, standard protocols were applied. Poly-A mRNA was selected using poly-T oligo-attached magnetic beads. The recovered mRNA was fragmented into smaller pieces using divalent cations under increased temperature. cDNA was produced using reverse transcriptase and random primers, followed by second-strand cDNA synthesis using DNA polymerase

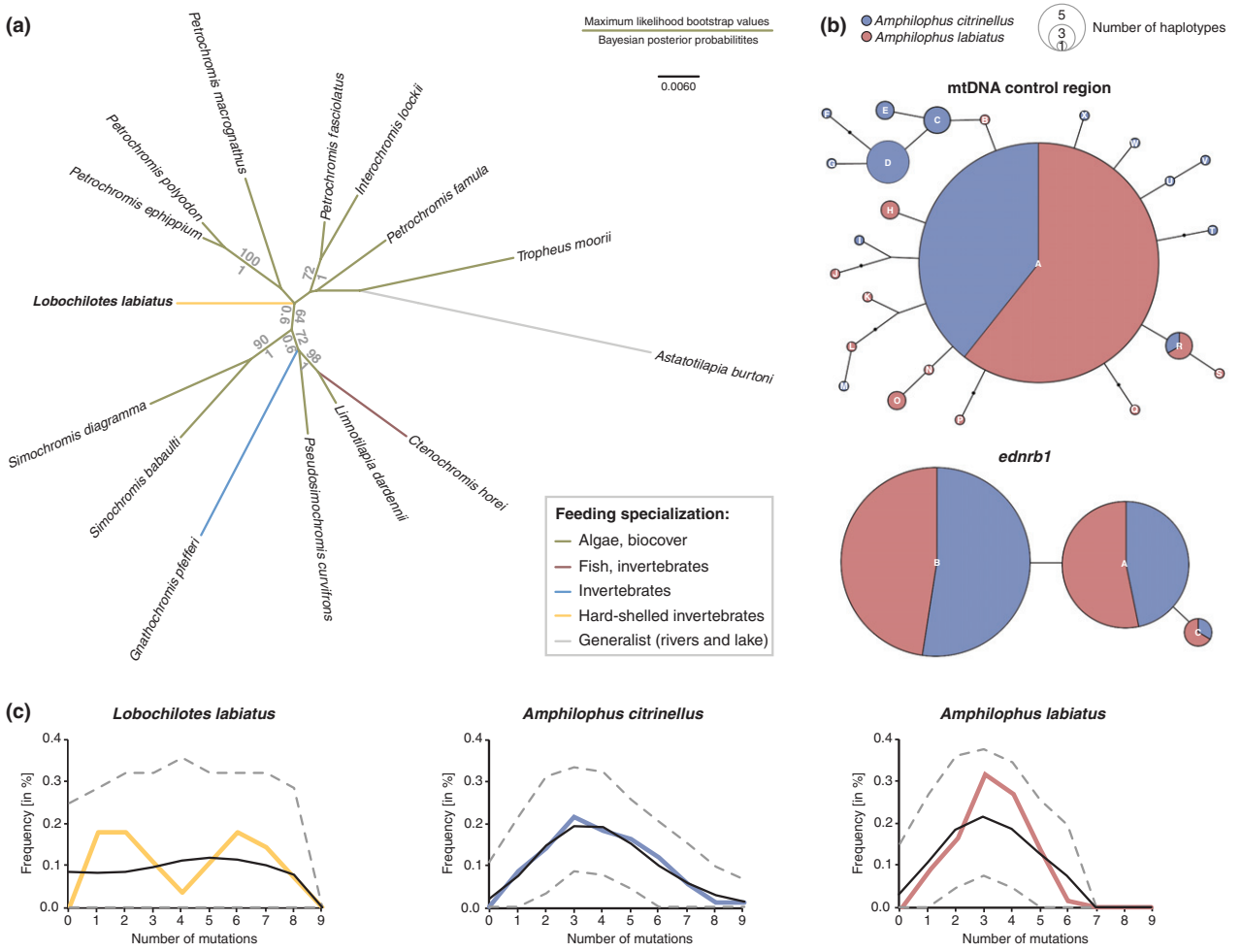


Fig. 2 Evolutionary origin of the thick-lipped species in East African Lake Tanganyika and in the Great Lakes of Nicaragua. (a) Maximum-likelihood tree of the Tropheini from Lake Tanganyika based on two mitochondrial (control region and ND2) and two nuclear (*ednrb1* and *phpt1*) gene segments (2345 bp in total) and the GTR+G+I model of molecular evolution. Numbers above the branches refer to maximum-likelihood bootstrap values, and numbers below are Bayesian posterior probabilities (note that support values are only shown for branches with bootstrap values >60). Branches are colour-coded according to feeding specializations; the trait values for internal branches have been reconstructed with MESQUITE. (b) Haplotype genealogies of the two *Amphilophus* species based on the mitochondrial control region and the nuclear *endrb1* gene. A large fraction of the haplotypes is shared between *A. citrinellus* and *A. labiatus*. (c) Results from the mismatch analysis on the basis of the mitochondrial control region showing the inferred demographic histories for *L. labiatus*, *A. citrinellus* and *A. labiatus*. Coloured lines represent the observed data, the black line indicates the best-fit model, and the dashed lines in grey indicate the upper and lower boundaries from the simulations in ARLEQUIN.

I and RNaseH. cDNA went through an end-repair process, the addition of a single 'A' base and ligation of the adapters. It was then purified and enriched with PCR to create the final cDNA library. Each library was sequenced in one lane on an Illumina Genome Analyzer IIx (read length was 76 bp). Illumina reads are available from the Sequence Read Archive (SRA) at NCBI under the accession number SRA052992.

The Illumina reads were assembled into three different data sets for further analyses: (i) a quality-filtered data set (Data set 1), where the quality of the reads was assessed with the FASTX toolkit tools implemented in GALAXY [version September/October 2011; available at http://main.g2.bx.psu.edu/ (Giardine *et al.* 2005; Blankenberg *et al.* 2010; Goecks *et al.* 2010)]; low-quality reads were discarded applying quality filter cut-off values of 22–33. (ii) a quality-filtered plus trimmed data set (Data set 2), in which all the reads were trimmed to a length of 42 bp to evaluate the effects of read length (iii) as a control for the effect of trimming and filtering, a nonquality-filtered, nontrimmed data set (Data set 3).

The reads of the three data sets were then aligned to a reference cichlid assembly (Baldo *et al.* 2011) with NOVOALIGN 2.07.06 (http://www.novocraft.com/) after indexing the reference sequences with NOVOINDEX (http://www.novocraft.com/) using default parame-

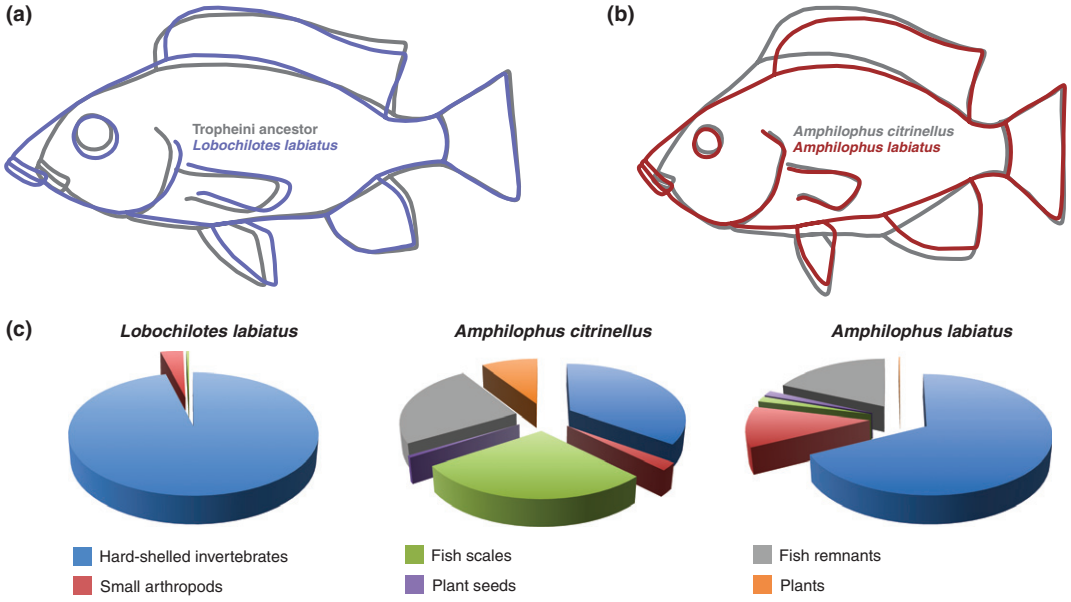


Fig. 3 Ecomorphology of the thick-lipped cichlid species in Central America and in Lake Tanganyika. (a) Body shape of *L. labiatus* in comparison with a reconstruction of the ancestor of *L. labiatus* and nine closely related Tropheini species. (b) Differences in body shape between *A. citrinellus* and *A. labiatus* along a discriminate function. In both plots, changes in landmark positions were increased threefold and interpolated outlines added for illustration purposes. Landmark locations are indicated in black on the reconstructed outlines in plot (a). (c) Analysis of stomach and gut content in the focal species. The fraction of each food category is shown.

ters. The alignment was performed using default settings with a maximum alignment score (t) of 180 and a maximum number of alignments for a single read (e) of 100; reads with multiple alignment locations were discarded. Next SAMTOOLS version 0.1.18 (Li *et al.* 2009) was used to sort and index the files and to generate count files, which were subsequently transformed into count tables and analysed in the R package DESEQ version 1.0.5 (Anders & Huber 2010). Differentially expressed genes between the four experimental groups were detected using a model based on a negative binomial distribution implemented in DESEQ. Differentially expressed genes with *P*-values (adjusted for multiple testing) >0.05 and/or a quotient of variance >1.00 were discarded to reduce the number of false positives. The remaining differentially expressed genes of all pairwise comparisons were tested for multiple hits. Next the hits of the three data sets were compared with each other to create a candidate gene list, consisting of genes that were found in multiple analyses in all three data sets. Lastly, these hits were compared to the annotated *A. burtoni* ESTs of Baldo *et al.* (2011).

Comparative gene expression assays using quantitative real-time PCR

Based on their function according to gene ontology terms (GO terms; http://www.geneontology.org/) and their putative involvement in lip formation and/or hypertrophy in other organisms, six candidate genes were selected out of the list of differentially expressed genes for further characterization by means of quantitative real-time PCR (qPCR). These candidate genes are the *Bcl2 adenovirus e1b 19-kda protein-interacting protein 3* (*BNIP3*), *long-chain-fatty-acid(CoA)-ligase 4* (*ACSL4*), *histone 3.3* (*His3*), *beta actin* (*Actb*), *coatomer subunit beta* (*Copb*) and *claudin 7* (*Cldn7*; see Table 1 for primer details). qPCR experiments were performed in total of 36 cichlid specimens: *L. labiatus* (six adults, six juveniles), *A. burtoni* (six adults, six juveniles), *A. labiatus* (six adults) and *A. citrinellus* (six adults). By performing two pairwise comparisons between a thick-lipped and a normal-lipped species (a species pair each from Africa and Nicaragua), we effectively control for species-specific expression differences, as genes specific to thick-lip tissue should be upregulated in both comparisons.

The experiments were conducted on a StepOnePlus Real-Time PCR system (Applied Biosystems) as described elsewhere (Diepeveen & Salzburger 2011) using the *elongation factor 1* (*EF1*) and the *ribosomal protein SA3* (*RpSA3*) as endogenous controls. Average relative quantifications (RQ) were calculated for the six experimental groups and subsequently analysed with a two-tailed unpaired t-test using GRAPHPAD PRISM version 5.0a for Mac OS X (www.graphpad.com). We compared the expression levels between the two thick-lipped species and a closely related normally lipped species (i.e. *L. labiatus* vs. *A. burtoni* and *A. labiatus* vs. *A. citrinellus*). We also compared adults vs.

Table 1 Primers used for the quantitative real-time PCR experiments

Locus	Forward (5′–3′)	Reverse (5′–3′)
Actb	CAGGCATCAGGGTGTAATGGTT	CAGGCATCAGGGTGTAATGGTT
Copb	GAGGCTACCTTGGCTGTCAAAG	GTGCTGGATGGTTTGAGGGTAA
His3	CATCTACTGGTGGAGTGAAGAAACC	GGATCTCACGCAGAGCAACA
ACSL4	TGGTTCTGCACCGGAGATG	TCTTGCGGTCAACAATTTGTAGA
BNIP3	AACAGTCCACCAAAGGAGTTCCT	CCTGATGCTGAGAGAGGTTGTG
Cldn7	GACATCATCCGGGCCTTCT	CACCGAACTCATACTTAGTGTTGACA
EF1	GCCCCTGCAGGACGTCTA	CGGCCGACGGGTACAGT
RpSA3	AGACCAATGACCTGAAGGAAGTG	TCTCGATGTCCTTGCCAACA

juveniles in the African species, as hypertrophy in lips is much less pronounced at juvenile stages, so that this experiment also captures ontogenetic changes in lip formation. As primer efficiency was lower in the Nicaraguan samples, no direct comparisons between African and Nicaraguan tissues were possible.

Results

Phylogenetic and demographic analyses

Our phylogenetic analysis of members of the Tanganyikan cichlid tribe Tropheini based on two mitochondrial and two nuclear DNA gene segments reveals only limited phylogenetic resolution between the main lineages of the tribe (Fig. 2a). This confirms an earlier analysis based on mitochondrial DNA only, which attributed the star-like phylogeny of the Tropheini to the rapidity of lineage formation in the early stages of the adaptive radiation of this clade (Sturmbauer *et al.* 2003). Just as in the previous study, the thick-lipped species *L. labiatus* represents a separate lineage (without a closely related sister-taxon) that branches off relatively early in the phylogeny, but shows affinities to the algae-eating genera *Petrochromis* and *Simochromis*.

The haplotype genealogies of the *Amphilophus* samples based on the mitochondrial control region and the nuclear *ednrb1* gene (Fig. 2b) revealed haplotype sharing between *A. citrinellus* and *A. labiatus* (see also Barluenga & Meyer 2010). While all *Amphilophus* sequences were identical in *phpt1*, we detected three shared haplotypes in ednrb1 and 24 haplotypes in the mitochondrial control region (two shared, ten unique to *A. labiatus* and twelve unique to *A. citrinellus*).

The mismatch analyses based on the mitochondrial control region sequences revealed unimodal distributions for the two sympatrically occurring *Amphilophus* species and a bimodal distribution for *L. labiatus* (Fig. 2c). According to this analysis, the demographic expansion of the two *Amphilophus* species happened at similar times, with the one of *A. citrinellus* being slightly older than that of *A. labiatus* (mean number of differences: 3.9 vs. 3.2; τ: 3.9 vs. 3.5; see also Barluenga & Meyer 2010, who provide a relative time frame for the evolution of the Midas Cichlid species complex); the mean number of differences in *L. labiatus* was 6.4 (τ: 6.5).

Geometric morphometric analyses

The PCA of overall body shape revealed substantial overlap between the two Nicaraguan species *A. citrinellus* and *A. labiatus* (Appendix S3). The African thick-lipped species *L. labiatus* is separated from these mainly by principal component 1 (accounting for 20.2% of the variance), whereas principal component 2 (covering 16.0% of the variance) did not discriminate much between species. The discriminant function analysis, in which we compared species in a pairwise manner, revealed the main morphological differences between species. Of the two Nicaraguan species, *A. labiatus* had a more acute head, less deep body and a larger mouth than *A. citrinellus* (Fig. 3) (see also Klingenberg *et al.* 2003). These characters were even more pronounced in *L. labiatus*, when compared to either of the *Amphilophus* species. However, the distance in morphospace between the two species with fleshy lips was somewhat smaller than between *A. citrinellus* and *L. labiatus* (procrustes distance 0.08 and 0.1, respectively). We also estimated the body shape of the ancestor of *L. labiatus* and the 9 most closely related Tropheini species. A comparison of this reconstructed shape and the mean shape of our *L. labiatus* samples highlighted similar morphological differences as the comparison of the Nicaraguan species (Fig. 3), especially in the mouth region.

Stomach and gut content analyses

The fractions of food categories in guts and stomachs differed between *A. citrinellus*, *A. labiatus* and *L. labiatus* (Fig. 3c). While the diet of *A. citrinellus* did not overlap with that of *A. labiatus* (Schoener's index: 0.58) or *L. labiatus* (Schoener's index: 0.38), we found significant overlap between the two thick-lipped species *A. labiatus* and *L. labiatus* (Schoener's index: 0.71) (note that any value >0.6 is considered 'biologically significant'; see Wallace 1981). The stomach and gut contents of both

thick-lipped species consisted of a substantial fraction of hard-shelled prey (*Lobochilotes labiatus* 96%, *Amphilophus labiatus* 67.6%, *Amphilophus citrinellus* 35%).

Field observations in Lobochilotes labiatus

A careful inspection of the video material confirmed the findings from the stomach and gut content analyses that *L. labiatus* regularly feeds on snails (more than 90% of the stomach and gut content of *L. labiatus* consisted of snail shells). Small snails were engulfed using suction feeding without the lips touching the prey item or the surface (rocks) on which the items were placed. When feeding on larger snails, however, *L. labiatus* exhibited a different feeding strategy and snails were no longer taken up using suction feeding. Instead, *L. labiatus* used their lips to snatch the snails and they turned the snails a few times before they either swallowed the snails or spat them out (see Appendix S4).

Comparative gene expression assays using RNAseq

On average, ca. 42 million total reads were retrieved for each of the four RNAseq samples (*A. burtoni* adult, *A. burtoni* juvenile, *L. labiatus* adult and *L. labiatus* juvenile). Quality filtering and trimming reduced this number so that on average 21.9 (Data set 1), 24.6 (Data set 2) and 23.5 (Data set 3) million reads were aligned to the reference cichlid assembly. Five different pairwise comparisons were made to obtain genes that are differentially expressed between thick lips and normal lips (see Table 2 for the three comparisons with the highest number of genes being different). The largest number of differentially expressed genes between *L. labiatus* and *A. burtoni* was detected in adult lip tissue, with the majority of the genes being upregulated in *L. labiatus*. The total number of differentially expressed genes ranged from 9050 (Data set 3; three pairwise comparisons) to 15230 (Data set 2; five pairwise comparisons). A substantial fraction of these differentially expressed genes appeared in at least two comparisons in each data set (Data set 1: 2085 [22.1% of all hits]; Data set 2: 8078 [53.0%]; Data set 3: 1693 [18.7%]). Of these 'multiple hits', 1463 were detected in all three data sets and 560 of those could be unequivocally annotated.

A more stringent analysis, in which only loci that appeared in at least three of five comparisons were included, resulted in 231 differentially expressed genes. A functional annotation of these 231 hits with Blast2GO resulted in a total of 141 annotations (122 upregulated and 19 downregulated in *L. labiatus*; see Appendix S3). Based on their annotations, known functions and/or exceptional fold change (>1000) between *A. burtoni* and *L. labiatus*, thirteen genes were identified as good candidates for being involved in the morphogenesis of fleshy lips (Table 3).

Comparative gene expression assays using quantitative real-time PCR

The results of the comparative gene expression assays between the thick-lipped species and the normal-lipped species are depicted in Fig. 4 and Appendix S5. Overall, the qPCR experiments largely validate differential gene expression in normal and hypertrophied lip tissue as indicated by RNAseq. In the African species pair *L. labiatus* and *A. burtoni*, which were the two species used for RNAseq, differences were highly significant in four of the six genes tested: *Actb* ($P = 0.0099$), *Cldn7* ($P = 0.004$), *ACSL4* ($P = 0.0005$) and *His3* ($P = 0.0003$). However, we would like to point out one inconsistency between RNAseq and qPCR. *Actb* was actually found to be downregulated in hypertrophied lips by RNAseq, while it shows significantly higher expression levels in lip tissue in the qPCR experiments (Fig. 4).

The comparison between lip tissue in adult and juvenile *L. labiatus* and *A. burtoni* further revealed a trend towards higher expression in lip tissue of adult *L. labiatus* in *Actb*, *BNIP3*, *Cldn7* and *Copb* (Appendix S5), whereas, generally, an opposite trend is observed in *A. burtoni*, although statistical support was only found in two cases [*Cldn7* ($P = 0.0063$) and *ACSL4* ($P = 0.0328$)]. This again suggests that these genes are involved in the formation of fleshy lips. In the Nicaraguan species pair, a similar trend was observed as in the African species pair, with four of the five genes tested appearing to be upregulated in lip tissue

Table 2 Pairwise comparisons of differentially expressed genes and total number of unique differentially expressed genes in the three data sets compiled in this study

Comparison	Data set 1	Data set 2	Data set 3
AB vs. LL	7120 (4606; 2514)	7080 (4689; 2391)	7285 (4665; 2620)
AB vs. LLjuv	3611 (3395; 216)	13747 (10683; 3064)	2618 (2514; 104)
ABjuv vs. LLjuv	1116 (792; 324)	3971 (2710; 1261)	986 (687; 298)
Total	9407	15225	9050

AB, *Astatotilapia burtoni;* LL, *Lobochilotes labiatus*; juv, juvenile; numbers in brackets denote the number of upregulated and downregulated genes in *L. labiatus*.

Table 3 Thirteen candidate loci for the genetic basis of lip development in the East African cichlid *Lobochilotes labiatus*, based on RNAseq and qPCR in comparison with *Astatotilapia burtoni*, in combination with information on gene functions (in alphabetical order)

Locus	Abbreviation
ATPase mitochondrial precursor	*ATPmp*
Bcl2 adenovirus e1b 19-kda protein-interacting protein 3	*BNIP3*
Beta actin	*Actb*
Caspase-8	*Casp8*
Claudin 7	*Cldn7*
Coatomer subunit beta	*Copb*
Grainyhead-like protein 1 homolog	*Grhl1*
Heat-shock 70-kda protein 12a-like	*Hspa12al*
Histone 3.3	*His3*
Laminin subunit gamma-2	*Lamc2*
Long-chain-fatty-acid(CoA)-ligase 4	*ACSL4*
Sodium-dependent phosphate transporter 1	*Slc17a1*
Transcription factor ap-2 gamma	*Tfap2*

of *A. labiatus* as compared to *A. citrinellus* (Fig. 4; we could not amplify *BNIP3* here). We would like to note, however, that qPCR efficiency was less good in the *Amphilophus* samples, most likely because we used primers designed for the African species pair based on the available genomic resources, which also explains the limited statistical support for these comparisons. Interestingly, it seems that several loci (i.e. *Actb*, *Cldn7*, *Copb*, *His3*) are upregulated in both thick-lipped species when compared to their normally lipped relatives.

Discussion

The species flocks of cichlid fishes in the East African Great Lakes Victoria, Malawi and Tanganyika, counting hundreds of endemic species each, are prime examples of adaptive radiation and explosive speciation (see e.g. Kocher 2004; Seehausen 2006; Salzburger 2009). Interestingly, the cichlid adaptive radiations in East Africa have independently produced ecomorphs with highly similar colour patterns and (mouth) morphologies (Kocher *et al.* 1993). Here, we explore the ecological and genetic basis of one of the particular trophic structures of cichlids, which has evolved convergently in various cichlid assemblages: fleshy lips. Instead of focusing on species with hypertrophied lips between the radiations in the East African lakes, we compare the thick-lipped phenotype between a cichlid assemblage in East African (Lake Tanganyika) and in Central American (the lake Nicaragua/Managua system), where thick-lipped species have evolved in parallel (see Fig. 1).

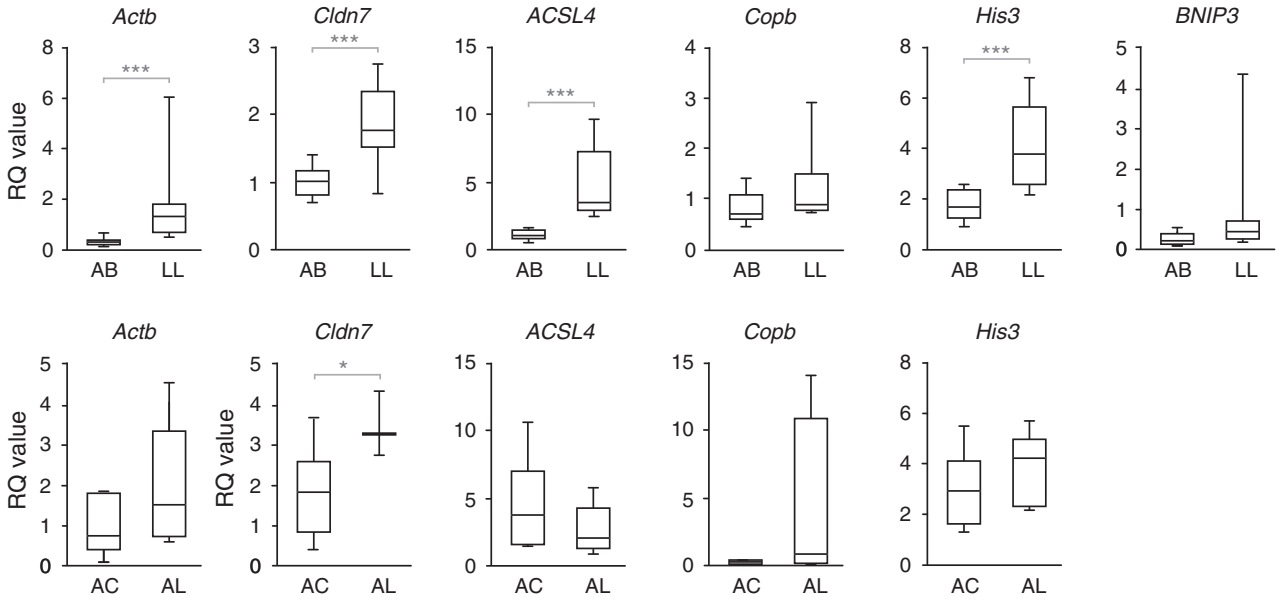


Fig. 4 Results from the comparative gene expression experiments via quantitative real-time PCR. The six genes tested in this experiment were selected on the basis of comparative RNA sequencing. All genes tested show a higher expression level in lip tissue of the Tanganyikan thick-lipped species *L. labiatus* as compared to *A. burtoni* (top panel; note that we used both juvenile and adult samples in these analyses to increase statistical power). A similar trend was found when comparing the Nicaraguan thick-lipped species *A. labiatus* to its sister species *A. citrinellus* (with the exception of *ACSL4*; lower panel). Note that *BNIP3* could not amplified in the *Amphilophus* species. *Astatotilapia burtoni* (AB); *Lobochilotes labiatus* (LL); *Amphilophus citrinellus* (AC); *Amphilophus labiatus* (AL); $*P < 0.05$; $***P < 0.01$.

The evolution of hypertrophied lips in cichlid adaptive radiations

Our phylogenetic and demographic analyses in the Tanganyikan Tropheini and the Nicaragua Midas Cichlid species complex reveal that the thick-lipped species are nested within their respective clade. The molecular phylogeny of 14 Tropheini species (Fig. 2a) shows a footprint characteristic for adaptive radiations: a 'bottom heavy' topology with only limited phylogenetic resolution at the deeper nodes due to rapid lineage formation (Gavrilets & Vose 2005). Our new analysis thus confirms previous results based on mtDNA only (Sturmbauer *et al.* 2003) or a combination of mtDNA and AFLPs (Koblmuller *et al.* 2010). In all analyses thus far, the thick-lipped species *L. labiatus* forms an independent evolutionary lineage that branches off deep in the Tropheini. Its exact position remains unclear, though. In the AFLP phylogeny of Koblmuller *et al.* (2010), *L. labiatus* appears as sister group to all Tropheini except for the genus *Tropheus*, which is sister to all other representatives of that clade (the topology has very little support, though). In our new phylogeny and the previous mtDNA trees of Sturmbauer *et al.* (2003), *L. labiatus* shows affinities to *Simochromis* and *Petrochromis* (with moderate support). In all phylogenies, however, *L. labiatus* is nested within a clade formed by various species that feed on algae and biocover (see our character state reconstruction in Fig. 2a).

In the Midas Cichlid species complex from Central America, a phylogenetic approach is not applicable with the available molecular markers. There is simply too little genetic variation, even in the rapidly evolving mitochondrial control region, as a consequence of the young age of the assemblage (see Barluenga & Meyer 2004, 2010; Barluenga *et al.* 2006). The structures of our haplotype genealogies, which now also include the analysis of a nuclear gene (Fig. 2b), confirm this scenario. In combination with the mismatch analyses (Fig. 2c), these data suggest that *A. labiatus* underwent its main demographic expansion soon after the expansion of the sympatric *A. citrinellus* populations (see Barluenga & Meyer 2010 for a large-scale analysis of the Midas Cichlid species complex).

In both species assemblages, the evolution of the thick-lipped phenotype was associated with similar modifications of overall body shape (Fig. 3a,b). Reduced body depth, a more acute head shape and a larger mouth, along with the prominently enlarged lips, can be hypothesized to be adaptations to the species' microhabitat and trophic niche. If individuals search for food in narrow rock crevices, these modifications appear advantageous. Klingenberg *et al.* (2003) already suggested that the elongation of the head, as observed in both '*labiatus*' species, increases suction power. Other morphological differences between the two thick-lipped species, such as eye size or the length of anal fin insertion, might be either due to adaptations to the specific environments or due to phylogenetic effects. Inclusion of other thick-lipped species in future studies focusing on the ecology and morphological evolution of this trait might answer this question.

The function of hypertrophied lips in cichlids

Hypertrophied lips in cichlids have been implicated in several functions. For example, it has been suggested that fleshy lips are used to seal cracks and grooves to facilitate sucking of invertebrates (Barlow & Munsey 1976; Ribbink *et al.* 1983; Seehausen 1996; Konings 1998), that they act as bumpers to protect from mechanical shock (Greenwood 1974; Yamaoka 1997) or that they function as taste (Arnegard *et al.* 2001) or mechanoreceptors (Fryer 1959; Fryer & Iles 1972). Previous food web analyses on *L. labiatus* identified this species as mollusc eater (Nori 1997).

Our ecomorphological analysis of the thick-lipped species *L. labiatus* from Lake Tanganyika and *A. labiatus* from the large lakes in Nicaragua suggests that this phenotype is indeed associated with feeding on hard-shelled prey such as snails, mussels and crustaceans in rocky habitats (Fig. 3c). We cannot, however, conclusively answer the question whether the lips are used to seal rock crevices or whether they serve as bumpers or receptors. In the underwater observations at our field site at Lake Tanganyika, small snails were usually engulfed by *L. labiatus* via suction feeding, whereas larger snails were turned around several times before being swallowed or spit out (see Appendix S4). This would classify the lips as instrument to handle hard-shelled invertebrate food (mostly molluscs). Note, however, that our observations were made in semi-natural conditions only, in the form of concrete ponds equipped with stones from the lake and filled with lake water.

Our experimental set-up could not address the possibility that phenotypic plasticity plays a role in the formation of fleshy lips, as has previously been shown in certain foraging traits in cichlid fishes (oral jaws: Meyer 1987; pharyngeal jaws: e.g. Greenwood 1965; Huysseune 1995; Muschick *et al.* 2011). Interestingly, it has been reported that thick-lipped cichlid species lose their fleshy lips under unnatural conditions in captivity (when fed with standard food; Barlow & Munsey 1976; Barlow 1976; Loiselle 1998). So far, there is no evidence for the opposite process, the plastic development of fleshy lips due to environmental or feeding properties. In the common garden experiment of Muschick *et al.*

(2011), one group of normally lipped *A. citrinellus* individuals was fed with whole snails over a period of several months, and—although not formally assessed—no increase in lip size was apparent (compared to the other two treatment groups peeled snails and crushed snails). Another study on a snail crusher (Huysseune 1995) did not report such changes either, which seems to suggest that phenotypic plasticity in the lips, if at all present, is specific to thick-lipped species only. Future common garden and feeding experiments should thus expand on this question. Such experiments, combined with molecular analyses, should focus on the plastic component of this trait and its genomic basis.

Insights into the genetic basis of hypertrophied lips in cichlids

Our comparative gene expression assays with RNA sequencing between tissue from thick and normal lips identified a set of 141 candidate genes that might be responsible for the morphogenesis or the maintenance of fleshy lips in (East African) cichlid fish (Appendix S3). Six genes were tested further by means of quantitative real-time PCR, and these experiments largely confirm the results obtained from RNAseq (Fig. 4). While there is no obvious functional connection to fleshy lips for three of these differentially expressed genes (*ACSL4*, *His3* and *BNIP3*), the observed upregulation of the remaining three (*Actb*, *Cldn7* and *Copb*) makes sense in the light of the structure of hypertrophied lips. These three genes (together with *BNIP3*) also show a higher expression in lip tissue from adult vs. juvenile *L. labiatus* (Appendix S5).

It has previously been shown that the 'fleshy' lips of the Lake Malawi cichlid *Otopharynx pachycheilus* mainly consist of loose connective tissue covered by dermis and a layer of epithelial cells (Arnegard *et al.* 2001). Interestingly, the known functions of *Actb*, *Cldn7* and *Copb* can be directly implicated in cell and/or intercell or membrane structure. The cytoplasmic *Actb* is found in high abundance in nonmuscle cells, where it promotes cell surface and cell thickness (Schevzov *et al.* 1992), which is also consistent with its upregulation in the more massive adult compared to juvenile *L. labiatus* lips (Appendix S5). The integral membrane protein *Cldn7* (among other *claudin* gene family members) constitutes the backbone of tide junctions between epithelial cells (Tsukita *et al.* 2001). The coatomer coat proteins (such as *Copb*) are involved in protein and membrane trafficking via vesicle secreting between the endoplasmic reticulum and the Golgi apparatus, plus the intra-Golgi transport (Duden 2003). In addition, they mediate lipid homoeostasis and lipid storage for energy use and membrane assembly (Soni *et al.* 2009). *Copb* might thus be involved in cellular (membrane) development but possibly also in the formation of fat cells that compose adipose tissue, a specific subtype of connective tissue. Clearly, much more work will be necessary to unravel the development and genetic basis of hypertrophied lips in cichlids, for which we herewith established a valuable starting ground.

Our results, especially the comparison of gene expression levels between the thick-lipped species in East Africa and Central America (Fig. 4), allow us to touch on ongoing discussions related to the genetic basis of convergent morphologies (reviewed in Brakefield 2006; Arendt & Reznick 2008; Elmer & Meyer 2011). Although our qPCR results in Midas Cichlid (*Amphilophus* spp.) species must be taken with caution (efficiency was lower as a consequence of using molecular tools developed for the African species leading to a lack of statistical power), we find rather similar trends in gene expression. Our results seem to indicate that a largely overlapping set of genes was recruited to develop the hypertrophied lips in Nicaraguan and African species, which are—according to most authors—separated by ~ 100 million years of evolution. This important question about the basis of convergent phenotypes should be addressed in future studies, and thick-lipped fish species, including those outside the family Cichlidae, appear as an excellent model system.

Conclusion

Our integrative evolutionary, ecological, morphological, observational and genomic analysis of thick-lipped species in East Africa and in Nicaragua reveals stunning similarities between these convergent morphs. Both thick-lipped species appear to have evolved early in the respective clade, they seem to have adapted to the same habitat (rocks) and food source (hard-shelled prey), and their evolution was associated with comparable morphological trajectories, especially in the mouth and head region. Importantly, we also show that the expression patterns of at least some genes are similar, too. We thus provide valuable resource for future studies focusing on the development of this trait and genetic basis of convergence.

Acknowledgements

We would like to thank our helpers in the field, V. Campos, C. Heule, B. Meyer, M. Roesti; E. P. van den Berghe and G. Tembo and his crew for their logistic support in Nicaragua and Africa, respectively; the Ministerio del Ambiente y los Recursos Naturales Nicaragua (MARENA) and the Lake Tanganyika Research Unit, Department of Fisheries, Republic of Zambia,

for research permits; and three anonymous referees and the Subject Editor, C. Eizaguirre, for valuable comments. This study was supported by grants from the Fundacao para Ciencia e a Tecnologia (FCT, Portugal) to M. E. S., the Swiss Academy of Sciences to A. I., the Spanish Minsiterio de Economía y Competitividad to M. B., and the European Research Council (ERC, Starting Grant 'INTERGENADAPT'), the University of Basel and the Swiss National Science Foundation (SNF, grants 3100A0_122458 and CRSII3_136293) to W. S.

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.

Arendt J, Reznick D (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology and Evolution*, **23**, 26–32.

Arnegard ME, Snoeks J, Schaefer SA (2001) New three-spotted cichlid species with hypertrophied lips (Teleostei: Cichlidae) from the deep waters of Lake Malawî/Nyasa, Africa. *Copeia*, **2001**, 705–717.

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

Barlow GW (1976) The Midas cichlid in Nicaragua. In: *Investigations of the Ichthyofauna of Nicaraguan lakes* (ed. Thorson TB), pp. 333–358. School of Life Sciences, University of Nebraska, Lincoln, Nebraska.

Barlow GW, Munsey JW (1976) The Red Devil Midas Cichlid species complex in Nicaragua. In: *Investigations of the Ichthyofauna of Nicaraguan Lakes* (ed. Thorson TB), pp. 359–370. School of Life Sciences, University of Nebraska, Lincoln, Nebraska.

Barluenga M, Meyer A (2004) The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? *Molecular Ecology*, **13**, 2061–2076.

Barluenga M, Meyer A (2010) Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus* spp.) in the Nicaraguan crater lakes. *BMC Evolutionary Biology*, **10**, 326.

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

Berner D, Roesti M, Hendry AP, Salzburger W (2010) Constraints on speciation suggested by comparing lake-stream stickleback divergence across two continents. *Molecular Ecology*, **19**, 4963–4978.

Blankenberg D, Von Kuster G, Coraor N (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Current protocols in molecular biology/edited by Frederick M. Ausubel et al.* Chapter 19, Unit 19 10 11-21.

Brakefield PM (2006) Evo-devo and constraints on selection. *Trends in Ecology and Evolution*, **21**, 362–368.

Brichard P (1989) Cichlids and all Other Fishes of Lake Tanganyika, T.H.F. Publications, Neptune City, New Jersey.

Bruford MW, Hanotte O, Brookfield JFY, Burke T (1998) Multilocus and single-locus DNA fingerprinting. In: *Molecular Analysis of Populations* (ed. Hoelzel AR), pp. 283–336. Oxford University Press, New York.

Butler MA, Sawyer SA, Losos JB (2007) Sexual dimorphism and adaptive radiation in Anolis lizards. *Nature*, **447**, 202–205.

Diepeveen ET, Salzburger W (2011) Molecular characterization of two endothelin pathways in East African cichlid fishes. *Journal of Molecular Evolution*, **73**, 355–368.

Duden R (2003) ER-to-Golgi transport: COP I and COP II function (Review). *Molecular Membrane Biology*, **20**, 197–207.

Elmer KR, Meyer A (2011) Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology and Evolution*, **26**, 298–306.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fryer G (1959) The trophic interrelationships and ecology of some littoral communities of Lake Nyasa with a special reference to the fishes, and a discussion on the evolution of a group of rock-frequenting Cichlidae. *Proceedings of the Zoological Society of London*, **132**, 153–281.

Fryer G, Iles TD (1972) The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution. Oliver & Boyd, Edinburgh.

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

Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceeding of the National Academy of Sciences U S A*, **102**, 18040–18045.

Genner MJ, Seehausen O, Lunt DH *et al.* (2007) Age of cichlids: new dates for ancient lake fish radiations. *Molecular Biology and Evolution*, **24**, 1269–1282.

Giardine B, Riemer C, Hardison RC *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, **15**, 1451–1455.

Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, **11**, R86.

de Graaf M, Dejen E, Osse JWM, Sibbing FA (2008) Adaptive radiation of Lake Tana's (Ethiopia) *Labeobarbus* species flock (Pisces, Cyprinidae). *Marine and Freshwater Research*, **59**, 391–407.

Grant PR, Grant BR (2008) How and Why Species Multiply: The Radiations of Darwin's Finches. Princeton University Press, Princeton, USA.

Greenwood PH (1965) Environmental effects on the pharyngeal mill of a cichlid fish, *Astatoreochromis alluaudi*, and their taxonomic implications. *Proceedings of the Linnean Society London*, **176**, 1–10.

Greenwood PH (1974) The cichlid fishes of Lake Victoria East Africa: the biology and evolution of a species flock. *Bulletin of the British Museum for Natural History (Zool.) Supplementary*, **6**, 1–134.

Harmon LJ, Kolbe JJ, Cheverud JM, Losos JB (2005) Convergence and the multidimensional niche. *Evolution*, **59**, 409–421.

Herder F, Schwarzer J, Pfaender J, Hadiaty RK, Schliewen UK (2006) Preliminary checklist of sailfin silversides (Teleostei: Telmatherinidae) in the Malili Lakes of Sulawesi (Indonesia), with a synopsis of systematics and threats. *Verhandlungen der Gesellschaft für Ichthyologie*, **5**, 139–163.

Huysseune A (1995) Phenotypic plasticity in the lower pharyngeal jaw dentition of *Astatoreochromis alluaudi* (Teleostei: Cichlidae). *Archives in Oral Biology*, **40**, 1005–1014.

Johnson MA, Revell LJ, Losos JB (2009) Behavioral convergence and adaptive radiation: effects of habitat use on territorial behavior in Anolis lizards. *Evolution*, **64**, 1151–1159.

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.

Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, **11**, 353–357.

Klingenberg CP, Barluenga M, Meyer A (2003) Body shape variation in cichlid fishes of the *Amphilophus citrinellus* species complex. *Biological Journal of the Linnean Society*, **80**, 397–408.

Koblmuller S, Egger B, Sturmbauer C, Sefc KM (2010) Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Molecular Phylogenetics and Evolution*, **55**, 318–334.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **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. *Molecular Phylogenetics and Evolution*, **2**, 158–165.

Konings A (1998) Tanganyikan Cichlids in their Natural Habitat. Cichlid Press, El Paso.

Lang M, Miyake T, Braasch I *et al.* (2006) A BAC library of the East African haplochromine cichlid fish *Astatotilapia burtoni*. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, **306B**, 35–44.

Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54–66.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Loiselle PV (1998) The *Amphilophus labiatus* Species Complex. The Cichlid Room Companion. Retrieved on June 05, 2012, from: http://www.cichlidae.com/article.php?id=106.

Losos JB (2009) Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles. University of California Press, Berkeley.

Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. American Society of Naturalists E. O. Wilson award address. *American Naturalist*, **175**, 623–639.

Losos JB (2011) Convergence, adaptation and constraint. *Evolution*, **65**, 1827–1840.

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature*, **457**, 830–836.

Losos JB, Jackmann TR, Larson A, De Queiroz K, Rodriguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**, 2115–2118.

Matschiner M, Hanel R, Salzburger W (2011) On the origin and trigger of the notothenioid adaptive radiation. *PLoS One*, **6**, e18911.

Meyer A (1987) Phenotypic plasticity and heterochrony in *Cichlasoma managuense* (Pisces, Cichlidae) and their implications for speciation in cichlid fishes. *Evolution*, **41**, 1357–1369.

Meyer A, Morrissey JM, Schartl M (1994) Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. *Nature*, **368**, 539–542.

Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evolutionary Biology*, **11**, 116.

Nori M (1997) Structure of littoral fish communities. In: *Fish Communities in Lake Tanganyika* (eds Kawanabe H, Hori M, Nagoshi M), pp. 277–298. Kyoto University Press, Kyoto.

Nosil P, Crespi BJ, Sandoval DP (2002) Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, **417**, 440–443.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Ribbink AJ, Marsch BA, Marsch AC, Ribbink AC, Sharp BJ (1983) A preliminary survey of the cichlid fishes of rocky habitats in Lake Malawi. *South African Journal of Zoology*, **18**, 149–310.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Rohlf FJ (2006) DIG, Version 2.10.0. Department of Ecology and Evolution. State University of New York, Stony Brook, NY.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

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

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**, 169–185.

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. *Systematic Biology*, **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 Evolutionary Biology*, **5**, 17.

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.

Salzburger W, Ewing GB, Von Haeseler A (2011) The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology*, **20**, 1952–1963.

Schevzov G, Lloyd C, Gunning P (1992) High level expression of transfected beta- and gamma-actin genes differentially impacts on myoblast cytoarchitecture. *The Journal of Cell Biology*, **117**, 775–785.

Schluter D (2000) The Ecology of Adaptive Radiation. Oxford University Press, New York.

Schluter D, Nagel LM (1995) Parallel speciation by natural selection. *American Naturalist*, **146**, 292–301.

Schoener TW (1970) Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, **51**, 408–418.

Seehausen O (1996) Lake Victoria Rock Cichlids, Verdujin Cichlids, Zevenhuizen, The Netherlands.
Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceeding of the Royal Society London B*, **273**, 1987–1998.
Sereno PC, Wilson JA, Conrad JL (2004) New dinosaurs link southern landmasses in the Mid-Cretaceous. *Proceeding of the Royal Society London B*, **271**, 1325–1330.
Sibbing FA, Nagelkerke LAJ, Stet RJM, Osse JWM (1998) Speciation of endemic Lake Tana barbs (*Cyprinidae*, Ethiopia) driven by trophic resource partitioning: a molecular and ecomorphological approach. *Aquatic Ecology*, **32**, 217–227.
Simpson GG (1953) The Major Features of Evolution, Columbia University Press, New York.
Soni KG, Mardones GA, Sougrat R, Smirnova E, Jackson CL, Bonifacino JS (2009) Coatomer-dependent protein delivery to lipid droplets. *Journal of Cell Science*, **122**, 1834–1841.
Stiassny MLJ, Meyer A (1999) Cichlids of the Rift Lakes. *Scientific American*, **280**, 64–69.
Sturmbauer C, Hainz U, Baric S, Verheyen E, Salzburger W (2003) Evolution of the tribe Tropheini from Lake Tanganyika: synchronized explosive speciation producing multiple evolutionary parallelism. *Hydrobiologia*, **500**, 51–64.
Swofford DL (2003) PAUP*—Phylogenetic Analyses Using Parsimony and Other Methods, Version 4.0, Sinauer, Sunderland, Massachusetts.
Tsukita S, Furuse M, Itoh M (2001) Multifunctional strands in tight junctions. *Nature Reviews in Molecular Cell Biology*, **2**, 285–293.
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.
Villa J (1982) Peces Nicaragüenses de Agua Dulce, Fondo de Promoción Cultural, Banco de América, Managua.
Wake DB, Wake MH, Specht CD (2011) Homoplasy: from detecting pattern to determining process and mechanism of evolution. *Science*, **331**, 1032–1035.
Wallace RK (1981) An assessment of diet-overlap indexes. *Transactions of the American Fisheries Society*, **110**, 72–76.
Yamaoka K (1997) Trophic ecomorphology of Tanganyikan cichlids. In: *Fish Communities in Lake Tanganyika* (eds Kawanabe H, Hori M, Nagoshi M), pp. 27–56. Kyoto University Press, Kyoto.
Yoder JB, Clancey E, Des Roches S *et al.* (2010) Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, **23**, 1581–1596.

M.C., E.T.D., M.E.S. and A.I. are PhD students in the group of W.S. M.C. is interested in parallel evolution events as natural replicates to test hypotheses about trait evolution and the different (or similar) genetic bases that underlie these phenotypes. E.T.D. is interested in the genetic basis of adaptive traits and the selective forces acting upon these genes. M.E.S. is interested in the ecological and developmental mechanisms underlying the emergence and diversification of novel adaptive traits. A.I. is interested in ecomorphological adaptations, phylogeography and taxonomy in cichlid fishes. M.M. recently finished his PhD in the group of W.S. and is now postdoctoral fellow with Patrik Nosil in Sheffied. His research is concerned with morphological and genomic evolution in adaptive radiations. N.B. is a technical assistant who is involved in several projects of the SalzburgerLab. M.B. is a group leader at the Natural History Museum in Madrid. Her research focuses on understanding incipient stages of speciation and the sequence of adaptations and specializations that organisms undergo after the colonization of new habitats. W.S. is Professor of Zoology and Evolutionary Biology at the University of Basel. The research of his team focuses on the genetic basis of adaptation, evolutionary innovation and animal diversification. The main model systems in the laboratory are threespine stickleback fish, Antarctic notothenioids and the exceptionally diverse assemblages of cichlid fishes. The laboratory's homepage at http://www.evolution.unibas.ch/salzburger/ provides further details on the group's (research) activities.

Data accessibility

Newly generated DNA sequences for phylogenetic and haplotype analyses have been deposited in GenBank under accession numbers JX402217–JX402407 (see Appendix S1 for details). Illumina reads from the RNA-seq experiments are available from the Sequence Read Archive (SRA) at NCBI under the accession number SRA052992. Data from the stomach and gut content analyses, the MorphoJ input files and the quantitative real-time PCR experiments have been deposited at Dryad (doi:10.5061/dryad.vf1ms).

Supporting information

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

Appendix S1 List of specimens used in this study including sampling date and location and GenBank accession numbers.

Appendix S2 PCA of overall body shape of the African cichlid *Lobochilotes labiatus* and the Nicaraguan species *Amphilophus labiatus* and *A. citrinellus* (a) and distribution of landmarks for morphometric analyses (b).

Appendix S3 Blast2GO annotations of genes with differential expression between lip tissue from thick-lipped and normal-lipped cichlid species.

Appendix S4 Underwater video showing snail feeding in *Lobochilotes labiatus*.

Appendix S5 Results of the quantitative real-time PCR experiments comparing adult and juvenile lip tissue of the African cichlid species *Lobochilotes labiatus* and *Astatotilapia burtoni*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

CHAPTER 6

THE ECOLOGICAL AND GENETIC BASIS OF CONVERGENT THICK-LIPPED PHENOTYPES IN CICHLID FISHES

SUPPORTING MATERIAL

S1. LIST OF SPECIMENS
S2. PCA OF OVERALL BODY SHAPES
S3. BLAST2GO ANNOTATIONS
S4. UNDERWATER VIDEO (NOT INCLUDED)
S5. QUANTITATIVE REAL-TIME PCR RESULTS

Appendix S1 List of specimens used in this study including sampling date and location and GenBank accession numbers

Sample ID		Species	Sampling	Location	DNA/RNA sequencing/GenBank accession numbers				qPCR					
DNA	RNA		Date		mtControl	ednrb1	phpt1	RNA sequencing	BNIP3	Actb	Cldn7	Copb	His3	ACSL4
23H8		*Amphilophus citrinellus*	9/5/09	Isletas		JX402280	JX402217							
23H9		*Amphilophus citrinellus*	9/5/09	Isletas		JX402281	JX402218							
23I1		*Amphilophus citrinellus*	9/5/09	Isletas		JX402282	JX402219							
23I2		*Amphilophus citrinellus*	9/5/09	Isletas			JX402220							
24A2		*Amphilophus citrinellus*	9/5/09	Isletas		JX402283	JX402221							
24A3		*Amphilophus citrinellus*	9/5/09	Isletas		JX402284	JX402222							
24A4		*Amphilophus citrinellus*	9/5/09	Isletas		JX402285	JX402223							
24A5		*Amphilophus citrinellus*	9/5/09	Isletas		JX402286	JX402224							
25D1	31I8	*Amphilophus citrinellus*	9/9/09	Isletas		JX402288	JX402225							
25D2	30D7	*Amphilophus citrinellus*	9/9/09	Isletas		JX402289	JX402226							
25D6	30F1	*Amphilophus citrinellus*	9/9/09	Isletas			JX402227		y	y	y	y	y	y
25D7	30F3	*Amphilophus citrinellus*	9/9/09	Isletas		JX402298	JX402228							
25D8	30F9	*Amphilophus citrinellus*	9/9/09	Isletas		JX402299	JX402229							
25E1	30F6	*Amphilophus citrinellus*	9/9/09	Isletas			JX402230							
25E2	30G2	*Amphilophus citrinellus*	9/9/09	Isletas		JX402292	JX402231							
25E3	30G4	*Amphilophus citrinellus*	9/9/09	Isletas		JX402287	JX402232		y	y	y	y	y	y
25E4	30G6	*Amphilophus citrinellus*	9/9/09	Isletas		JX402293	JX402233							
25E6	30H3	*Amphilophus citrinellus*	9/9/09	Isletas			JX402234							
25E8	30H6	*Amphilophus citrinellus*	9/9/09	Isletas		JX402296	JX402235							
25E9	30H7	*Amphilophus citrinellus*	9/9/09	Isletas		JX402297	JX402236							
25D3	30E2	*Amphilophus citrinellus*	9/9/09	Isletas		JX402290								
25D4	30E7	*Amphilophus citrinellus*	9/9/09	Isletas					y	y	y	y	y	y
25E5	30G8	*Amphilophus citrinellus*	9/9/09	Isletas		JX402294			y	y	y	y	y	y
25E7	30H5	*Amphilophus citrinellus*	9/9/09	Isletas		JX402295								
25F1	30H9	*Amphilophus citrinellus*	9/9/09	Isletas		JX402298								
26I3	32A7	*Amphilophus citrinellus*	9/14/09	Managua Miraflores					y	y	y	y	y	y
26I4	32B1	*Amphilophus citrinellus*	9/14/09	Managua Miraflores					y	y	y	y	y	y
23F9		*Amphilophus labiatus*	9/5/09	Isletas	JX402360	JX402301	JX402237							
23G1		*Amphilophus labiatus*	9/5/09	Isletas	JX402354	JX402302	JX402238							
23G3		*Amphilophus labiatus*	9/5/09	Isletas			JX402239							
23G6		*Amphilophus labiatus*	9/5/09	Isletas	JX402367		JX402240							
23G9		*Amphilophus labiatus*	9/5/09	Isletas	JX402361	JX402305	JX402241							
23H3		*Amphilophus labiatus*	9/5/09	Isletas	JX402365	JX402306	JX402242							
23H6		*Amphilophus labiatus*	9/5/09	Isletas	JX402368		JX402243							
23H7		*Amphilophus labiatus*	9/5/09	Isletas	JX402375	JX402308	JX402244							
25A1	31C3	*Amphilophus labiatus*	9/9/09	Isletas	JX402371	JX402310	JX402245							
25A5	31D6	*Amphilophus labiatus*	9/9/09	Isletas	JX402364	JX402311	JX402246							
25A6	31D9	*Amphilophus labiatus*	9/9/09	Isletas	JX402369		JX402247							
25A7	31E2	*Amphilophus labiatus*	9/9/09	Isletas	JX402378	JX402312	JX402248							
25A8	31E3	*Amphilophus labiatus*	9/9/09	Isletas	JX402376	JX402313	JX402249							
25B1	31E7	*Amphilophus labiatus*	9/9/09	Isletas	JX402363	JX402314	JX402250		y	y	y	y	y	y
25B3	31F2	*Amphilophus labiatus*	9/9/09	Isletas	JX402366	JX402316	JX402251							
25B4	31F4	*Amphilophus labiatus*	9/9/09	Isletas	JX402351	JX402309	JX402252							
25B5	31F5	*Amphilophus labiatus*	9/9/09	Isletas	JX402372		JX402253							
25B6	31F7	*Amphilophus labiatus*	9/9/09	Isletas	JX402359	JX402317	JX402254							
25B7	31F9	*Amphilophus labiatus*	9/9/09	Isletas	JX402377	JX402318	JX402255							
25B9	31G3	*Amphilophus labiatus*	9/9/09	Isletas	JX402349	JX402320	JX402256							
23F8		*Amphilophus labiatus*	9/5/09	Isletas	JX402353	JX402300								
23G2		*Amphilophus labiatus*	9/5/09	Isletas	JX402374	JX402303								
23H5		*Amphilophus labiatus*	9/5/09	Isletas	JX402350	JX402307								
25B2	31E9	*Amphilophus labiatus*	9/9/09	Isletas	JX402358	JX402315								

25B8	31G2	*Amphilophus labiatus*	9/9/09	Isletas	JX402370	JX402319										
23G5		*Amphilophus labiatus*	9/5/09	Isletas	JX402373											
23G7		*Amphilophus labiatus*	9/5/09	Isletas	JX402355											
23G8		*Amphilophus labiatus*	9/5/09	Isletas	JX402356	JX402304										
23H1		*Amphilophus labiatus*	9/5/09	Isletas	JX402357											
25A3	31C9	*Amphilophus labiatus*	9/9/09	Isletas	JX402352											
25A4	31D4	*Amphilophus labiatus*	9/9/09	Isletas	JX402348											
25A9	31E5	*Amphilophus labiatus*	9/9/09	Isletas	JX402362					y	y	y	y	y	y	
25C1	31G6	*Amphilophus labiatus*	9/9/09	Isletas						y	y	y	y	y	y	
25C6	31H7	*Amphilophus labiatus*	9/9/09	Isletas						y	y	y	y	y	y	
28A3	32G6	*Amphilophus labiatus*	9/18/09	Managua Miraflores						y	y	y	y	y	y	
28C6	32H7	*Amphilophus labiatus*	9/20/09	Ometepe, San Ramon	J					y	y	y	y	y	y	
35A1		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402388	JX402321	JX402257									
35A2		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402383	JX402322	JX402278									
35A3		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402398	JX402323	JX402263									
35A4		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402399	JX402324	JX402266									
35A5		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402400	JX402325	JX402268									
35A6		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402405	JX402326	JX402262									
35A7		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402391	JX402327	JX402279									
35A8		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402392	JX402328	JX402259									
35B6		*Lobochilotes labiatus*	2/21/10	Mbita Island W	JX402401	JX402329	JX402258									
36B3		*Lobochilotes labiatus*	2/22/10	Mpulungu area	JX402404	JX402330	JX402267									
36B4		*Lobochilotes labiatus*	2/22/10	Mpulungu area	JX402406	JX402331	JX402275									
36H7		*Lobochilotes labiatus*	2/23/10	Kasakalawe Lodge	JX402403	JX402333	JX402277									
43D7	81A2	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402379	JX402334	JX402261	y	y	y	y	y	y	y		
43D8	81A5	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402407	JX402335	JX402260	y	y	y	y	y	y	y		
43E4	81B9	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402390	JX402340	JX402264		y	y	y	y	y	y		
43E5	81B4	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402381	JX402341	JX402273	y	y	y	y	y	y	y		
43E6		*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402382	JX402342	JX402265									
44G8		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402395	JX402343	JX402270									
44G9		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402386	JX402344	JX402271									
44H1		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402380	JX402345	JX402269									
44H3		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402396		JX402276									
44H5		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402387	JX402346	JX402272									
44H6		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402402	JX402347	JX402274									
36H8		*Lobochilotes labiatus*	2/23/10	Kasakalawe Lodge	JX402389	JX402332										
43D9	81A7	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402384	JX402336										
43E1	81B1	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402385	JX402337										
43E2	81B3	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402393	JX402338										
43E3	81B7	*Lobochilotes labiatus*	2/28/10	Toby's Place	JX402394	JX402339										
44H4		*Lobochilotes labiatus*	3/1/10	Toby's Place	JX402397											
43D6	81A3	*Lobochilotes labiatus*	2/28/10	Toby's Place				y	y	y	y	y	y	y		
55D1	81H9	*Lobochilotes labiatus*	3/11/10	Mpulungu area					y	y	y	y	y	y		
55D2	81I2	*Lobochilotes labiatus*	3/11/10	Mpulungu area					y	y	y	y	y	y		
55D8	79A5	*Lobochilotes labiatus*	3/11/10	Mukaka					y	y	y	y	y	y		
55D9	79A9	*Lobochilotes labiatus*	3/11/10	Mukaka				y	y	y	y	y	y	y		
BIB4	BMB4	*Lobochilotes labiatus*	7/20/11	Toby`s place					y	y	y	y	y	y		
BIC4	BMC4	*Lobochilotes labiatus*	7/20/11	Toby`s place					y	y	y	y	y	y		
BIC6	BMC6	*Lobochilotes labiatus*	7/20/11	Toby`s place					y	y	y	y	y	y		
BIC7	BMC7	*Lobochilotes labiatus*	7/20/11	Toby`s place					y	y	y	y	y	y		
BIC9	BMC9	*Lobochilotes labiatus*	7/20/11	Toby`s place					y	y	y	y	y	y		

additional Tropheini species for phylogenetic analyses:

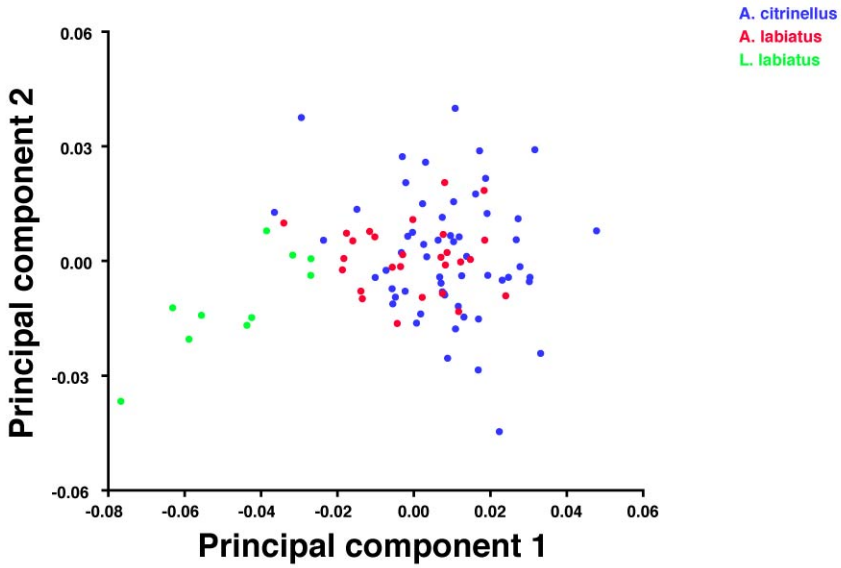
	Species	**mtControl**	**ednrb1**	**phpt1**	**ND2**
	Astatotilapia burtoni	AY930000	JF900252	JF900181	JF900319
	Ctenochromis horei	AY301952	JF900262	JF900191	EU753935

Gnathochromis pfefferi	AY301954	JF900269	JF900198	U07248
Interochromis loockii	GQ995855	JF900304	JF900232	JF900322
Limnotilapia dardennii	AY301956	JF900285	JF900214	GQ995724
Lobochilotes labiatus	AY301958	JF900286	JF900215	U07254
Petrochromis ephippium	AY301959	JF900300	JF900229	JF900323
Petrochromis famula	AY301960	JF900301	JF900230	JF900324
Petrochromis fasciolatus	GQ995911	JF900325	JF900231	JF900325
Petrochromis macrognathus	AY929963	JF900304	JF900233	AY930068
Petrochromis polyodon	AY301967	JF900305	JF900234	JF900326
Pseudosimochromis curvifrons	AY301973	JF900307	JF900236	GQ995777
Simochromis babaulti	AY301975	JF900309	JF900238	GQ995782
Simochromis diagramma	AY301977	JF900310	JF900239	AY930087
Tropheus moorii	AY930020	JF900314	JF900243	AY930093

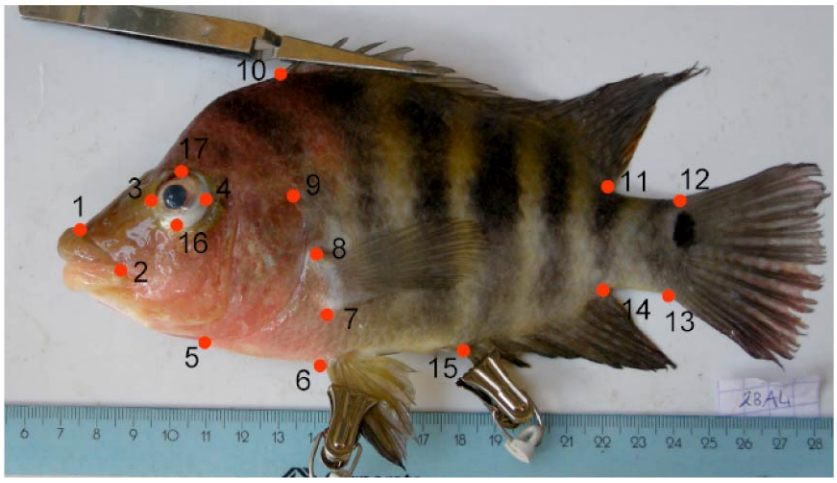
additional *Amphilophus citrniellus* mitochondrial control region sequences:

ampcit1	AY567011
ampcit2	AY567018
ampcit3	AY567012
ampcit4	AY567013
ampcit5	AY567014
ampcit6	AY567015
ampcit7	AY567016
ampcit8	AY567017
ampcit9	EF219270
ampcit10	EF219269
ampcit11	EF219268
ampcit12	EF219252
ampcit13	EF219251
ampcit14	EF219250
ampcit15	EF219249
ampcit16	EF219248
ampcit17	EF219247
ampcit18	AY567470
ampcit19	AY567469
ampcit20	AY567468
ampcit21	AY567467
ampcit22	AY567466
ampcit23	AY567465
ampcit24	AY567464
ampcit25	AY567463
ampcit26	AY567462
ampcit27	AY567461
ampcit28	AY567460
ampcit29	AY567459
ampcit30	AY567458
ampcit31	AY567457

(a)



(b)



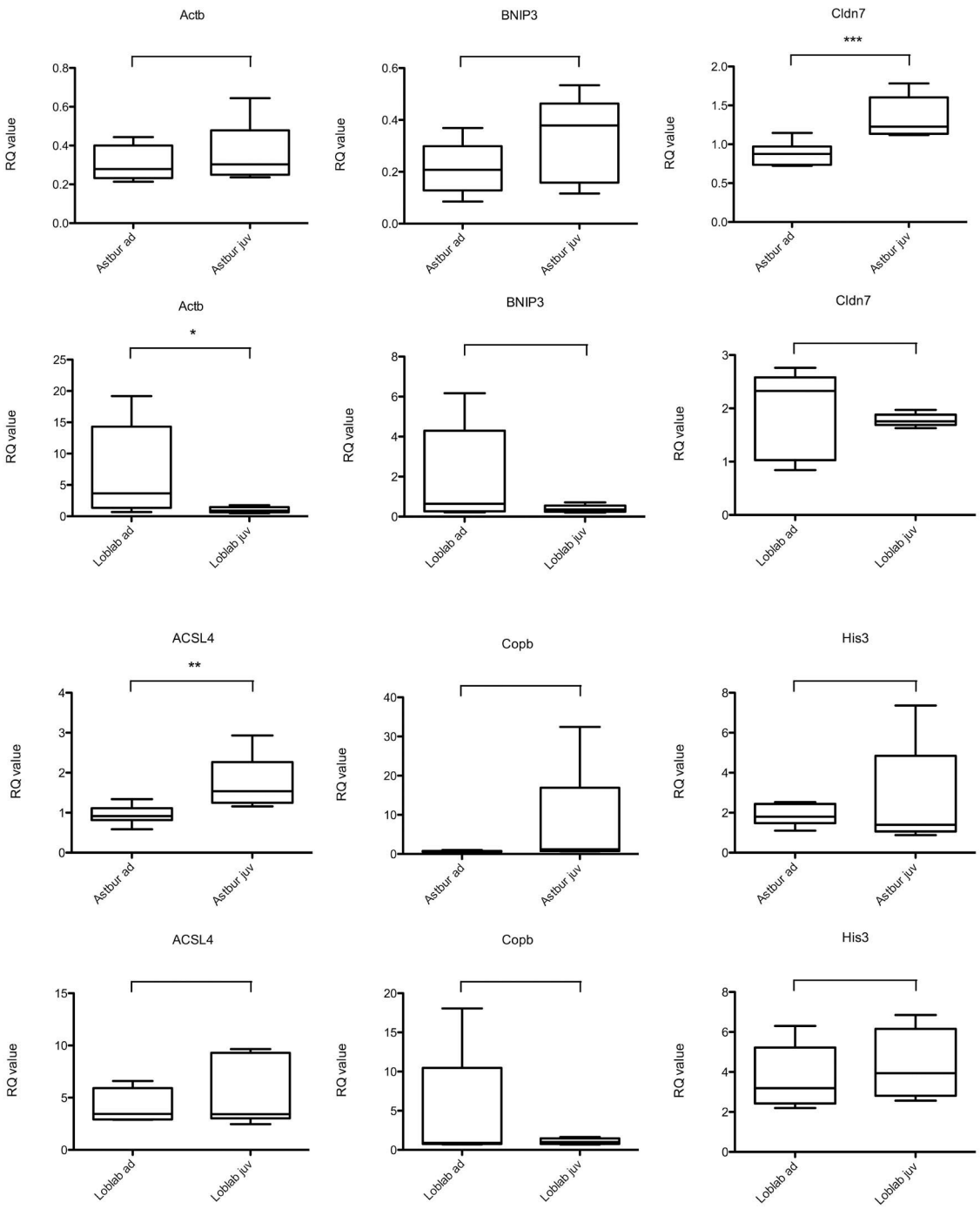
Appendix S2. PCA of overall body shape of the African cichlid *Lobochilotus labiatus* and the Nicaraguan species *Amphilophus labiatus* and *A. citrinellus* (a) and distribution of landmarks for morphometric analyses (b).

Blast2GO annotations of genes with differential expression between lip-tissue from thick-lipped and normal-lipped cichlid species

26s protease regulatory subunit 8
3-hydroxyanthranilate -dioxygenase
60s acidic ribosomal protein p2
actin-related protein 2-a
actin-related protein 3
activating transcription factor 4
acyl carrier mitochondrial precursor
acyl- -binding protein
adaptor-related protein complex mu 1 isoform cra_a
adaptor-related protein complex mu 1 subunit
adp-dependent glucokinase-like
adp-ribose mitochondrial-like
atp synthase subunit mitochondrial precursor
atpase mitochondrial precursor
baculoviral iap repeat-containing protein 4
bcl2 adenovirus e1b 19 kda protein-interacting protein 3
bcl2 adenovirus e1b 19 kda protein-interacting protein 3-like
beta actin
calpastatin
carboxypeptidase z-like
caspase-8
chaperonin containing subunit 6a (zeta 1)
chromobox protein homolog 3
claudin 7
cmp-n-acetylneuraminate-beta-galactosamide-alpha- -sialyltransferase 1-like
coatomer subunit beta
comm domain-containing protein 9
complement c1q tumor necrosis factor-related protein 3-like
cop9 signalosome complex subunit 8
coproporphyrinogen oxidase
cystathionine gamma-lyase
cystatin precursor
cysteine and glycine-rich protein 1
cytochrome c oxidase polypeptide viia-liver mitochondrial precursor
dcn1-like protein 1
dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase mitochondrial
dnaj homolog subfamily c member 9-like
dynactin subunit 5
ectonucleoside triphosphate diphosphohydrolase 3
estradiol 17-beta-dehydrogenase 12-b
eukaryotic translation initiation factor 3 subunit i
eukaryotic translation initiation factor 3 subunit k
eukaryotic translation initiation factor 3 subunit l
eukaryotic translation initiation factor 3 subunit m
eukaryotic translation initiation factor 4 gamma 2-like
eukaryotic translation initiation factor 4h
excitatory amino acid transporter 1 isoform 1
fk506-binding protein 2 precursor
forkhead box q1

glutamate dehydrogenase
glyoxalase domain-containing protein 4-like
grainyhead-like protein 1 homolog
granulins precursor
gtpase imap family member 4-like
gtpase imap family member 7-like
gtpase imap family member 8-like
gtpase imap family member 8-like
h1 histone
heat shock 70 kda protein 12a-like
histone
iars protein
importin-7
integrin beta-4-like
interferon-induced protein 35
isocitrate dehydrogenase
l _3
lamin b1
laminin subunit gamma-2
loc100127300 protein
long-chain-fatty-acid-- ligase 4
low quality protein: coronin-1c-like
lrr and pyd domains-containing protein 3-like
magnesium transporter 1
major vault protein
membrane magnesium transporter 1-like
methylmalonyl epimerase
microfibril-associated glycoprotein 4-like
mortality factor 4 like 1
myosin regulatory light chain smooth muscle isoform
nadh dehydrogenase
nadh dehydrogenase 1 alpha subcomplex subunit 11
nedd4 family-interacting protein 1
nedd4 family-interacting protein 1
nuclear factor erythroid 2-related factor 1-like
ornithine decarboxylase
pancreatic progenitor cell differentiation and proliferation factor
peptidylprolyl isomerase b (cyclophilin b)
phosphoglycerate kinase 1
piggybac transposable element-derived protein 4-like
pre-mrna splicing factor
PREDICTED: galectin-3-like [Oreochromis niloticus]
PREDICTED: hypothetical protein LOC100704514 [Oreochromis niloticus]
prefoldin subunit 4
probable glutathione peroxidase 8-like
programmed cell death 6-interacting protein
proteasome subunit alpha type-1
proteasome subunit alpha type-6
protein disulfide isomerase family member 4
protein fam100a-like
protein fam176b-like
protein kiaa0664-like
protein rer1

rab acceptor 1
ras-related protein rab-11b
regulator of g-protein signaling 2
renin receptor isoform 3
ribosomal l1 domain-containing protein 1-like
rilp-like protein 1
scinderin like a
scinderin like a
secretory carrier-associated membrane protein 2-like
septin 10
signal peptidase complex catalytic subunit sec11a
signal peptide peptidase-like 2a-like
small 1
sodium-dependent phosphate transporter 1
solute carrier family facilitated glucose transporter member 11-like
solute carrier family member 30
splicing factor 3b subunit 1
subfamily member 11
syntaxin 12
t-cell receptor type 1
t-complex protein 1 subunit alpha-like
t-complex protein 1 subunit theta
tbc1 domain member 15
thioredoxin domain containing 4 (endoplasmic reticulum)
threonyl-trna cytoplasmic
transaldolase
transcription factor ap-2 gamma (activating enhancer binding protein 2 gamma)
transmembrane protein 214
transmembrane protein 59 precursor
transmembrane protein 79
transposon tx1 uncharacterized 149 kda
tumor protein 63
tumor-associated calcium signal transducer 2 precursor
u6 snrna-associated sm-like protein lsm8
uap56-interacting factor-like
uncharacterized protein c22orf25-like
upf0510 protein inm02 precursor
v-type proton atpase catalytic subunit a
v-type proton atpase subunit d 1



Appendix S5. Results of the quantitative real-time PCR experiments comparing adult and juvenile lip tissue of the African cichlid species *Lobochilotus labiatus* and *Astatotilapia burtoni*.

CHAPTER 7

DISCUSSION

The work presented in my doctoral thesis focuses on inferring the molecular evolutionary genetics of three candidate gene families (i.e., the endothelin, *dlx* and *Hivep* families; see Figure 7.1) in relation to their (presumed) roles in either evolutionary key innovations or other adaptations in East African cichlid fishes (chapters 2, 3, 4). Furthermore, a top-down approach, by means of RNA sequencing (chapter 6), was used to generate a candidate gene list for the thick-lipped phenotype, which represents the first insight into the genetic basis of this trait that evolved convergently in several cichlid adaptive radiations. In the following I will discuss the genetic basis, the role and mechanisms of gen(om)e duplication and the observed signatures of selection pressure of genes involved in naturally and sexually selected traits in cichlid fishes and put the obtained results in a broader context.

The genetic basis of naturally and sexually selected traits

As highlighted in the previous chapters, the great ecological diversity of cichlid fishes is brought forward by various naturally and sexually selected traits, which are thought to have played an important role in the evolutionary success of cichlids and which offers great opportunities to study the genetic basis of e.g., adaptation and diversification (e.g., Salzburger 2009).
Although in this work no actual functional studies were performed to uncover the exact biochemical, physiological or fitness consequences of the studied loci contributing to the respective traits, my results do give some general insights into the genetics of naturally and sexually selected traits. First of all, several conserved functions within vertebrate lineages and even between vertebrates and insects were previously found for the *dlx*, *Hivep* and endothelin gene family members, indicating that these functions are very likely to be conserved in cichlids as well. Functional studies, such as e.g., gene knockout or knockdown studies are often not feasible in non-model organism, including cichlid fishes. Therefore, I used other methods to verify the described and often well-studied functions of the vertebrate or insect orthologs in cichlids. With quantitative real-time PCR I showed that the *Edn1*/*EdnrAa* pathway is expressed in the pharyngeal jaw tissue and the *Edn3b*/*EdnrB1a* pathway in egg-spot tissue (chapter 2). In chapter 4, I measured gene expression of the *Hivep* paralogs, in particular after an immune challenge in liver and gills, two tissues where reaction to infection is expected to take place. Here I found a significant positive correlation between expression level and immune response parameters, confirming immune-related functions for the *Hivep* paralogs in cichlids. *Therefore the results of the gene expression assays performed in chapters 2 and 4) together with the known functions in other vertebrates, suggest that (several of) the studied loci are indeed involved in these naturally and sexually selected traits.*

Specific insight into the genetic basis of the convergent thick-lipped phenotype was obtained by using RNAseq and subsequent comparative quantitative real-time PCR (chapter 6). This approach resulted into a candidate gene list of 141 loci that now can be studied in detail with respect to thick-lipped phenotype in cichlids, including the - so far - not studied species of Lake Malawi and Victoria, as well as other fish species with hypertrophied lips. The comparative gene expression assay of both African and Nicaraguan thick-lipped species revealed similar patterns of gene expression, which indicates an overlap in genes underlying this trait in both species that are believed to

have been separated by 100 million years of evolution (see chapter 6). It is thus plausible that some of the indentified genes are involved in the thick-lipped phenotype of other (cichlid) fishes as well. Interestingly, Manousaki and colleagues found only a small number of overlapping and differentially expressed genes in four convergent Nicaraguan thick-lipped and thin-lipped species pairs (Manousaki *et al*. 2013), of which only one loci matched with our list of 141 candidate genes.
These results clearly demonstrate that many genes are likely to be involved in this morphological trait and that more work is needed to uncover, especially the overlap of genes in convergent cases. Furthermore, identifying the genetic pathways, to which these genes belong, should become the focus, as well as screening major candidate genes (i.e., genes typically differentially expressed in multiple thick-lipped species), and their regulatory regions, for adaptive changes. These efforts should help to get a better understanding of the complex genetic basis underlying this convergent morphological trait.

The role and mechanisms of gen(om)e duplication in the evolution of naturally and sexually selected traits

It is clear that gen(om)e duplications play a major role in the evolution of phenotypic diversity across species (Ohno 1970; Sidow 1996). Gen(om)e duplications provide the raw genetic material on which mutation, drift and selection can act (Ohno 1970; Force *et al*. 1999; Prince & Pickett 2002; Zhang 2003). *The three gene families studied in chapters 2, 3 and 4 are hypothesized to have evolved by whole genome duplication events (Stock et al. 1996; Braasch et al. 2009; Renz et al. 2011; Figure 7.1) and based on their (presumed) functions, offer valuable insights into the evolution of duplicates in relation to naturally and sexually selected traits in cichlid fishes.*

A variety of mechanisms is known by which duplicated genes can contribute to the evolution of phenotypic diversity. Although most of the duplicated genes become pseudogenes through nonfunctionalization (see Lynch 2000; Zhang 2003), several other mechanisms are described that maintain the duplicated, or paralogous, genes. Functional conservation of the duplicates can evolve through gene conversion (concerted evolution) or strong purifying selection, in which the sequences of the duplicates are either highly similar or only differ at synonymous positions (Zhang 2003). This way higher levels of gene products are produced via gene dosage. The ancestral functions can, furthermore, be divided over the two paralogous genes via subfunctionalization (Force *et al*. 1999), while novel functions can arise via neofunctionalization by the fixation of beneficial mutations (Ohno 1970). Especially the latter is commonly regarded as major cause underlying the evolution of phenotypic diversity.

Teleost fishes are characterized by three rounds of whole genome duplication events, which have been hypothesized to underlie the great species diversity: two in the lineage towards vertebrates (Sidow 1996) followed by a fish specific whole genome duplication (Taylor *et al*. 2001; 2003; Meyer & Van de Peer 2005; Volff 2005; Braasch & Postlethwait 2012; Figure 7.1). *Cichlid fishes, having an exceptionally high species diversity, are great model systems to study the role of gene duplication during the evolution of naturally and sexually selected traits.* In fact, many of the previously identified loci involved in naturally and sexually selected traits of

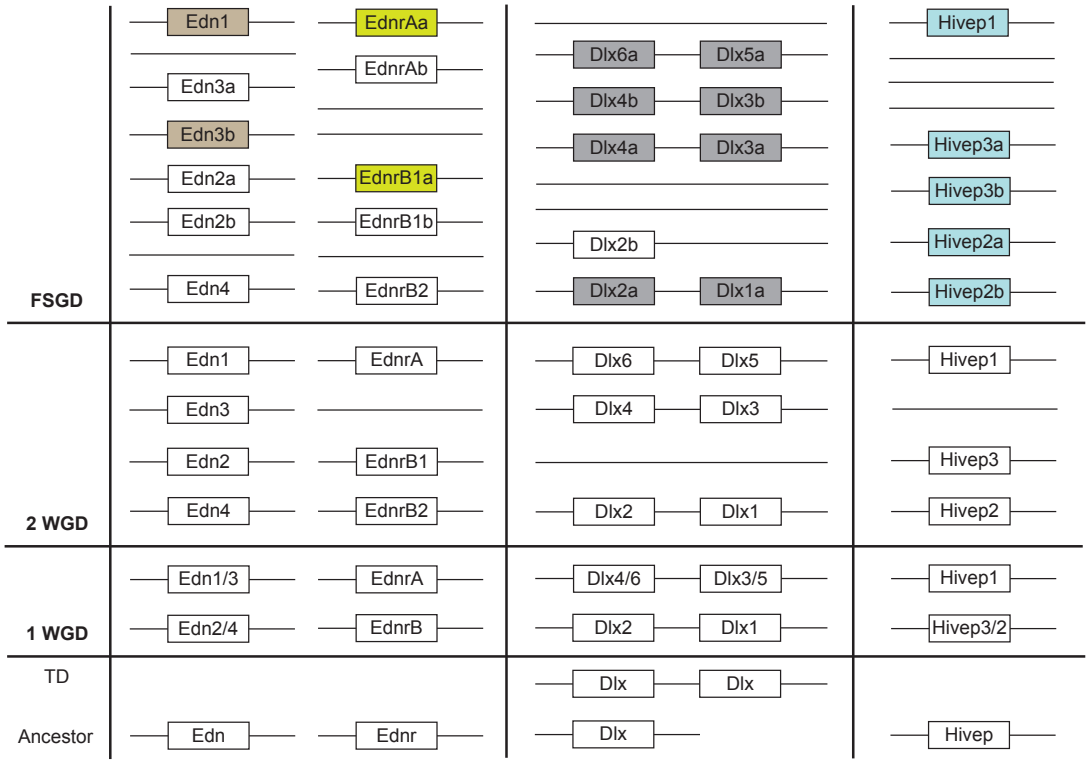


Figure 7.1. Gen(om)e duplication hypotheses for the candidate gene families. Depicted are the presumed ancestral states of the loci ('Ancestor'), the initial tandem duplication (TD) of the *dlx* locus and the three rounds of whole genome duplication (First round of Whole Genome Duplication (1 WGD), Second round of Whole Genome Duplication (2 WGD) and the Fish Specific Genome Duplication (FSGD)). The top part of the figure represents the teleost repertoire, although species-specific losses are known for e.g., the *dlx* gene family. Genes are shown as boxes, while nonfunctionalization is depicted as an empty line/genomic region. Colored boxes are the loci studied in chapters 2, 3 and 4. After Braasch *et al*. 2009; Renz *et al*. 2011.

cichlids, belong to different gene families (e.g., *bmp4* is involved in jaw formation (Albertson *et al*. 2005), *csf1ra* is expressed in egg-dummies (Salzburger *et al*. 2007) and a *fhl2* duplicate is involved in the evolution of egg-spots (Santos *et al*. *in prep*)).

My work on the Hivep, dlx and endothelin gene families provides further indications that loci produced by gen(om)e duplication are of particular relevance for naturally and sexually selected traits in cichlid fishes, since several of the studied loci can be associated with such traits in cichlids (see chapter 2, 3, 4). While the studied endothelin paralogs (ligands and receptors) are involved in distinct traits (i.e., mouth morphology versus pigmentation; see Braasch *et al*. 2009; chapter 2), it seems that several *dlx* paralogs are involved in the same traits (e.g., pharyngeal arches and egg-spot development; Renz *et al*. 2011; chapter 3). This is also the case for the *Hivep* paralogs, for which immune-related functions for three of the paralogs were found (chapter 4). *However, inferring the precise mechanisms by which these individual paralogs have been retained is hard without accurate and complete functional analyses. It is clear that nonfunctionalization occurred, as expected, in all three families and that concerted evolution is not observed in these families. Neo- and/or subfunctionalization are thus believed to play an important role in the evolution of these three gene families in cichlid fishes.*

Determining the signatures of selection pressure in naturally and sexually selected traits

In the previous I have proposed putative roles for the candidate gene family members in natural and sexually selected traits Since molecular evolution is – partly – responsible for evolutionary innovations (Yang 2002), genes underlying such traits are hypothesized to be subject to molecular selection favoring new or adaptive alleles (i.e., molecular adaptation; Yang 2006). Therefore, detailed screens of possible effects of selection can give insights into the mode of selection pressure (i.e., purifying, neutral or positive) on these candidate genes.

My approach consisted of comparing nuclear sequences from different cichlid species, in which each species was represented by a single individual, to calculate nonsynonymous/synonymous substitution rate ratios. By using this method I focus on divergence between species instead of polymorphisms within a species and it is, thus, impossible to differentiate between fixed and non-fixed substitutions per species, which might have affected – to some degree – the results presented here. More comprehensive population genetic approaches are available which take into account both the polymorphisms within species and the divergence between species, such as the McDonald-Kreitman test (McDonald & Kreitman 1991), which needs at least two samples per species. *In this work I was primarily interested in the divergence between cichlid lineages to study the molecular adaptation among closely related species.* In fact, the *dN*/*dS* approach is widely used, and originally developed, for interspecific comparisons to detect positive selection (Goldman & Yang 1994; Yang 2006; Jensen *et al*. 2007; Kryazhimskiy & Plotkin 2008). I thus, assume that the observed variable sites are fixed substitutions between species, although this assumption is expected to be violated, to some extend.

I found different signatures of molecular selection acting upon the candidate loci, both within and between gene families. These results generally imply that selection pressure is very different among loci involved in naturally and sexually selected traits, although certain families, *Hivep* in particular, seem to be characterized by positive selection, with only one locus evolving more conserved (chapter 4). In fact, it is known that complex phenotypes are often controlled by many interacting genes (Nadeau 2001) and that few changes in a small number of genes can change the spatio and/or temporal expression of signaling pathways, which can subsequently become co-opted for a new function that underlies the evolution of novelties (Orr 2005; Abzhanov *et al*. 2008). Selection on such beneficial changes may then, leave only local signatures of positive selection in these few regions, while the rest of the pathway is not characterized by elevated *dN*/*dS* ratios. It could thus be that e.g., beneficial substitutions occurred in other members of the endothelin pathways, slightly changing its functions, which led to the co-option of the whole pathway for functions in the morphogenesis of the pharyngeal jaw apparatus and egg-spots (chapter 2).
Also, I did not sequence the full coding regions, potentially missing important amino acid substitutions, although most loci which were characterized by positive selection had several regions with elevated *dN*/*dS* ratios (see sliding window plots in chapters 2 and 4). *Taken my results together then it is clear that positive selection is commonly observed in the Hivep immune-related genes, part of the dlx repertoire and absent in the endothelin family of ligands and receptors (although variation between individual*

loci within families is observed). Thus, different types of selection pressure are observed in these candidate genes for naturally and sexually selected traits in cichlids.

In conclusion, the results of my doctoral thesis provide valuable insights into the molecular evolutionary genetics of three specific gene families and offers exciting new insights into the genetic basis of one specific convergent adaptive trait: the thick-lipped phenotype observed in both African and Nicaraguan species.

References

Abzhanov A, Extavour CG, Groover A, *et al*. (2008) Are we there yet? Tracking the development of new model systems. *Trends in Genetics*, **24**, 353–360.

Albertson RC, Streelman JT, Kocher TD, Yelick PC (2005) Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 16287–16292.

Braasch I, Postlethwait J (2012) Polyploidy in fish and the teleost genome duplication. In: *Polyploidy and genome evolution* (eds Soltis PS, Soltis DE), pp. 341–383. Springer Verlag, Berlin Heidelberg.

Braasch I, Volff JN, Schartl M (2009) The endothelin system: Evolution of vertebrate-specific ligand-receptor Interactions by three rounds of genome duplication. *Mol Biol Evol*, **26**, 783–799.

Force A, Lynch M, Pickett FB, *et al*. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531–1545.

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*, **11**, 725–736.

Jensen JD, Wong A, Aquadro CF (2007) Approaches for identifying targets of positive selection. *Trends in Genetics*, **23**, 569–577.

Kryazhimskiy S, Plotkin JB (2008) The Population Genetics of dN/dS. *PLoS Genetics*, **4**, e1000304.

Lynch M (2000) The Evolutionary Fate and Consequences of Duplicate Genes. *Science*, **290**, 1151–1155.

Manousaki T, Hull PM, Kusche H, *et al*. (2013) Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology*, **22**, 650-669.

McDonald, Kreitman (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, **351**, 652–654.

Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays*, **27**, 937–945.

Nadeau JH (2001) Modifier genes in mice and humans. *Nature Reviews Genetics*, **2**, 165–174.

Ohno S (1970) *Evolution by gene duplication*. Springer Verlag, New York.

Orr HA (2005) The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, **6**, 119–127.

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics*, **3**, 827–837.

Renz AJ, Gunter HM, Fischer JM, *et al*. (2011) Ancestral and derived attributes of the dlx gene repertoire, cluster structure and expression patterns in an African cichlid fish. *EvoDevo*, **2**, 1.

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**, 169–185.

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

Santos ME, Braasch I, Boileau N, *et al*. On the origin and molecular basis of an evolutionary novelty in cichlid fishes. *in prep*.

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. *Current opinion in genetics & development*, **6**, 715–722.

Stock DW, Ellies DL, Zhao ZY, *et al*. (1996) The evolution of the vertebrate Dlx gene family. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 10858–10863.

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Research*, **13**, 382–390.

Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Phil. Trans. R. Soc. Lond. B*, **356**, 1661–1679.

Volff JN (2005) Genome evolution and biodiversity in teleost fish. *Heredity*, **94**, 280–294.

Yang Z (2002) Inference of selection from multiple species alignments. *Current opinion in genetics & development*, **12**, 688–694.

Yang Z (2006) *Computational molecular evolution*. Oxford University Press, Oxford.

Zhang J (2003) Evolution by gene duplication: an update. *Trends in ecology & evolution*, **18**, 292–298.

ACKNOWLEDGEMENTS

Naturally, this doctoral thesis would not have been possible without the help, support and advise of many people.

First and foremost I am very grateful to **Walter Salzburger** for being my PhD advisor. The help, advice and suggestions on experiments and manuscripts were very useful. Also your endless optimism when things did not go as planned and your door that was – literally – always open, helped me to continue the work and ultimately write this thesis. Thank you very much for given me the opportunity to work in your group.

I am also very thankful to **Dick Roelofs**, who instantly agreed to be part of my committee. Thank you for all the positive responses to my emails, the reference letters that you wrote for me and the fun times during the Gordon Research Conferences in New England.

Thanks to **Mathias Kölliker** for being the chair of my committee, the reference letters and general interest in my work.

I would like to thank all my co-authors on the different projects, especially **Olivia Roth** for such a good and fun collaboration on the immune challenge that we designed and performed together in Kiel. I am still a bit afraid to put my hand under blacklight and find out that I too still have some VIE tags in me.
Marco Colombo for preparing the RNAseq samples together and the subsequent analyses. Without the help of **Nicolas Boileau** we could not have finished all the qPCR analyses. Also I enjoyed supervising **Fabienne Kim** in her masterproject and working together on solving problems during the analyses of the results.

Furthermore, I would like to thank my current and past office mates, **Michael Matschiner**, **Moritz Muschick**, **Nicolas Boileau**, **Marius Rösti**, **Sara Mae Stieb** and **Marco Colombo** for the pleasant (and quiet) atmosphere in the office. Thanks for all the conversations, sharing food items and the many laughs.

Many people in the lab advised me, or commented on analyses, techniques, manuscripts, applications, project plans etc or shared data, pictures, protocols etc. Special thanks to **Daniel Berner**, **Hugo Gante**, **Laura Baldo**, **Emilia Santos**, **Britta Meyer**, **Michael Matschiner**, **Moritz Muschick**, **Adrian Indermauer**, **Corina Heule**, **Anya Theis**, **Fabio Cortesi** and **Sara Mae Stieb**.

Brigitte Aeschbach and **Nicolas Boileau** were great lab managers who always were able to provide answers and solutions to problems in the lab.

Thank to all the members of the **Salzburger group** for the great times during and after work, including the fieldwork trips to Nicaragua and Zambia. I did not have colleagues in this group, I have friends!

Lastly I would like to thank my family and friends, especially **my parents** and **my brother** and **sister** for their support, visits to Basel and interest in work. I would like to thank **Richard Kluin** who always encouraged me to go to meetings, workshops, fieldwork trips, and listened to me talking about work *–again-*.

This doctoral work was financially supported by the **European Research Council** and the **Swiss National Science Foundation** (grants to W. Salzburger) and a dissertation support grant from the **Freiwillige Akademische Gesellschaft Basel**.

CURRICULUM VITAE

Curriculum Vitae

Eveline T. Diepeveen

2013-current
Postdoctoral Researcher **Rosenblum Lab**, UC Berkeley (USA)

2009-2013
Ph.D. in Zoology* **Salzburger Lab**, University of Basel (CH)
Thesis: *Evolutionary genetics of naturally and sexually selected traits in cichlid fishes*
***** *Magna cum laude*

2008-2009
Lab assistant (HLA diagnostics) Sanquin Blood Supply Foundation, Amsterdam (NL)

2005-2008
M.Sc. in Biology **VU University**, Amsterdam (NL)
M.Sc. thesis (evo-devo) Beldade Lab, Leiden University (NL)
M.Sc. internship (evolutionary genetics) Hoekstra Lab, UC San Diego (USA)
M.Sc. internship (ethology and welfare) Dep Animal, Science and Society, Utrecht University (NL)

2002-2005
B.Sc. in Biology **VU University**, Amsterdam (NL)
B.Sc. internship (molecular evolution) National Museum of Natural History Naturalis, Leiden (NL)

Publications

Diepeveen ET, Roth O and Salzburger W (2013) Immune-related functions of the Hivep gene family in East African cichlid fishes. *G3: Genes, Genomics, Genetics* 3: 2205-2217.

Diepeveen ET, Kim FD and Salzburger W (2013) Sequence analyses of the distal-less homeobox gene family in East African cichlid fishes reveal signatures of positive selection. *BMC Evolutionary Biology* 13: 153.

Colombo M*, **Diepeveen ET***, Muschick M, Santos E, Indermaur A, Boileau N, Barluenga M and Salzburger W (2013) The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology* 22: 670-684. **Equal contribution*.

Diepeveen ET and Salzburger W (2012) Two decades of molecular ecology: where are we and where are we heading? *Molecular Ecology* 21: 5656-5659.

Diepeveen ET and Salzburger W (2011) Molecular characterization of two endothelin pathways in East African cichlid fishes. *Journal of Molecular Evolution* 73: 355-368.

Invited Talks

2012 Research seminar, Helmholtz Centre for Ocean Research (GEOMAR), Kiel (D)

Teaching

2014-current Grad student supervisor, UC Berkeley (USA); 1 student
2013-current Undergrad student supervisor, UC Berkeley (USA); 2 students
2011-2012 M.Sc student supervisor, University of Basel (CH), 1-year M.Sc project; 1 student
2010 B.Sc student supervisor, University of Basel (CH), 4-week B.Sc project; 2 students

Conference participation (with contribution)

2014 21st-Century Naturalists: integrating pattern and process to understand biodiversity, Pacific Grove (USA)
2012 Evolution Meeting, Ottawa (CA)
2011 Gordon Research Conference on Ecological and Evolutionary Genomics, University of New England, Biddeford (USA)
Biology 2011, University of Zurich (CH)
2010 Evolutionary and Ecological Genomics of Adaptation Symposium, University of Fribourg (CH)
Cichlid Science 2010 Meeting, Basel (CH)
European Meeting of PhD Students in Evolutionary Biology, Wierzba (PL)

Grants

2013 Basler Stiftung für Biologische Forschung (CH), Research Support Grant (CHF 10'000)
2012 Freiwillige Akademische Gesellschaft Basel (CH), Support Grant for Dissertation Completion (CHF 10'000)
Reisefonds University of Basel (CH), Travel Grant (CHF2'000)
2011 Reisefonds University of Basel (CH), Travel Grant (CHF 1'580)
2010 Reisefonds University of Basel (CH), Travel Grant (CHF 424)
Conference Funding University of Fribourg (CH), Travel Grant (CHF 105)
2007 VU-Fonds (NL), Travel Grant (€ 400)
St. Fonds Dr. Christine Buisman (NL), Travel Grant (€ 250)
St. Bekker-La Bastide-Fonds (NL), Travel Grant (€ 1'000)
St. Dr. Hendrik Muller's Vaderlands Fonds (NL), Travel Grant (€ 2'000)

Fieldwork experience

2011 Lake Tanganyika, Zambia (2,5 weeks). DNA, RNA tissue extractions in Tanganyikan cichlids, fishing with gill nets.
2009 Lake Apoyo and Xiloa, Nicaragua (3 weeks). DNA, RNA and several other tissue extractions in Nicaraguan cichlid species, fishing with gill nets.

Review Activities

BMC Evolutionary Biology

Memberships

American Society of Naturalists
Canadian Society for Ecology and Evolution
European Society for Evolutionary Biology