The complexity of alternative splicing of *hagoromo* mRNAs is increased in an explosively speciated lineage in East African cichlids

Yohey Terai*, Naoko Morikawa*, Koichi Kawakami†, and Norihiro Okada*‡§

*Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan; †Department of Developmental Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan; and ‡Division of Cell Fusion, National Institute of Basic Biology, 38 Nishigonaka, Myodaiji, Okazaki 444-8585 Aichi, Japan

Edited by Tomoko Ohta, National Institute of Genetics, Mishima, Japan, and approved August 29, 2003 (received for review May 12, 2003)

The adaptive radiation of cichlid fishes in the lakes of East Africa is a prime example of speciation. The choice of cichlid mates on the basis of a variety of coloration represents a potential basis for speciation that led to adaptive radiation. Here, we characterize the cichlid homolog of the zebrafish *hagoromo* (*hag*) gene that was recently cloned and characterized from a pigmentation mutant. Although only one *hag* mRNA was reported in zebrafish, cichlids express nine different *hag* mRNAs resulting from alternative splicing. The *hag* mRNAs are expressed between the myotome and the epidermis where pigment cells are located, suggesting the cichlid *hag* gene is involved in pigmentation. The *hag* mRNA splicing pattern does not fluctuate among individuals from each of two species, suggesting that alternative splice site choice is fixed within species. Furthermore, cichlids in lineages that underwent explosive speciation expressed a greater variety of *hag* mRNAs than those in lineages that did not undergo such a degree of speciation, suggesting that species in the explosively speciated lineage acquired a complex regulatory mechanism of alternative splicing over a very short evolutionary period. Here, we provide an example in which alternative splicing may play a role in mate choice, leading to cichlid speciation through diversification of gene function by production of multiple mRNAs from a single gene.

Lakes Victoria, Malawi, and Tanganyika in the East African Rift Valley harbor roughly >500, >500, and 200 endemic species of cichlid fishes, respectively (1–5). These fishes provide a spectacular example of adaptive radiation of living vertebrates (1, 2). The fishes exploit almost all resources available to freshwater fishes (1), and they are ecologically and morphologically highly diverse despite having evolved during a very short evolutionary period (6, 7).

The formation of color patterns in these fishes is important for mate choice and is therefore of particular interest from an evolutionary standpoint (8, 9). Cichlids in Lake Victoria can potentially interbreed without loss of fertility (10), but species are sexually isolated as a consequence of mate choice (8, 9). Male cichlids are known for their display of bright breeding coloration, and the ability of females to choose conspecific males on the basis of their color pattern has been documented extensively by observations in natural environments and behavioral experiments (8, 9). It is therefore expected that the genes responsible for formation of cichlid color pattern diversity must have either evolved rapidly via mutations or undergone dynamic changes such as alternative splicing that broadened their functional potential (11, 12) during adaptive radiation in each lake. Although cichlid body color is based on pigment pattern formation, the mechanisms underlying this process remain unknown. In zebrafish, various mutations affecting pigmentation have been described (13, 14), and some relevant genes have been cloned and characterized (e.g., *sparse*, *nacre*, and *hagoromo*; refs. 15–17).

The genes and mutations involved in zebrafish pigmentation may provide potential clues to pigmentation mechanisms in other fish, including cichlids. With insertional mutagenesis, the zebrafish gene *hagoromo* (*hag*) was shown to be responsible for pigment pattern formation (17). Recently, we cloned a cichlid homolog of *hag* and showed a correlation between the morphological diversity of the Great Lakes lineage and the evolutionary rate of amino acid sequence changes in the *hag* gene (18). Here we show a correlation between the explosive speciation of cichlids and the complexity of alternative splicing variants of this putative pigmentation gene. Diversification of the function of the gene in cichlids has been achieved by alternative splicing that evolved over a very short evolutionary period. Together, our results suggest that pigmentation diversity was attained through both amino acid substitutions and posttranscriptional regulatory mechanisms.

Materials and Methods

Determination of the Structure of the Cichlid *hag* Gene. Isolation of the cichlid *hag* cDNA was described in our previous work (18). Briefly, the cichlid *hag* gene was mapped by using inter-exon PCR with genomic DNA of *Labidochromis caeruleus* as a template. We used pairs of forward (F) and reverse (R) primers to amplify introns 1 (e1F, e2R); 2 (e2F, e3R); 3 (e3F, e4R); 4 (e4F, e5R); 5 (e5F, e6R); 6 (e6F, e7R); 7 (e7F, e8R); and 8 (e8F, e9R).

Detection of Each Type of *hag* mRNA. Total RNA (5 μg) extracted from whole bodies of several cichlid species by using TRIzol reagent (Invitrogen) was used for synthesis of cDNA with an oligo(dT) primer. The first round of PCR was performed to amplify all cDNAs by using 1% of the synthesized cDNAs as a template. Primers for the e9terminal type were e1F and e9R2, whereas those for the i5terminal type were e1F and XgoodR1. The PCR conditions were 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C for 35 cycles. By using 1 μl of 1% of the product of the first PCR as template, a second PCR was performed with the following conditions: 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C for 20 cycles, with pairs of primers to detect e9terminal-i2ret (i2F, e9R2); e9terminal-i3ret (i3F, e9R2); e9terminal-e3less (we3F, e9R2); i5terminal coding region (e1F, e5R); i5terminal-i1ret (i1F or i1FL for Lamprologini species, XgoodR1); i5terminal-i2ret (i2F, XgoodR1); i5terminal-i3ret (i3F, XgoodR1); and i5terminal-e3less (we3F, XgoodR1) (see Fig. 1*d*). mRNAs retaining intron 4 were probed by using primers i4F (specific to intron 4) and e9R2 under the same conditions described above.

***In Situ* Hybridization.** We used a 1-kb fragment of e9terminal-f1 (full length) cDNA to generate sense and antisense RNA probes

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: *hag*, *hagoromo*; F, forward; R, reverse.

§To whom correspondence should be addressed. E-mail: nokada@bio.titech.ac.jp.

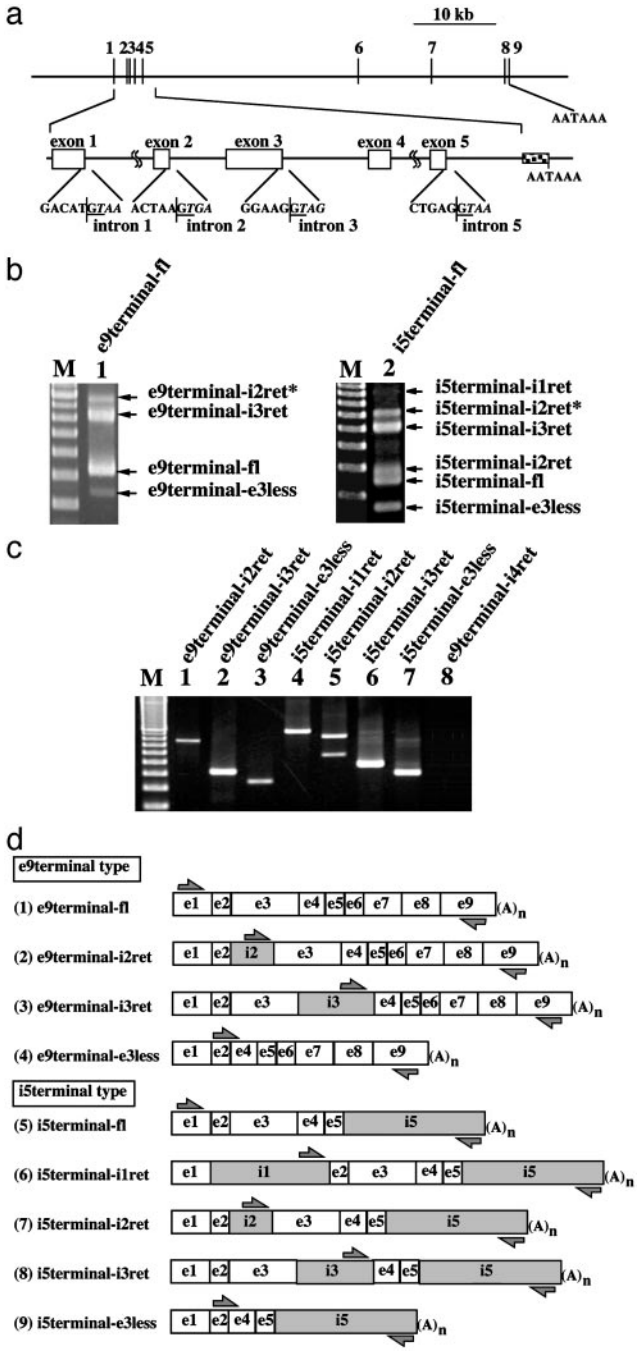

Fig. 1. (*a*) Structure of the cichlid *hag*. Positions of exons are indicated by numbers. Sequences at exon–intron boundaries are given for exons 1–5. Vertical lines delineate exon–intron junctions; conserved GT sequences for splicing are underlined, and termination codons are italicized. The checkered box indicates the conserved 3′UTR in the mRNAs terminating within intron 5 (i5terminal). (*b*) RT-PCR amplification of *hag* mRNAs. mRNAs terminating at exon 9 (e9terminal) and i5terminal mRNAs were amplified from *M. auratus* cDNA by RT-PCR (lanes 1 and 2). Positions of expected mobilities of amplified products of mRNAs are indicated to the right of lanes 1 and 2, respectively. Asterisks indicate intron 2- and 3-containing cDNA in lanes 1 and 2. (*c*) Nested PCR amplification of *hag* mRNAs. Intron-containing and exon-skipped mRNAs were amplified by nested PCR by using the RT-PCR products of e9terminal and i5terminal mRNAs as templates (lanes 1–8). The fragment was intron 2- and 3-containing e9terminal cDNA in lane 1. The longer fragments were intron 2- and 3-containing i5terminal cDNA in lane 5. ''M'' indicates a 200-bp ladder of size markers. (*d*) Schematic representation of *hag* mRNA structures. Exons are indicated by ''e,'' and introns are indicated by ''i.'' Half-arrows indicate positions of primers specific to each mRNA. Gray boxes represent introns.

with a digoxigenin labeling kit (Roche Applied Science). *In situ* hybridization was performed as described (19), except that 30 min was added to each washing time, and each sample was incubated in a solution of proteinase K until just before the sample began to fall apart (≈1 h, depending on the sample size).

Primers. Primers were designed on the basis of cichlid *hag* cDNA sequences (GenBank accession nos. AB107792 and AB107793). The sequences of the primers were as follows: e1F (5′-AGCCGCCTCTCTCAAGTTTGTAGAAGC-3′), e2R (5′-CATCTCAGAGGAATTCTTCTTCTGCAGAC-3′), e2F (5′-TGCAGAAGAAGAATTCCTCTGAGATGGA-3′), e3R (5′-CTTTGTGGATATGGTACACCTTTATGTCAG-3′), e3F (5′-CTGCTGACATAAAGGTGTACCATATCCACA-3′), e4R (5′-TCCGTCTTTAGCATCCACACAGTTCAC-3′), e4F (5′-TTCTCAGGTCATAACCAGGAGGTGAACT-3′), e5R (5′-ACTCTGTCACACATGGGAATTGTATCTCT-3′), e5F (5′-CAGAGATACAATTCCCATGTGTGACAGAGT-3′), e6R (5′-AGTGCCTAAGAATATCGATCGTTCAACATC-3′), e6F (5′-GAACTTCTCCCCTCTCTGCATCTGG-3′), e7R2 (5′-ACAGTCGAATAAAGGTATCATAACCACAGG-3′), e7F (5′-TCTCCGTTCCAGCTCTTCACCTGT-3′), e8R2 (5′-ATTCAGTCTGCCGTTTGTCCCAA-3′), e8F (5′-CCAGACAGATGGAAACCACATGATAGC-3′), e9R2 (5′-TCTGAAGTCCAGCGAATGCACAG-3′), XgoodR1 (5′-GTATCGCCCAGATTTCCTACAACAA-3′), i1F (5′-GATGAAAGGGTCAGATGTGCAT-3′), i1FL (5′-ATGTTGATGTGTCATCTGAACATGT-3′), i2F (5′-AACTGTTCATCATATCTCTGCCTGA-3′), i3F (5′-CTCCTGTCAGCATAAGAGAGGTCAAGACT-3′), i4F (5′-ATCTAAATCATCATGTGTTCAGTTATTGGT-3′), and we3F (5′-GAATTCCTCTGAGATGGAAAACTAATGAC-3′).

Results

Isolation of Various Types of *hag* mRNA from Cichlid Fishes. The exon–intron structure of the cichlid *hag* cDNA was determined (Fig. 1*a*) by using the *hag* gene cDNA from *L. caeruleus* (18). Unlike the zebrafish *hag* cDNA, the cichlid *hag* cDNA comprised two major types terminating at different polyadenylation sites (Fig. 1*a*). One type, designated ''e9terminal,'' ended at exon 9, whereas the other, designated ''i5terminal,'' ended within intron 5 with the polyadenylation site located ≈600 bp downstream of exon 5 (Fig. 1*a*). RNA of the e9terminal type and the i5terminal type was amplified from the cichlids *L. caeruleus* (data not shown) and *Melanochromis auratus* by using RT-PCR with primers specific to each type. Several shorter and longer fragments were detected in addition to the cDNA containing the expected number of exons (e9terminal-fl and i5terminal-fl, Fig. 1*b*). The relative levels of the different mRNAs were presumed to be comparable, because PCR for each analysis was performed under almost identical conditions.

Further cloning and sequence analyses revealed the presence of nine different *hag* mRNAs that could be grouped into four subtypes of the e9terminal type and five subtypes of the i5terminal type. The e9terminal types included e9terminal-fl that contained nine exons and resembled the zebrafish cDNA (17) [(1) in Fig. 1*d*]; e9terminal-i2ret that retained intron 2 [(2) in Fig. 1*d*]; e9terminal-i3ret that retained intron 3 [(3) in Fig. 1*d*]; and e9terminal-e3less that lacked exon 3 [(4) in Fig. 1*d*]. The i5terminal types included i5terminal-fl that contained the first five exons and a partial intron 5 sequence [(5) in Fig. 1*d*]; i5terminal-i1ret that retained intron 1 [(6) in Fig. 1*d*]; i5terminal-i2ret that retained intron 2 [(7) in Fig. 1*d*]; i5terminal-i3ret that retained intron 3 [(8) in Fig. 1*d*]; and i5terminal-e3less that lacked exon 3 ([(9) in Fig. 1*d*]).

Because a *hag* cDNA probe containing exons 2 and 3 hybridized to a single band on a genomic Southern blot (see Fig. 5, which is published as supporting information on the PNAS web site, www.pnas.org), we concluded that these various subtypes of

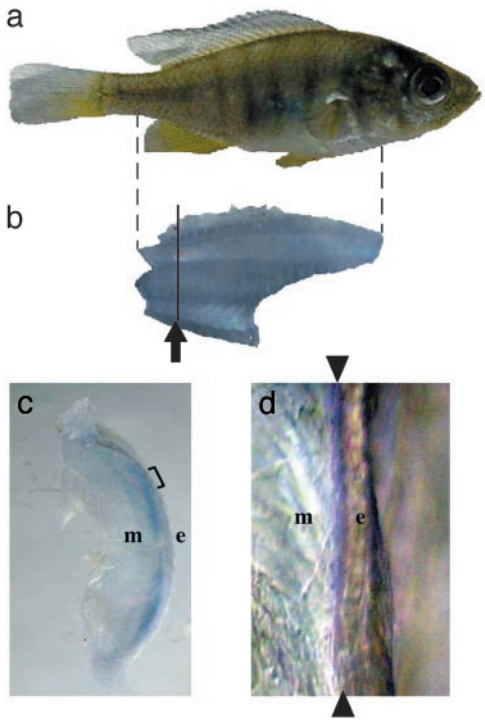

Fig. 2. Expression of *hag*. (*a*) A juvenile specimen of *Haplochromis sp*. from Lake Victoria at day 60 was used for *in situ* hybridization. (*b*) *In situ* hybridization by using full-length antisense RNA terminating at exon 9 (e9terminal-fl) as a probe (head and venter removed from specimen). The signal (blue) was detected in almost all areas of the skin. (*c*) Cross section through the solid line with arrow in *b*. The expression of *hag* (blue) was detected beneath the epidermis (e). The myotome is indicated by ''m.'' (*d*) Higher magnification of the bracketed region in *c*, showing *hag* expression between the m and the e between the arrowheads.

mRNA were generated from a single *hag* gene by alternative splicing by intron retention and/or exon skipping.

Expression of *hag* mRNAs. We confirmed the expression of each of the various mRNAs by PCR, some of which retained introns 1, 2, or 3, and some of which did not contain exon 3. In this experiment, we used the products amplified by PCR representing e9terminal-f1 and i5terminal-fl, respectively (Fig. 1*b*, lanes 1 and 2), as templates. The primers were designed to be specific for intron 1–3 or the exon 2–4 boundary (positions of primers are indicated in Fig. 1*d*). All of the mRNAs were clearly detected in our analysis (Fig. 1*c*, lanes 1–8). Because we detected retention of introns 1–3, we examined whether introns 4 and 8 could also be retained (introns 5–7 were too long to test). Introns 4 and 8 were not detected in any cichlid species, including *M. auratus* (Fig. 1*c*, lane 8 is an example of intron 4) and five species from Lake Malawi (data not shown; the species are described in Figs. 3 and 4). These data suggested that only introns 1–3 were retained.

To examine the pattern of *hag* gene expression in cichlids, we performed *in situ* hybridization with digoxigenin-labeled sense and antisense RNAs of e9terminal-f1 as probes. The antisense probe can potentially hybridize to all *hag* transcripts. In zebrafish, expression of the *hag* gene is ubiquitous during all stages of development and in almost all adult tissues (17). The expression of *hag* was detected in a skin sample from a whole-body skin preparation from *Haplochromis* sp. (the head was removed before analysis) (Fig. 2 *b* and *c*). At higher magnification, *hag* expression was detected between the myotome and the epidermis where the pigment cells are located (Fig. 2*d*), demonstrating that *hag* mRNAs are expressed specifically in this region. No hybridization was detected with the sense e9terminal-f1 RNA probe (data not shown).

Detection of *hag* mRNA Variants from 14 African Cichlid Species. *hag* mRNAs from 14 different cichlid species were analyzed to examine the extent of *hag* gene alternative splicing among African cichlids. The e9terminal-f1 mRNA (filled arrow, Fig. 3*a*) and i5terminal-f1 mRNA were detected in all 14 species (filled arrows, Fig. 3 *e* and *f*). Because the length of the 3′UTR of i5terminal mRNAs varied in three species (Fig. 3*e*, lanes 1–3), only the coding region of i5terminal-f1 mRNA was amplified (Fig. 3*f*). All species of the recently evolved cichlids from the Tropheini tribe in Lake Tanganyika, from the haplochromine in Lake Victoria and Lake Malawi, exhibited at least six of nine *hag* mRNAs (Figs. 3 *a–j*, lanes 7–24), whereas fewer types of *hag* mRNAs were detected in cichlids from rivers and Lake Tanganyika (Figs. 3 *a–j*, lanes 1–6). The amount of each mRNA was estimated over a range extending from 1/100 to 10 times that of the e9terminal-f1 subtype (Figs. 1*b* and 3 *a* and *e*).

To examine whether the alternative splicing pattern of *hag* may vary between two genetically closely related species displaying different color morphs, we analyzed the splicing pattern of *Dimidiochromis compressiceps* and *Dimidichromis strigatus* in Lake Malawi (Fig. 3, lanes 14 and 20, and Fig. 4). *D. compressiceps* and *D. strigatus* expressed seven and six splicing variants of *hag*, respectively, and their splicing patterns differed (Fig. 3, lanes 14 and 20). In *D. compressiceps*, the variants of e9terminal-e3less and i5terminal-i2ret were expressed, and those variants were not expressed in *D. strigatus* (Fig. 3 *d* and *h*, lanes 14 and 20). In *D. strigatus*, the variant of i5terminal-i1ret was expressed, and that variant was not expressed in *D. compressiceps* (Fig. 3*g*, lanes 14 and 20).

Splicing patterns were examined in *D. compressiceps* and *D. strigatus* (five specimens each) and found to be identical within each species (Fig. 3 *a–j*, lanes 15–19 for *D. compressiceps*, lanes 20–24 for *D. strigatus*). These results provide evidence that the splicing pattern is species-specific, and consistent within each of these two species.

Discussion

The Expression of Cichlid *hag* Is Controlled by a Complex Posttranscriptional Regulatory Mechanism. Alternative splicing constitutes a mechanism by which multiple mature mRNAs may be generated from a single pre-mRNA transcript. This mechanism is widely used to enhance protein diversity because the mature mRNAs may encode proteins with subtle or opposing functional differences (20). Several types of alternative splicing have been identified (20). Three types of *hag* transcript variants were found in African cichlid species, namely those with different transcriptional termination sites, those that retained introns (mRNAs are mature, but introns are not removed from pre-mRNA), and those that skipped exons. Ultimately, nine distinct transcripts were generated (Fig. 1*d*). Three lines of evidence suggest that the intron-containing mRNAs most likely represent completely processed products rather than pre-mRNAs isolated before complete splicing. First, the amount of each splicing variant is comparable with that of e9terminal-f1 and i5terminal-f1 types (Fig. 1*d*). Second, no mRNAs that contained either intron 4 or 8 were identified (e.g., Fig. 1*c*, lane 8), and if intron-containing *hag* mRNAs constitute partially processed premRNAs, then we should have detected variants containing each individual intron. Third, the retained introns differed among species (Fig. 3). For example, no intron-containing mRNA was detected in *Altolamprologus calvus* (Fig. 3, lane 3), although many types of intron-containing *hag* transcripts were found in Lake Malawi species (Fig. 3, lanes 9–14). If intron-containing *hag* mRNAs are truly pre-mRNAs, then they should be detectable in all species.

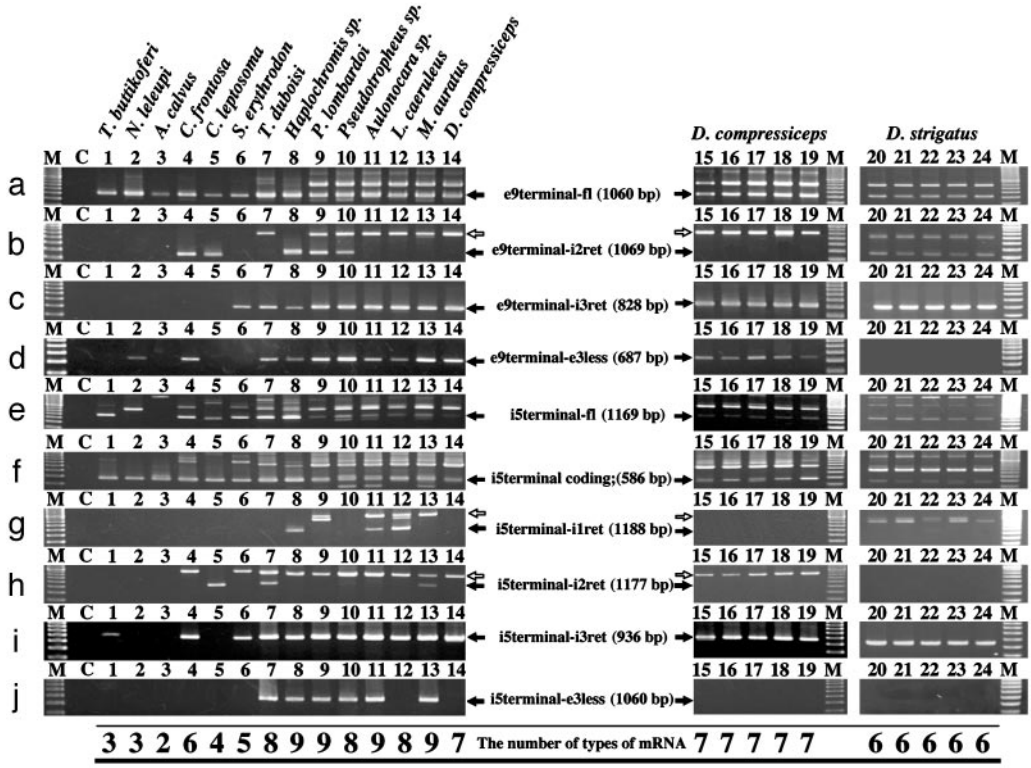

Fig. 3. Examination of intron and exon content from 14 cichlid species by using RT-PCR and nested PCR. The following mRNAs from various species were amplified by RT-PCR and nested PCR: (*a*) e9terminal-fl, (*b*) e9terminal-i2ret, (*c*) e9terminal-i3ret, (*d*) e9terminal-e3less, (*e*) i5terminal-fl, (*f*) the coding region of i5terminal, (*g*) i5terminal-i1ret, (*h*) i5terminal-i2ret, (*i*) i5terminal-i3ret, and (*j*) i5terminal-e3less. Filled arrows indicate expected mobilities of amplified products of each mRNA. (*a*) The longer and shorter fragments relative to e9terminal-fl were intron-containing and exon 3-less e9terminal cDNAs, respectively. (*b*) The longer fragments (open arrow), relative to e9terminal-i2ret, were intron 2- and 3-containing e9terminal cDNAs. (e) The longer and shorter fragments, relative to i5terminal-fl, were intron-containing and exon 3-less i5terminal cDNAs. (*f*) The longer and shorter fragments, relative to i5terminal coding region, were intron-containing and exon 3-less i5terminal cDNAs. (*g*) The longer fragments (open arrow), relative to i5terminal-i1ret, were intron 1- and 3-containing i5terminal cDNAs. (*h*) The longer fragments (open arrow), relative to i5terminal-i2ret, were intron 2- and 3-containing i5terminal cDNAs. Each of five individual *D. compressiceps* and *D. strigatus* specimens were used for lanes 15–19 and lanes 20–24, respectively. Refer to Fig. 4 for full names of species.

Translation of the different *hag* mRNA species would presumably result in multiple *hag* proteins. The e9terminal-fl cDNA encodes a putative protein of 389 amino acids that is 64% identical to the zebrafish *hag* gene product. Sequence analysis indicates that the exon 3-less e9terminal mRNA may be translated to the end of exon 9 without any frame shift. In the case of the e9terminal-i2ret, e9terminal-i3ret, i5terminal-fl, i5terminal-i1ret, i5terminal-i2ret, and i5terminal-i3ret mRNAs, translation may yield truncated proteins that terminate at a termination codon adjacent to the 5′ splice site of the retained introns (Fig. 1 *a* and *c*). Exon 3-less i5terminal mRNA may be translated to the end of exon 5 without any frame shift. To our surprise, we noted that the T residues in the conserved GT sequence at the 5′ splice site of introns 1, 2, 3, and 5 are used as the first nucleotide of the termination codons TAA, TGA, or TAG (Fig. 1*a*). The termination codons in these introns are conserved in the cichlid species examined in this study. The distance between exon 5 and the polyadenylation signal within intron 5 varied, but the sequence ≈120 bp upstream from the first polyadenylation signal (Fig. 1*a*, checkered box) in the 3′UTR sequences is strongly conserved in various cichlid species (data not shown). The conservation of the 3′UTR sequence in mRNAs of the i5terminal type may be related to some unidentified function, such as termination of transcription, control of translation, or mRNA stabilization. The functions of these putative truncated proteins are unknown, but our results indicate that expression of the cichlid *hag* gene is controlled by extremely complex posttranscriptional regulatory mechanisms.

As shown in Fig. 3, the splicing pattern is species-specific and is consistent among individuals within each of two species that exhibit distinct color patterns. Thus, the mechanism of alternative splicing of *hag* is species-specific, indicating that the pattern of alternative splicing may correlate with species-specific characteristics of color pattern.

Fig. 2 *b* and *c* show the expression of *hag* in the skin. Expression was observed between the myotome and the epidermis, where the pigment cells are located (Fig. 2*d*), suggesting that the cichlid *hag* gene is indeed involved in pigmentation. The product of the *hag* gene is a member of the family of F-box/WD-repeat proteins in which the F-box and WD-repeats are fused (17). The proteins in the F-box/WD-repeat family function in a variety of regulatory pathways including cell division control (21, 22), sulfur metabolism (23), and differentiation (24). Moreover, all F-box/WD-repeat proteins are involved in protein degradation. The localized expression of *hag* suggests that it may represent one of the genes responsible for the regulation of pigment pattern formation via a protein degradation pathway.

The production of multiple mRNAs by alternative splicing from genes responsible for morphogenesis may be a common phenomenon in cichlids. To examine this possibility, we used RT-PCR to analyze three morphogenetic genes (*Bmp4*, *Bmp2*, and *Pax9*) in cichlids but failed to detect any alternative splicing (data not shown). Thus, the generation of multiple mRNAs from a single gene might have evolved specifically in the cichlid *hag* gene.

Alternative splicing is known to expand the diversity of

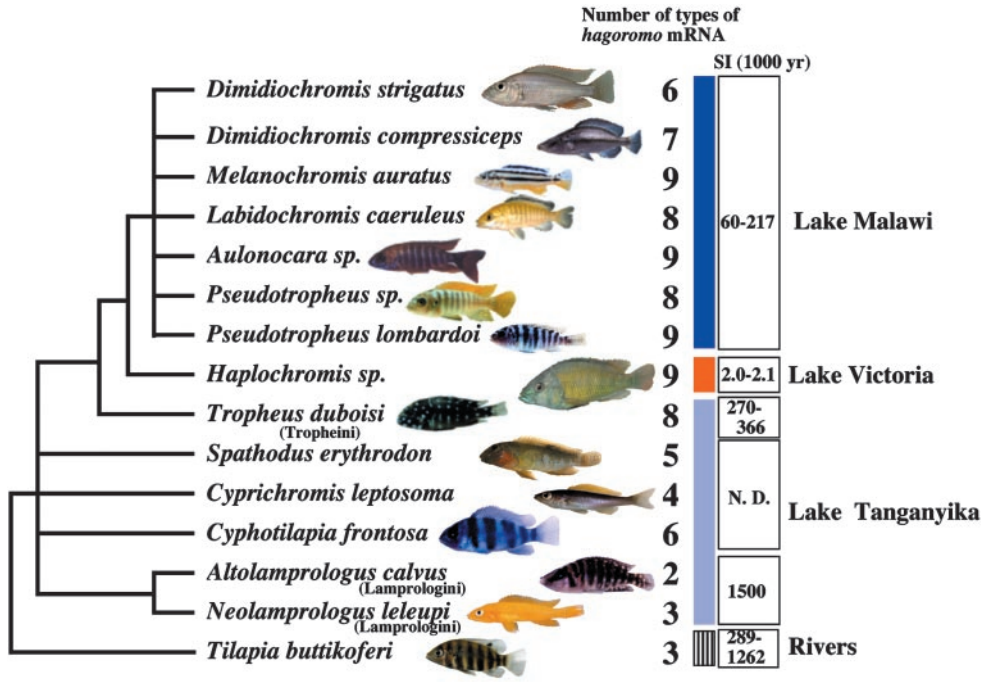

Fig. 4. Phylogeny of the cichlid species examined in this study, showing the numbers of *hag* mRNA types found in each species and the speciation rates in each lineage. The phylogenetic tree is based on molecular data from several sources (29, 41–47). The net speciation intervals (SI) for Lakes Malawi and Victoria haplochromine, Lamprologini, and river species (tilapiine) were calculated in previous works (31, 32). The SI for *Tropheus* was calculated on the basis of the method described in refs. 30 and 32 by using the age of radiation for *Tropheus* (48). N.D., no data, because the age of radiation for each lineage was not deduced.

proteins. For example, alternative splicing of a *Drosophila* homolog of the cell adhesion molecule, *Dscam* (associated with Down syndrome), may potentially generate >38,000 isoforms (25). The potential molecular diversity of Dscam may contribute to the specificity of neuronal connectivity (25). The estimated >500 mRNAs of the calcium-activated potassium channel gene, *slo*, are thought to mediate inner-ear hair cell tuning in response to a unique range of sound (26). More than 1,000 different mRNAs for neurexins, a family of neuropeptide receptor genes, may be expressed by alternative splicing from three genes (27). Alternative splicing of these genes was shown to impart functional diversity of the various gene products in different organs.

At least nine different mRNAs that can be generated from the *hag* gene are expressed in cichlids. The organ in which *hag* expression was detected, namely the skin, is also the organ that exhibits morphological diversity. Skin pigmentation represents one of the most diverse morphological characteristics among these fishes (Figs. 2 and 4). The possible regulation of intron-less mRNA gene products by intron-containing mRNA gene products has been reported with regard to phenylethanolamine *N*-methyltransferase (28). Thus, it is possible that truncated *hag* proteins translated from intron-containing mRNAs might functionally counteract *hag* proteins translated from intron-less mRNA types.

Species in the Explosively Speciated Lineage Have Acquired Complex Splicing Mechanisms Within a Very Short Evolutionary Period. It has been estimated that Lake Malawi and Lake Victoria flocks are 700,000 (7) and <12,400 (6) years old, respectively, and the species in these lakes are believed to have speciated within these very short time periods. Fig. 4 shows the phylogenetic relationship among the cichlid species used in this study, along with the number of *hag* mRNA types. We detected six to nine *hag* mRNA types in species from the Tropheini tribe in Lake Tanganyika, from the haplochromine in Lake Victoria and Lake Malawi (Fig. 3, lanes 7–24, and Fig. 4) but only two or three in riverine and old-lineage species (Lamprologini) from Lake Tanganyika (ref. 29; Fig. 3, lanes 1–3; Fig. 4). To test a correlation between the species diversity and the complexity of alternative splicing of the *hag* gene, we compared speciation rates for each lineage with the complexity of alternative splicing (Fig. 4). Net speciation intervals (mean doubling time of species number; refs. 30–32) were shorter in Lakes Victoria and Malawi lineages and longer in Lamprologini and river lineages (Fig. 4). Thus, there appears to be a correlation between species diversity in the explosively speciated lineage and the complexity of alternative splicing of the *hag* gene. Species in the explosively speciated lineage appear to have acquired more complex mechanisms for regulating *hag* mRNA splicing.

How does the complexity of alternative splicing of *hag* correlate to species diversity? The most species-rich lineages of cichlids are polygynous with male nuptial coloration that leads to speciation by sexual selection, and the speciation rates in polygynous lineages are higher than in monogamous lineages (30). The differences in the complexity of *hag* alternative splicing between Lamprologini (monogamous lineage) and Lakes Malawi and Victoria haplochromine species (polygynous lineage) may correlate with differences in the intensity of sexual selection within the different mating systems.

Cichlids, especially those in Lake Victoria, show many color variations that lead to mate choice among individuals in one species and/or in a local population (5). It is believed that the color morph variations evolved in <12,400 years (6), although the diversity of the color morph variations and the genetic diversity of the color pattern formation might have been acquired in ancestral populations (33). Several examples of changes in splicing patterns are caused by mutations in splicing motifs leading to phenotypic changes (34). Sex determination in *Drosophila melanogaster* is one of the well known examples. Sex of *D. melanogaster* is phenotypically regulated by inhibition or activation of regulatory protein(s) binding at splice sites of genes such as sex-lethal (*Sxl*), transformer (*tra*), and double sex (*dsx*),

resulting in alternative splicing (35–37). Thus, a few mutations in the splicing motif may change the alternative splicing pattern, thereby dramatically altering the phenotype. This mechanism makes it feasible to evolve complex splicing mechanisms to generate phenotypic changes dramatically over a very short time period.

The regulatory mechanisms of *hag* splicing are expected to be similar to those of other alternatively spliced genes. It is possible that as-yet-unknown RNA-binding proteins that inhibit or activate *hag* pre-mRNA splicing regulate *hag* alternative splicing. A minimal number of mutations within mRNA-binding proteins or in the protein-binding sites within the *hag* gene may either induce new splice variants or change the ratio of splice variants. Relatively large alterations in pigment pattern phenotypes could arise from these mechanisms. Thus, we may now investigate the mutations responsible for generating new alternatively spliced transcripts in different cichlid species and study their effects by introducing mutated *hag* genes into cichlids by microinjection (38).

Our recent work demonstrated an accelerated rate of amino acid changes within *hag* gene products in species of the Great Lakes lineage (namely, species from Lakes Tanganyika, Malawi, and Victoria) (18). In addition, our present work demonstrates that the complexity of alternative splicing mechanisms within the *hag* gene increases in an explosively speciated lineage (namely, species from the Tropheini tribe in Lake Tanganyika, from the haplochromine in Lake Victoria and Lake Malawi). Thus, the diversity of the *hag* gene has been achieved both by amino acid replacements (18) and by increasing the complexity of alternative splicing during adaptive radiation of lake cichlids. Perhaps the increased diversity of the *hag* gene accelerated the frequency of speciation through the generation of multiple color varieties leading to mate choice in cichlids.

The segregation of a population through reproductive isolation is the first step toward speciation, and a few genes thought to participate in this process have been identified (39, 40). The *hag* gene may be a strong candidate for such a gene because its expression, known to affect pigment pattern formation in zebrafish, is regulated uniquely in different cichlid species. The present study provides a molecular clue toward elucidating the mechanism of explosive speciation in cichlids in the lakes of the East African Rift Valley.

This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

1. Fryer, G. & Iles, T. D. (1972) *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution* (Oliver and Boyd, Edinburgh).
2. Greenwood, P. H. (1991) in *Cichlid Fishes: Behaviour, Ecology and Evolution*, ed. Keenleyside, M. H. A. (Chapman & Hall, London), pp. 86–102.
3. Snoeks, J., Ruber, L. & Verheyen, E. (1994) *Arch. Hydrobiol. Beiheft. Ergebnisse Limnol.* **44,** 355–372.
4. Konings, A. (1995) *Malawi Cichlids in Their Natural Habitat* (Cichlid Press, St. Leon-Rot, Germany), 2nd Ed.
5. Seehausen, O. (1996) *Lake Victoria Rock Cichlids: Taxonomy, Ecology, and Distribution* (Verduijn Cichlids, Zevenhuizen, The Netherlands).
6. Johnson, T. C., Scholz, C. A., Talbot, M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & Mcgill, J. W. (1996) *Science* **273,** 1091–1093.
7. Mayer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. (1990) *Nature* **347,** 550–553.
8. Seehausen, O., van Alphen, J. J. M. & Witte, F. (1997) *Science* **277,** 1808–1811.
9. Seehausen, O. & van Alphen, J. J. M. (1998) *Behav. Ecol. Sociobiol.* **42,** 1–8.
10. Crapon de Caprona, M. D. (1996) *Ann. Mus. R. Afr. Cent. Sci. Zool.* **251,** 117–120.
11. Leff, F. E., Rosenfeld, M. G. & Evans, R. M. (1986) *Annu. Rev. Biochem.* **55,** 1091–1118.
12. Breibart, R. E., Andreadis, A. & Nadal-Ginard, B. (1987) *Annu. Rev. Biochem.* **56,** 467–495.
13. Johnson, S. L., Africa, D., Walker, C. & Weston, J. A. (1995) *Dev. Biol.* **167,** 27–33.
14. Haffter, P., Odenthal, J., Mullins, M. C., Lin, S., Farrell, M. J., Vogelsang, E., Haas, F., Brand, M., van Eeden, F. J. M., Furutani-Seiki, M., *et al.* (1996) *Dev. Genes Evol.* **206,** 260–276.
15. Parichy, D. M., Rawls, J. F., Pratt, S. J., Whitfield, T. T. & Johnson, S. L. (1999) *Development (Cambridge, U.K.)* **126,** 3425–3436.
16. Lister, J. A., Robertson, C. P., Lepage, T., Johnson, S. L. & Raible, D. W. (1999) *Development (Cambridge, U.K.)* **126,** 3757–3767.
17. Kawakami, K., Amsterdam, A., Shimoda, N., Becker, T., Mugg, J., Shima, A. & Hopkins, N. (2000) *Curr. Biol.* **10,** 463–466.
18. Terai, Y., Morikawa, N., Kawakami, K. & Okada, N. (2001) *Mol. Biol. Evol.* **19,** 574–578.
19. Poss, K. D., Shen, J., Nechiporuk, A., McMahon, G., Thisse, B., Thisse, C. & Keating, M. T. (2000) *Dev. Biol.* **222,** 347–358.
20. Graveley, B. R. (2001) *Trends Genet.* **17,** 100–107.
21. Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. (1997) *Cell* **91,** 221–230.
22. Kominami, K. & Toda., T. (1997) *Genes Dev.* **11,** 1548–1560.
23. Natorff, R., Piotrowska, M. & Paszewski, A. (1998) *Mol. Gen. Genet.* **257,** 255–263.
24. Jiang, J. & Struhl, G. (1998) *Nature* **391,** 493–496.
25. Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E. & Zipursky, S. L. (2000) *Cell* **101,** 671–684.
26. Black, D. L. (1998) *Neuron* **20,** 165–168.
27. Ullrich, B. Ushkaryov, Y. A. & Sudhof, T. C. (1995) *Neuron* **14,** 497–507.
28. Unsworth, B. R., Hayman, G. T., Carroll, A. & Lelkes, P. I. (1999) *Int. J. Dev. Neurosci.* **17,** 45–55.
29. Nishida, M. (1991) *Experientia* **47,** 974–979.
30. Turner, G. F. (1999) in *The Evolution of Biological Diversity*, eds. Magurran, A. E. & May, R. M. (Oxford Univ. Press, Oxford), pp. 217–229.
31. Seehausen, O. (2000) *Adv. Ecol. Res.* **31,** 237–274.
32. Seehausen, O. (2002) *Proc. R. Soc. London Ser. B* **269,** 491–497.
33. Seehausen, O., Koetsier, E., Schneider, M. V., Chapman, L. J., Chapman, C. A., Knight, M. E., Turner, G. F., van Alphen, J. J. M. & Bills, R. (2003) *Proc. R. Soc. London Ser. B* **270,** 129–137.
34. Nissim-Rafinia, M. & Kerem, B. (2002) *Trends Genet.* **18,** 123–127.
35. Inoue, K., Hoshijima, K., Sakamoto, H. & Shimura, Y. (1990) *Nature* **344,** 461–463.
36. Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H. & Shimura, Y. (1991) *Science* **252,** 833–836.
37. Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H. & Shimura, Y. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8092–8096.
38. Houdebine, L. M. & Chourrout, D. (1991) *Experientia* **15,** 891–897.
39. Ting, C. T., Tsaur, S. C., Wu, M. L. & Wu., C. I. (1998) *Science* **282**, 1501–1504.
40. Terai, Y., Mayer, W. E., Klein, J., Tichy H. & Okada, N. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 15501–15506.
41. Kocher, T. D., Conroy, J. A., McKaye, K. R. & Stauffer, J. R. (1993) *Mol. Phylogenet. Evol.* **2,** 158–165.
42. Sturmbauer, C. & Meyer, A. (1993) *Mol. Biol. Evol.* **10,** 751–768.
43. Meyer, A. (1993) *Trends Ecol. Evol.* **8,** 279–284.
44. Mayer, W. E., Tichy, H. & Klein, J. (1998) *Heredity* **80,** 702–714.
45. Takahashi, K., Terai, Y., Nishida, M. & Okada, N. (2001) *Mol. Biol. Evol.* **18,** 2057–2066.
46. Salzburger, W., Meyer, A., Baric, S., Verheyen E. & Sturmbauer, C. (2002) *Syst. Biol.* **51,** 113–135.
47. Terai, Y., Takahashi, K., Nishida, M., Sato, T. & Okada, N. (2003) *Mol. Biol. Evol.* **6,** 924–930.
48. Baric, S., Salzburger, W. & Sturmbauer, C. (2003) *J. Mol. Evol.* **56,** 54–68.